ADVANCES IN INTELLIGENT SYSTEMS AND COMPUTING 174

Aswatha Kumar M.
Selvarani R.
T.V. Suresh Kumar (Eds.)

Proceedings of International Conference on Advances in Computing

Springer

Advances in Intelligent Systems and Computing 174

For further volumes:
http://www.springer.com/series/11156

Aswatha Kumar M., Selvarani R.,
and T.V. Suresh Kumar (Eds.)

Proceedings of International Conference on Advances in Computing

Springer

Editors
Dr. Aswatha Kumar M.
M.S. Ramaiah Institute of Technology
Bengaluru
India

Dr. T.V. Suresh Kumar
M.S. Ramaiah Institute of Technology
Bengaluru
India

Dr. Selvarani R.
M.S. Ramaiah Institute of Technology
Bengaluru
India

ISSN 2194-5357
e-ISSN 2194-5365
ISBN 978-81-322-0739-9
e-ISBN 978-81-322-0740-5
DOI 10.1007/978-81-322-0740-5
Springer New Delhi Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012941516

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

We are very happy to bring this Proceedings of International Conference on Advances in Computing 2012 (ICAdC 2012). This Conference is being organized by the departments of Information Science & Engineering, Computer Science & Engineering, and Computer Applications of M.S. Ramaiah Institute of Technology (MSRIT), Bengaluru, Karnataka, India. This programme is being organized under the Golden Jubilee celebrations of MSRIT (Estd. 1962). We are happy to organize the event at MSRIT, Bengaluru, where the idea for this Conference was blossomed. The participants to visit Bengaluru in July may experience rain and wind opportunities. The themes of the Conference, viz., Theoretical Computer Science, Systems and Software, and Intelligent Systems, cover the broader spectrum of computing and its advances; these are the common factors of all the three departments. The Conference tries to provide the platform to all senior and young researchers to share their knowledge as well to have networking for strengthening their research activities. We have received overwhelming response from all corners but could not accommodate them all. We are extremely delighted to receive research contributions from different parts of the world, including Australia, United Kingdom, South Africa, Ethiopia, Iceland, Vietnam, Iran, and Mauritius. It is true that the success of such events depends on the quality of the papers and on the efforts of the Conference organizers. All the contributions are peer reviewed by program committee members/technical committee reviewers. The reviews are focussed primarily on originality, quality and relevance to the theme of the Conference. As large numbers of contributions were accepted, it was decided to have parallel sessions so that interested groups may interact with other researchers in their interested research areas. We convey our special thanks to all the authors for submitting their research outcomes to this Conference, to the program committee/technical review committee, and to numerous reviewers who did an excellent job in guaranteeing that the articles in this volume are of good quality.

We also thank all the committee members for their efforts to make the Conference a grand success. The editors profoundly thank the management for the generous financial support extended to this Conference.

Aswatha Kumar M.
Selvarani R.
T.V. Suresh Kumar

Organization

M.S. Ramaiah Institute of Technology
International Conference on Advances in Computing -2012 (ICAdC-2012)

Conference General Chair
K. Rajanikanth, MSRIT, Bengaluru, India

Program Chairs
Aswatha Kumar M., MSRIT, Bengaluru
Selva Rani, MSRIT, Bengaluru
T.V. Suresh Kumar, MSRIT, Bengaluru

Advisory Committee

Prof. B.N. Chatterji	Indian Institute of Technology Kharagpur, Kolkata
Dr. V. Gopalakrishna	Director, Integra Micro Systems
Jayanta Mukhopadhyay	Indian Institute of Technology, Kharagpur
Y. Narahari	Indian Institute of Science, Bengaluru
S.S. Iyengar	Florida International University, Miami, USA
Dr. C.P. Ravikumar	Texas Instruments India, Bangalore 560093, India
Raghu Nambiar	Siemens Ag, Bangalore, India
Ravi Ramaswamy	Philips Electronics India Ltd. Bangalore, India
Dr. Shantanuchaudhury	Indian Institute of Technology, Delhi
Dr. N.L. Sarda	Indian Institute of Technology Bombay, Mumbai, India
Krishna M. Kavi	Dept. of CSE, Denton, Texas, USA
K. Chidanandagowda	Former Vc, Kuvempu University, Mysore
Bhabatoshchanda	Indian Statistical Institute, Kolkata, India
K.R. Murali Mohan	Department of Science e& Technology New
S. Ramesh Babu	Infosys, Bangalore
John Tsitsiklis	Massachusetts Institute Of Technology, Cambridge, USA
Dr. G. Athithan	Centre for Ai and Robotics, Bangalore, India
Ajith Abraham	Machine Intelligence Research Lab, Washington, USA
Meldal, Sigurd	San Jose State University, USA

Program Committee

Prof. Vasanth Honnavar	Iowa State University Ames, Iowa, USA
Prof. Rajkumar Buyya	The University of Melbourne, Melbourne, Australia
Prof. N. Sundararajan	Nanyang Technological University, Singapore
Prof. Ruppa Thulasiram	University of Manitoba, Canada
Dr. Rajib Mall	Indian Institute of Technology, Kharagpur, India
Dr. Lokesh Boregowda	Honeywell Technology Solutions Lab, Bangalore
Dr. Shyam Vasudevarao	Philips Healthcare, Philips Electronics India Ltd, Bangalore, India
Dr. David K. Kahaner	Asian Technology Information Program (ATIP), Japan
S.R. Mahadeva Prasanna	Indian Institute of Technology, Guwahati
C.V. Jawahar	International Institute of Information Technology, Hyderabad, India
Francis Chi Moon Lau	The University of Hong Kong, Hong Kong
Bhabani P. Sinha	Indian Statistical Institute(ISI), Kolkata, India
Muthu Ramachandran	Leeds Metropolitan University, Leeds, United Kingdom
Umapada Pal	Indian Statistical Institute(ISI), Kolkata, India
ShrishaRao	International Institute of Information Technology, Bangalore, India
C.B. Akki	Wipro Technologies, Electronics City, Bengaluru-560100
M. Khalid	Professor & Dean, VIT University, Vellore-632014, Tamilnadu
N.K. Srinath	R V College of Engg., Mysore Road, Bengaluru
Niranjan U.C.	Research and Training, Manipal-576104
T.N. Nagabhushan	S.J. College of Engg., Mysore, India
V.R. Udupi	Gogte Institute of Technology, Belgaum, India
K.S. Shreedhara	UBDT College of Engineering, Davangere
Sriman Narayana Iyengar	VIT University, Vellore, India
Umesh Bellur	Indian Institute of Technology, Bombay, India
Nitin Auluck	Indian Institute of Technology, Ropar, India
Venkatesh Prasad Ranganath	Microsoft Research, India (Software Engg.), Bangalore, India
R. Venkatesh Babu	Indian Institute of Science, Bengaluru, India
Sundaram Suresh	Nanyang Technological University, Singapore
Ranga V. Narayanan	Samsung INDIA SOFTWARE OPERATIONS PVT LTD., Bangalore, India
Ajith Bopardikar	Samsung India Software Operations, Bangalore, India
M.S. Dinesh	Siemens Information Systems, Bengaluru, India

Organizing Committee

Department of Computer Science & Engg

Dr. R. Selvarani
Prof. S.M. Narayana
Dr. Vijaya Kumar B.P.
Dr. K.G. Srinivasa
Anita Kanavalli
Seema S.
Annapurna P. Patil
Jagadish S.K.
Jayalakshmi D.S.
Monica R. Mundada
Sanjeetha R.
A. Parkavi
Veena G.S.
J. Geetha
T.N.R. Kumar
Ramanagouda S. Patil
Mamatha Jadhav V.
Suvarna
Chethan C.T.
Sini Anna Alex
Sardar Vandana Sudhakar
Meera Devi A. Kawalgi
Malle Gowda M.
R. Manoranjitham
H.V. Divakar
Chandrika Prasad
Leelavathi Rathod
Rajarajeshwari
Sowmyarani C.N.
Sunanda V.K.

Department of Information Science & Engineering

Dr. Aswatha Kumar M.
Dr. Lingaraju G.M.
N. Ramesh
Rajaram M. Gowda
Mydhili K. Nair
S.R. Chickerur
Shashidhara H.S.
George Philip
T. Tamilarasi
Savita K. Shetty
Myna A.N.
Deepthi K.
P.M. Krishnaraj
Rajeshwari S.B.
Siddesh G.M.
Pushpalatha M.N.
Mohan Kumar S.
Sumana M.
Prashanth Kambli
Naresh E.
Jagadeesh Sai D.
Mani Sekhar S.R.
Suresh Kumar K.R.
Sunitha R.S.
Sandeep B.L.
Dayananda P.

Department of MCA

Dr . T.V. Suresh Kumar
S. Ajitha
S. Jagannatha
D. Evangelin Geetha
Madhu Bhan
M. Mrunalini
Chethan Venkatesh
Sailaja Kumar
Seema D.
Manish Kumar
Niranjana Murthy M.

Technical Note

Computing power is increasing day by day, making the impossible become possible. We can create and visualize virtual world, speeding up time to simulate the creation of the universe or slowing down time to understand the interaction of the most basic particles of matter. Technology has grown to a wider dimension in all the sectors and has made everything virtual, is driven by the advancement in technology. However, this seems to apply mainly to developed nations, while developing countries still struggle with antediluvian machinery and systems. The primary sector that supplicates rapid development is intelligent computing and ICT, all are interdependent. The role of Intelligent Computing and ICT in a developing society cannot be exaggerated. Hence there is a need for academicians to take a major role in providing a smooth takeover of advanced intelligent computing and ICT in the young mindsets and interested researchers.

The development of technology in the domain of intelligent computing, communication and Information providing an end user/terminal requirement has given the way for integrated system design and development. Similarly, rapid advances in modern high speed networks and wireless/mobile networks, with the support of Internet growth, have produced tremendous research and commercial opportunities in the areas of mobile multimedia Networks (2G, 3G, and 4G), ubiquitous and pervasive computing systems. The main focus of this International Conference – ICAdC-2012 is to provide the benefits to the academicians, researchers and industrialist to bring in a smooth transition in the developing countries. Some of them include the following.

- Provide a platform to engineers, academicians, budding research scholars, Ph.D. scholars of various Engineering Colleges/ Universities/ Institutes to show case their research in the field of education and develop aptitude for writing technical papers.
- Discuss the future direction of research and new technologies, which can be helpful in research.
- Networking the professionals from various sectors with academic institutions for better education and work flow.
- Promote the benefits of the applications of Standards and Ethics in the government, academia and industry.

ICAdC-2012 seeks to bring together international researchers to present paper and generate discussions on current research and development in all aspects of IT.

Special Emphasis will be made on the aspects of R&D development through several presentations of research papers and key note addresses. The conference addresses the following topics: New Theoretical Computer Science, Systems and Software, Intelligent Systems.

We wish the dignitaries, participants and specially the organizing team in making this event a grand success.

About the Editors

Dr. Aswatha Kumar M. is working as Professor & Head of the Department of Information Science & Engineering since 2007. Before joining MSRIT, he has worked as Professor & Head of the Department of Computer Science & Engineering in JNN College of Engineering, Shimoga, Karnataka. He has also served two decades in Government of Karnataka. His research interests are in the areas of Image Processing. He has published more than 40 papers in Journals and Conferences. His research papers has been awarded as Best Papers in National as well as International Conferences. He has established new centres in association with various industries. He is also a board member of studies/examination boards of various universities. At present his research focuses are in the areas of document processing and medical imaging.

Dr. R. Selvarani is currently working as the Head of Department of Computer science and engineering in M.S. Ramaiah Institute of Technology, Bangalore. Earlier to this, she was the Dean Research in Dayananda Sagar Institutions, Bangalore. She has a dedicated teaching and research carrier spanning over 20 years in various universities including the position of Head of Department of Computer Science and Engineering for a period of 9 years. Two times she has been awarded the best teacher award from different Institutions. Her research interest spreads over the modern approaches in the area of Software Design and Architecture, CBSD, software quality estimation techniques, Analytical Software Engineering, Metrics and Measurements, software design quality estimation and software industry process and computational models for real time systems, service oriented architecture and cloud computing technology, Distributed networks and computer networks. She has more than 40 research publications in peer reviewed international journals and conferences. Her research interest has lead to 2 patents on technology innovation. She has published a book in Computer science and Engineering and currently working on 2 research handbooks in multidisciplinary domain. She is the Managing Editor of 2 international journals published from the research centre. She is the chairman/member of different board of studies/examinations of various universities.

Dr. T.V. Suresh Kumar is the Professor and Head of the Department of Master of Computer Applications in M S Ramaiah Institute of Technlogy, Bangalore. He received his Ph.D in Applied Mathematics from S K University, Anantapur, India. He has 20 years of experience in academics and 15 years of experience in research. He is supervising 5 Ph.D theses under Visvesvaraya Technological University, Belgaum and he has supervised M.S, M.Tech and over 10 M.Phil theses in Computer Science. He has published over 55 research papers and authored books on Java, J2EE and Data Mining. He has carried out funded projects from CASSA, DRDL, DRDO and UGC and consultancy projects from 7H Group of Companies and M S Ramaiah Dental College and Hospital. He is a visiting faculty for several renowned industries such as Intel, Honey Well, SAP Labs, Wipro Technologies, Jataayu Soft, Mphasis, Integra Micro Systems, HCL Technologies, Blue Star Infotech, L&T, Nokia, DRDO-CAIR, DRDO-CASSA, Indian Institute of Science (Proficience) and Various Universities/Academic Institutions. He is a member of Board of Studies and Board of Exams in various institutions. He is a life member of ISTE and Secretary for IEEE SMC Society Bangalore Chapter. His research and teaching interests include Object Technology, Software Engineering, Software Performance Engineering, System Simulation and Reliability Engineering.

Contents

New Theoretical Computer Science

Multilevel Feedback Queue Scheduling Technique for Grid Computing Environments 1
Dharamendra Chouhan, S.M. Dilip Kumar, B.P. Vijaya Kumar

Location and Detection of a Text in a Video 9
T.N.R. Kumar, S.K. Srivatsa, S. Murali

Hidden Markov Model with Computational Intelligence for Dynamic Clustering in Wireless Sensor Networks 19
Veena K.N., Vijaya Kumar B.P.

Assessment of Workload Using Shapely Value in Distributed Database 31
S. Jagannatha, T.V. Suresh Kumar, D.E. Geetha, K. Rajani Kanth

Modeling and Estimation of Cooperative Index for Multi-Agent Systems Using Execution Graph 41
S. Ajitha, T.V. Suresh Kumar, D. Evangelin Geetha, K. Rajani Kanth

QoS Multicast Routing Using Teaching Learning Based Optimization 49
Anima Naik, K. Parvathi, Suresh Chandra Satapathy, Ramanuja Nayak, B.S. Panda

A New Privacy Preserving Measure: *p*-Sensitive, t-Closeness 57
Sowmyarani C.N., G.N. Srinivasan, Sukanya K.

Indic Language Translation in CLIR Using Virtual Keyboard 63
Mallamma V. Reddy, M. Hanumanthappa

Energy Efficient Clustering and Grid Based Routing in Wireless Sensor Networks 69
Amrutha K.M., Ashwini P., Divyashree K. Raj, Kavitha Rani G., Monica R. Mundada

A Comparative User-Centric Study of Digital Library Software Systems ... 75
Samaneh Ahmadi, Shiva Shirdavani, Srini Ramaswamy

Adaptive Hexa-Diamond Search (AHDS) Algorithm for Fast Block Matching Motion Estimation ... 85
M.K. Pushpa, S. Sethu Selvi

A Probabilistic Solution to Rendezvous Problem ... 95
Shivam Agarwal, Arup Kumar Pal, Vihang Gosavi, Hemant Gangolia

Satellite Image Feature Extraction Using Neural Network Technique ... 101
T. Karthikeya Sharma, Sarvesh Babu N.S., Y.N. Mamatha

Identifying Refactoring Opportunity in an Application: A Metric Based Approach ... 107
Syamala Kumari Dora, Debananda Kanhar

Technologies for Cost Efficient Enterprise Resource Planning: A Theoretical Perspective ... 113
Shivani Goel, Ravi Kiran, Deepak Garg

Test Case Generation Using Activity Diagram and Sequence Diagram ... 121
Abinash Tripathy, Anirban Mitra

User Authentication Using Keystroke Recognition ... 131
Urvashi Garg, Yogesh Kumar Meena

On Decidability and Matching Issues for Regex Languages ... 137
Praveen Alevoor, Pratik Sarda, Kalpesh Kapoor

Randomized Algorithms: On the Improvement of Searching Techniques Using Probabilistic Linear Linked Skip Lists ... 147
Yogeesha C.B., Ramachandra V. Pujeri, Veena R.S.

Review of Proposed Architectures for Automated Text Summarization ... 155
Tejas Yedke, Vishal Jain, R.S. Prasad

Steer-By-Wire Implementation Using Kinect ... 163
Rohan Sadale, Roshan Kolhe, Sachin Wathore, Jagannath Aghav, Saket Warade, Sandeep Udayagiri

An Efficient Incentive Compatible Mechanism to Motivate Wikipedia Contributors ... 171
Mane Pramod, Sajal Mukhopadhyay, D. Gosh

Simulating Spiking Neuron for Information Theoretic Analysis in Stochastic Neuronal System ... 183
Sanjeev Kumar

Nash Equilibrium and Marcov Chains to Enhance Game Theoretic Approach for Vanet Security 191
Prabhakar M., J.N. Singh, G. Mahadevan

Fast Computation of Image Scaling Algorithms Using Frequency Domain Approach 201
Prasantha H.S., Shashidhara H.L., K.N.B. Murthy

Word Level Script Identification of Text in Low Resolution Images of Display Boards Using Wavelet Features 209
S.A. Angadi, M.M. Kodabagi

Analytical Study Using Data Mining for Periodical Medical Examination of Employees 221
Kiran Waghmare, Anusha R. Pai

Syntactic Representation of Shape of Object Using Regular Grammar 229
Saket Jalan, Pinaki Roy Chowdhury, K.K. Shukla

Message Overhead Analysis of Quorum Protocols 237
Parul Pandey, Maheshwari Tripathi

Modified (Q, r) Policy for Stochastic Inventory Control Systems in Supply Chain 247
R. Bakthavachalam, S. Navaneethakrishnan, C. Elango

Single Input Variable Universe Fuzzy Controller with Contraction-Expansion Factor for Double Inverted Pendulum 257
Yogesh Kr. Dhanni, M.J. Nigam

Performance Evaluation of A Novel Most Recently Used Frequency Count (MRUFC) List Accessing Algorithm 267
Rakesh Mohanty, Ashirbad Mishra

Automatic Test Case Generation Using Sequence Diagram 277
Vikas Panthi, Durga Prasad Mohapatra

Normalized Wavelet Hybrid Feature for Consonant Classification in Noisy Environments 285
T.M. Thasleema, N.K. Narayanan

Cuckoo Search for Inverse Problems and Topology Optimization 291
Xin-She Yang, Suash Deb

A Lock Management Framework for a Class Hierarchy Tree 297
Arvind Mohan, Gaurav Singhal, Bhaskar Biswas

Systems and Software

ECG Arrhythmia Classification Using R-Peak Based Segmentation, Binary Particle Swarm Optimization and Absolute Euclidean Classifier 303
Milan S. Shet, Minal Patel, Aakarsh Rao, Chethana Kantharaj, Suma K.V.

Design of Low Power High Speed 4-Bit TIQ Based CMOS Flash ADC 319
Parvaiz Ahmad Bhat, Roohie Naaz Mir

A Reduced Complexity LDPC Decoding Algorithm Using Dynamic Bit Node Selection 329
Suvarna Hudgi, Siddram R. Patil

Memory Optimized Design of Reciprocal Unit 339
Mahmad M. Nadaf, R.M. Banakar, Saroja V. Siddamal

Enhanced LZW Algorithm with Less Compression Ratio 347
Amit Setia, Priyanka Ahlawat

Design of High Security and Performance System for Storage Devices Using AES 353
Vinodkumar I. Bellikatti, Chetan S., Shivaputra, Kushal K.S.

Implementation of Lifting Scheme Based DWT Architecture on FPGA 361
Naagesh S. Bhat

Mobile Based E-Mail Reading System 371
Azath M., Channamallikarjuna Mattihalli, Member IEEE

Design of ANFIS Controller Based on Fusion Function for Linear Inverted Pendulum 379
Abhishek Kumar, R. Mitra

Proposing Modified NSGA-II to Solve a Job Sequencing Problem 387
Susmita Bandyopadhyay, Arnab Das

Verification Platform for FPGA Based Architecture 393
Adesh Panwar

Implementation and Analysis of Downlink Scheduling for IEEE 802.16 Using Controlled Priority Queuing 399
Z.M. Patel, U.D. Dalal

Mechanism for Secure Content Publishing for Reporting Platform Hosted on Public Could Infrastructure 407
Bhanu Prakash Gopularam, Nalini N.

A Semi-Interquartile Min-Min Max-Min (SIM^2) Approach for Grid Task Scheduling 415
Sanjaya Kumar Panda, Sourav Kumar Bhoi, Pabitra Mohan Khilar

Simulation Based Performance Comparison of Reactive Routing Protocols in Mobile Ad-Hoc Network Using NS-2 423
G. Jose Moses, D. Sunil Kumar, P. Suresh Varma, N. Supriya

Fault Tolerance for Large Scale Storage Systems 429
Pradeep K.R., George Philip C.

Real Time Electro-Oculogram Driven Rehabilitation Aid 435
Anwesha Banerjee, Pratyusha Das, Shounak Datta, Amit Konar, R. Janarthanan, D.N. Tibarewala

A Quantitative Approach Using Goal-Oriented Requirements Engineering Methodology and Analytic Hierarchy Process in Selecting the Best Alternative 441
Vinay S., Shridhar Aithal, Sudhakara G.

Multilanguage Based SMS Encryption Techniques 455
M. Rajendiran, B. Syed Ibrahim, R. Pratheesh, C. Nelson Kennnedy Babu

High Speed Low Power VLSI Architecture for SPST Adder Using Modified Carry Look Ahead Adder 461
Narayan V.S., Pratima S.M., Saroja V.S., R.M. Banakar

ASIC Primitive Cells in Modified Gated Diffusion Input Technique 467
R. Uma, P. Dhavachelvan

Latent Dirichlet Allocation Model for Recognizing Emotion from Music . . . 475
S. Arulheethayadharthani, Rajeswari Sridhar

Investigations on the Routing Protocols for Wireless Body Area Networks 483
Jayanthi K. Murthy, Thimmappa P., V. Sambasiva Rao

Effect of Idle Mode on Power Saving in Mobile WiMAX Network 491
Thontadharya H.J., Shwetha D., Subramanya Bhat M., Devaraju J.T.

High Speed Programmable Digital Telemetry Filter for Flight Test 501
Navitha M.V., M.Z. Kurian, G. Koteswara Rao, Umashankar B.

Hierarchical Storage Technique for Maintaining Hop-Count to Prevent DDoS Attack in Cloud Computing 511
Vikas Chouhan, Sateesh Kumar Peddoju

VHDL Synthesis and Simulation of an Efficient Genetic Algorithm Based on FPGA 519
N. Rajeswaran, T. Madhu, M. Suryakalavathi

Forensic Sketch Matching Using SURF 527
Dileep Kumar Kotha, Santanu Rath

Comparison of Configurations of Data Path Architecture Developed Using Template 539
B. Bala Tripura Sundari, Varsha Krishnan

LDPC and SHA Based Iris Recognition for Smart Card Security 549
K. Seetharaman, R. Ragupathy

A Broker Based Architecture for Adaptive Load Balancing and Elastic Resource Provisioning and Deprovisioning in Multi-tenant Based Cloud Environments 561
Thamarai Selvi Somasundaram, Kannan Govindarajan, M.R. Rajagopalan, S. Madhusudhana Rao

Variation in Active Site Amino Residues of H1N1 Swine Flu Neuraminidase 575
G. Nageswara Rao, P. Srinivasarao, A. Apparao, T.K. Rama Krishna Rao

A Semantic Search Engine to Discover and Select Sensor Web Services for Wireless Sensor Network 585
Chinmohan Nayak, Manoranjan Parhi

Prevention of Man in the Middle Attack by Using Honeypot 593
Mayank Tiwari, Tushar Sharma, Pankaj Sharma, Shaivya Jindal, Priyanshu

Slicing of Programs Dynamically under Distributed Environment 601
Santosh Pani, Shashank Mouli Satapathy, G.B. Mund

An Efficient Incentive Compatible Mechanism for Paid Crowdsourcing 611
Shalini Gupta, Sajal Mukhopadhyay, D. Gosh

Doubling Runtime Estimations to Improve Performance of Backfill Algorithms in Cloud Metaschedular Considering Job Dependencies 621
Ankur Jindal, P. Sateesh Kumar

Self-managing the Performance of Distributed Computing Systems – An Expert Control Solution 629
Ravi Kumar G., C. Muthusamy, A. Vinaya Babu, Raj N. Marndi

An Embedded Navigation System for Aiding People with Alzheimer's Disease 639
Siddalingesh Navalgund, Jayashree Taralabenchi, Kavana Hegde, Soumya Hegde

Software Licensing Models and Benefits in Cloud Environment: A Survey 645
Mohan Murthy M.K., Mohd Noorul Ameen, Sanjay H.A., Patel Mohammed Yasser

VLSI Architecture of Spread Sprectrum Image Watermarking Decoder 651
Navonil Chatterjee, Moudud Sohid, Sudipta Chakraborty

Strategy Driven Approach for the AD HOC Network Participants Using the Notion of Trust and Activity 659
Shabana Sultana, C. Vidya Raj

Multicriteria Decision Analysis for Intrusion Detection Data 667
Sanjiban Sekhar Roy, Omsai Jadhav, Saptarshi Chakraborty, Swapnil Saurav, Madhu Viswanatham

Realization of the Cryptographic Processes in Privacy Preserving 673
Sumana M., Hareesh K.S.

A Privacy Preserved Integrated Framework for Location Based Tracking for Wireless Sensor Networks 679
Vinoth Kumar S., Suresh R.M., Govardhan A.

Cluster Allocation Strategies of the ExFAT and FAT File Systems: A Comparative Study in Embedded Storage Systems 691
Keshava Munegowda, G.T. Raju, Veera Manikandan Raju

Audio Steganography Used for Secure Data Transmission 699
Pooja P. Balgurgi, Sonal K. Jagtap

An XML Parser of Efficient Updates for a Binary String: A Case Study 707
J. Bhagyashala, S. Shefali

SVM-DSD: SVM Based Diagnostic System for the Detection of Pomegranate Leaf Diseases 715
Sanjeev S. Sannakki, Vijay S. Rajpurohit, V.B. Nargund

Encrypted Traffic and IPsec Challenges for Intrusion Detection System 721
Manish Kumar, M. Hanumanthappa, T.V. Suresh Kumar

Intelligent Systems

An Effective User Interface Tool for Retrieval of Heart Sound and Murmurs 729
Kiran Kumari Patil, B.S. Nagabhushana, Vijaya Kumar B.P.

DCell-IP: DCell Emboldened with IP Address Hierarchy for Efficient Routing 739
A.R. Ashok Kumar, S.V. Rao, Diganta Goswami, Ganesh Sahukari

An Approach to Securing Data in Hosted CRM Applications 747
Siddharth M. Pandya, Abhishek Srikumar, Chandrika T.

Alzheimer's Disease Detection Using Minimal Morphometric Features with an Extreme Learning Machine Classifier 753
M. Aswatha Kumar, B.S. Mahanand

Bidding Strategy in Simultaneous English Auctions Using Game Theory . . . 763
Nirupama Pavanje

Web Personalization Based on Short Term Navigational Behaviour and Meta Keywords 773
Siddu P. Algur, Nitin P. Jadhav, N.H. Ayachit

A Parallel Fuzzy C Means Algorithm for Brain Tumor Segmentation on Multiple MRI Images 787
Aarthi Ravi, Ananya Suvarna, Andrea D'Souza, G. Ram Mohana Reddy, Megha

Implementation of Web Search Result Clustering System 795
Hanumanthappa M., B.R. Prakash

Data Mining in Online Social Games 801
Nazneen Ansari, Maahi Talreja, Vaishali Desai

Data Mapping in Intelligent Form Using Random Hierarchical Bit Format Enhancing the Security in Data Retrieval 807
Rahul Gupta, Nidhi Garg, Preetham Kumar

Effective Unit Testing Framework for Automation of Windows Applications 813
A.N. Seshu Kumar, S. Vasavi

A New Optimization Method Based on Adaptive Social Behavior: ASBO . . . 823
Manoj Kumar Singh

Design and Implementation of Interval Type-2 Single Input Fuzzy Logic Controller for Magnetic Levitation System 833
Anupam Kumar, Manoj Kumar Panda, Vijay Kumar

Mining Negative Association Rules from Multiple Data Sources on the Basis of Local Pattern Analysis 841
T. Ramkumar, S. Selvamuthukumaran, S. Hariharan, V. Harikrishnan

Recognition of Hand Punched Kannada Braille Characters Using Knowledge Based Multi Decision Concept: Basic Symbols 847
Srinath S., C.N. Ravi Kumar

Vascular Tree Segmentation in Fundus Images Using Curvelet Transform 859
Rupu Kumari, Charul Bhatnagar, Anand Singh Jalal

A Density-Based Clustering Paradigm to Detect Faults in Wireless Sensor Network 865
Sourav Kumar Bhoi, Sanjaya Kumar Panda, Pabitra Mohan Khilar

Ant Colony Optimization for Data Cache Technique in MANET 873
R. Baskaran, P. Victer Paul, P. Dhavachelvan

Automatic Extraction of Kannada Complex Predicates from Corpora 879
S. Parameswarappa, V.N. Narayana

Texture Image Retrieval Using Greedy Method 885
Pushpa B. Patil, Manesh B. Kokare

Multi-lingual Speaker Identification with the Constraint of Limited Data . . . 893
B.G. Nagaraja, H.S. Jayanna

Improving Performance of K-Means Clustering by Initializing Cluster Centers Using Genetic Algorithm and Entropy Based Fuzzy Clustering for Categorization of Diabetic Patients 899
Asha Gowda Karegowda, Vidya T., Shama, M.A. Jayaram, A.S. Manjunath

Hindi and English Off-line Signature Identification and Verification 905
Srikanta Pal, Umapada Pal, Michael Blumenstein

A Robust Method of Image Based Coin Recognition 911
B.V. Chetan, P.A. Vijaya

N-Gram Based Approach to Automatic Tamil Lyric Generation by Identifying Emotion 919
Rajeswari Sridhar, Jalin Gladis D., Ganga K., Dhivya Prabha G.

Texture Analysis and Defect Classification for Fabric Images Using Regular Bands and Quadratic Programming 927
R. Obula Konda Reddy, B. Eswara Reddy, E. Keshava Reddy

Exploring the Pattern of Customer Purchase with Web Usage Mining 935
Paresh Tanna, Yogesh Ghodasara

Information Fusion from Mammogram and Ultrasound Images for Better Classification of Breast Mass 943
Minavathi, Murali S., M.S. Dinesh

Design of Fuzzy PD Controller for Inverted Pendulum in Real Time 955
Nidhi Patel, M.J. Nigam

Classification of Kannada Numerals Using Multi-layer Neural Network 963
Ravindra S. Hegadi

Content Based Image Retrieval by Combining Median Filtering, BEMD and Color Technique 969
Purohit Shrinivasacharya, M.V. Sudhamani

Fuzzy Geometric Face Model for Face Detection Based on Skin Color Fusion Model 977
P.S. Hiremath, Manjunath Hiremath

A Novel Approach for Prefetching of Web Pages through Clustering of Web Users to Reduce the Web Latency 983
G.T. Raju, M.V. Sudhamani

A Neuro-Fuzzy Based Intelligent Agent for Text Based Emotion Recognition 991
G. Sharada, O.B.V. Ramanaiah

Feature Selection for Decoding of Cognitive States in Multiple-Subject Functional Magnetic Resonance Imaging Data 997
Accamma I.V., H.N. Suma

A New Approach to Partial Image Encryption 1005
Parameshachari B.D., K.M.S. Soyjaudah

Fuzzy Number with Nonlinear Membership Functions to Provide Flexibility in a Multi Objective Travelling Salesman Problem 1011
Atul Kumar Tiwari, Cherian Samuel, Vinay Pratap Singh, Vivek Saraswati

A Novel Approach for Image Retrieval Based on ROI and Multifeatures Using Genetic Algorithm 1021
K.S.Md. Musa Mohinuddin, P. Subbaiah, S. Tipu Rahaman

Ship Detection from SAR and SO Images 1027
Y. Sreedevi, B. Eswar Reddy

Automatic Speaker Recognition System 1037
P.M. Ghate, Shraddha Chadha, Aparna Sundar, Ankita Kambale

An Approach for Document Image Based Printed Character Recognition 1045
Sushila Aghav, Shilpa Paygude

Flexibility in Supplier Selection Using Fuzzy Numbers with Nonlinear Membership Functions 1051
Atul Kumar Tiwari, Cherian Samuel, Anunay Tiwari

Fuzzy Based Interference Reduction in Cognitive Networks 1061
Lavanya G., Pandeeswari S., Shanmugapriya R.K., Umamaheswari A.

Managing Traffic Flow Based on Predictive Data Analysis 1069
Dhara J. Patel, Snoeji Varghese John, Fibinse Kaliangra

Digital Filter Approach for ECG in Signal Processing 1075
Sonal K. Jagtap, M.D. Uplane

A Pattern Recognition Approach of Japanese Text Recognition for Template Matching 1083
Soumendu Das, Sreeparna Banerjee

Extraction of Bacterial Clusters from Digital Microscopic Images through Statistical and Neural Network Approaches 1091
Chayadevi M.L., Raju G.T.

Analysis of Brain Activity for Motor Task Using Simultaneous EEG - fMRI 1101
Sandhya M., Rose Dawn, Rajanikant Panda

K-Means Clustering Microaggregation for Statistical Disclosure Control . . . 1109
Md. Enamul Kabir, Abdun Naser Mahmood, Abdul K. Mustafa

Content Based Image Retrieval Using Sketches 1117
M. Narayana, Subhash Kulkarni

Component Based Software Development Using Component Oriented Programming 1125
Ruchi Shukla, T. Marwala

Transformation of Artistic Form Text to Linear Form Text for OCR Systems 1135
Vishwanath C. Kagawade, Vijayashree C.S., Vasudev T.

A Fuzzy Sectional Real-Time Scheduling Algorithm Based on System Load 1145
Annappa B.

An Intelligent and Robust Single Input Interval Type-2 Fuzzy Logic Controller for Ball and Beam System 1155
Sumanta Kundu, M.J. Nigam

Adaptive Neuro Fuzzy Inference Structure Controller for Rotary Inverted Pendulum 1163
Rahul Agrawal, R. Mitra

Erratum

A Fuzzy Sectional Real-Time Scheduling Algorithm Based on System Load E1
Annappa B.

Author Index 1171

Multilevel Feedback Queue Scheduling Technique for Grid Computing Environments

Dharamendra Chouhan[1], S.M. Dilip Kumar[1], and B.P. Vijaya Kumar[2]

[1] Dept. of Computer Science and Engineering,
University Visvesvaraya College of Engineering,
Bangalore, India
[2] Dept. of Computer Science and Engineering,
M.S. Ramaiah Institute of Technology,
Bangalore, India

Abstract. Effective and efficient job scheduling is an important aspect of Grid computing. Task scheduling becomes more complicated in a Grid environment, due to geographically distribution, heterogeneity and dynamic nature of grid resources. In this paper, a new computational scheduling policy called Multilevel Feedback Queue (MLFQ) scheduling, which is designed to support the allocation of resources for gridlets (jobs) is proposed. Gridlets provided by the users are assigned to processing elements (PEs), and gridlets whose remaining service time is shifted between queues of the MLFQ scheduler to get completed. In MLFQ, the total architecture is divided into multiple prioritized queues. This approach provides gridlets which starve in the lower priority queue for long time to get resources. As a result, the response time of the starved gridlets decreases and overall turnaround time of the scheduling process decreases. This scheduling policy is simulated using Alea GridSim toolkit to test the performance.

Keywords: Grid computing, Job Scheduling, Multilevel feedback queue, GridSim.

1 Introduction

Grid computing is a distributed computing which has emerged for solving a large scale intensive data through sharing of resources over the network [1]. In grid computing systems, there are often large amounts of resources available to be used for computing jobs. Scheduling in a grid computing system is not as simple as scheduling on a multi-processor machine because of several factors. These factors include the fact that grid resources are sometimes used by paying customers who have interest in how their jobs are being scheduled [2]. However, grid computing systems usually operate in remote locations so scheduling tasks for the clusters may be occurring over a network [3]. Job scheduling algorithms are commonly applied to grid resources to optimally post jobs to grid resources [4, 5]. Usually, grid users submit their jobs to the grid manager to utilize and fulfill the facilities provided by grid. The grid manager distributes the submitted jobs among the grid resources to minimize the total response time.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1–7.
springerlink.com

In a Grid environment, there is moderately large number of job scheduling algorithms proposed to minimize the total completion time of the jobs [6, 7]. These algorithms works on minimizing the overall completion time of the jobs by analyzing the suitable resources to be assigned to the jobs. In contrast with minimizing the overall completion time of the jobs does not necessarily result in the minimization of execution time of each individual task. In this paper, we propose a new scheduling policy for grid computing which uses multilevel feedback queue technique concept to avoid the starvation of low priority jobs for a longer duration to get resources to complete their requested services. In this technique, jobs are scheduled according to their remaining service time and they are shifted down from queue to queue as they have some remaining service time. Every queue has unique time quanta that gradually increase from top level to bottom level queues so that longer jobs gradually moves from top to bottom level queues for getting completed. All low priority jobs will process on intermediate queues and gets completed with minimal duration, so that all jobs will get an equal opportunity to utilize grid resources efficiently. The rest of the paper is organized as follows. Section 2 presents the related works. In Section 3 the system model for scheduling in Grid computing environment is presented. In section 4, the MLFQ scheduling technique is proposed. The simulation of the MLFQ scheduling algorithm using Alea GridSim is presented in section 5. Finally, section 6 concludes the paper.

2 Related Work

There has been significant research continuing to attempt to devise scheduling algorithms for grid environments' problem of efficient job assignment. Some of the jobs scheduling algorithms in a grid environment are given below.

X. He et al. [9] have proposed an algorithm based on the conventional min-min algorithm known as QoS guided min-min which schedules the jobs requiring high bandwidth before others. F. Dong et al. [12] have proposed an algorithm called QoS priority grouping scheduling. This algorithm, considers completion time and acceptation rate of the jobs and the makespan of the entire system as key factors for job scheduling. E. Ullah Munir et al. [13] have proposed a new job scheduling algorithm which makes use of grid computing environments known as QoS Sufferage. K. Etminani et al. [14] have proposed an algorithm which provides a solution on basis of max-min and min-min algorithms. The algorithm discovers the situations where to adopt one of these two algorithms, based on the standard deviation of the estimated completion times of the jobs on every computing resources. L. Mohammad Khanli et al. [10, 11] have proposed a QoS based scheduling algorithm for an architecture called Grid-JQA. In this method the solution involves applying an aggregation formula which includes a combination of different parameters together with weighting factors to perform operations on QoS.

3 System Model

In this section, we present the system model of scheduling in Grid computing environment. An open queuing network model of Grid resources is considered as shown in Figure 1. There are input queues to store user gridlets waiting to be processed by one of the processing elements present in grid resource. They are connected with a high-speed network with negligible communication delay. The processing elements speed is measured in terms of million instructions per second (MIPS) rating. The processing elements can be either homogeneous or heterogeneous. For homogeneous PEs, the MIPS rating is same where as for heterogeneous it is different from each other. Each Grid resource system consists of machines having a set of PEs and each PE is having an independent multilevel feedback scheduling policy. The system comprises three levels of queues, namely, the first level with fixed time quanta which is double in the level two queues and a FIFO in the third level.

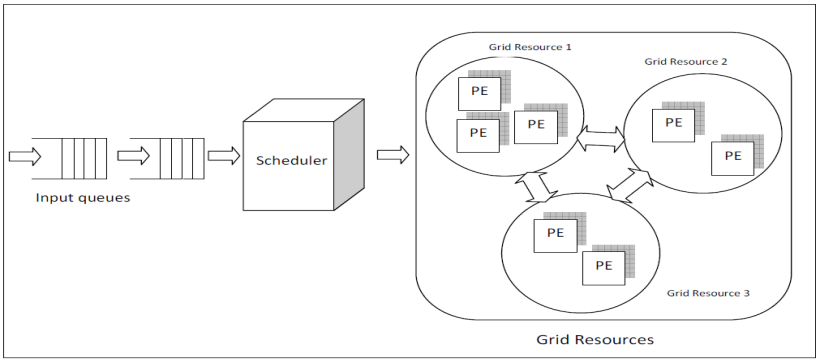


Fig. 1. A queuing network model for the GridSim scheduling system

The proposed model works under the following assumptions:

1. Gridlets arriving into the system are independent of one another.
2. When gridlets are mapped to the machines, based on their requirement, it checks for the (resource) availability list.
3. No information is available on the workload of incoming gridlets.
4. The initial processing speed of each PE is provided and processing capacity of Grid resources is updated from time to time based on last gridlet executed and time taken for task completion.

3.1 Multilevel Feedback Queue (MLFQ)

Multilevel feedback queue plays a significant role in multilevel queue scheduling. In MLFQ, jobs are scheduled according to their remaining CPU burst and they are shifted down from queue to queue as they have some remaining CPU burst. Every queue has unique time slice that gradually increases from upper level queue to lower level queue. So the CPU intensive jobs go down from upper queues to lower queues

gradually for getting completed. Thus, lower priority queues are filled with CPU intensive jobs and as a result these processes start to starve for getting CPU attention. The MLFQ scheduling organizes the queues to minimize the queuing delay and optimize the queuing environment efficiency [8].

3.2 State Diagram

The system is modeled in a state transition diagram as shown in Figure 2. As gridlets arrives to the input queue, each gridlet is selected and it acquires the requested resources from grid resource list. Once it acquires the requested resources, it finds the suitability of the resources and checks for the required PEs, MIPS, bandwidth and storage. If the suitability is fulfilled, the scheduler assigns gridlets to the resources selected from the resource list. Gridlets are scheduled according to their remaining service time and they are shifted down from queue to queue as they have some remaining service time. Every queue has unique time slice that gradually increases from upper level queue to lower level queue. So the PEs intensive gridlets go down from upper queues to lower level queues gradually for getting executed. If the gridlet fails to execute at this stage then it is placed back into input queue during the course of execution for later resumption.

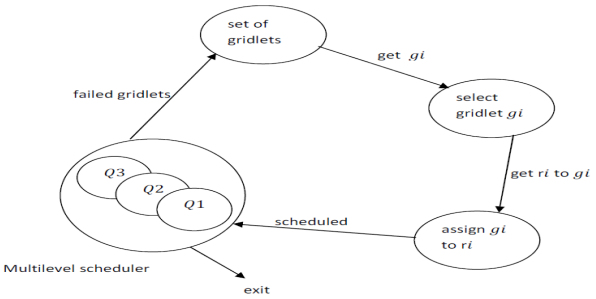


Fig. 2. State Diagram

4 Proposed Solution

In this section, we briefly explain the proposed solution for scheduling the jobs using MLFQ technique in Grid environment. The user submits gridlets along with the requirements to the Alea GridSim scheduling system. The submission of gridlets to the resources involves checking the suitability of the available PEs. If the requirement is satisfied, the gridlets are assigned to the respective resources. This technique uses a dynamic priority mechanism to schedule the gridlets to the system efficiently and maximize the resource utilization. The MLFQ scheduling model is depicted in the Figure 3. The gridlet waiting for the service is placed in the waiting queue. The gridlets that are scheduled in the queue Q_1 are executed. If the gridlets in Q_1 submitted for execution do not complete in the given time quanta of Q_1 then those gridlets are pushed onto the next level queue Q_2. Then the gridlets pushed on to Q_2 are executed

along with the gridlets present in queue Q_1. Similarly, if the gridlets in Q_2 submitted for execution do not complete in the fixed time quanta of Q_2 then those gridlets are pushed onto the next level queue Q_3. However, the gridlets present in Q_3 are executed based on FCFS scheduling policy. The shorter gridlets completes its execution quickly, without migrating to lower level queues. All gridlets gets an opportunity to execute and thus reduces starvation of gridlets.

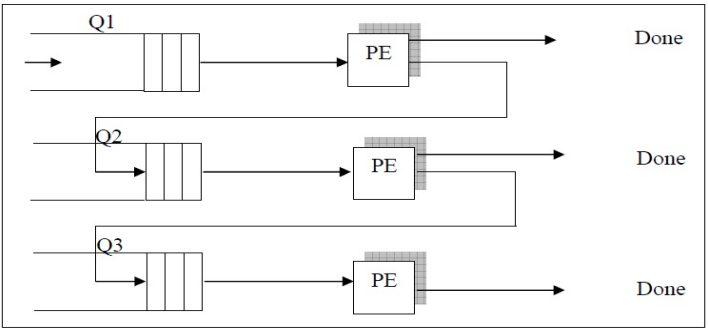


Fig. 3. Multilevel Feedback Queue (MLFQ) Scheduling model

5 Simulation

In this section we show the performance of MLFQ scheduling technique trough several experiments using Alea simulator, an extension of GridSim simulation toolkit. The experiment involved 5000 jobs that were executed on 14 clusters having 806 CPUs. We run the simulation by providing input data set and it completes all the jobs submitted to the grid over a span of time. MLFQ is able to increase the machine usage by using MLFQ approach and average machine usage per day as depicted in Figure 4. The number of waiting and running jobs on an average against each day is depicted in Figure 5. MLFQ is capable of a higher resource utilization and reduction of the

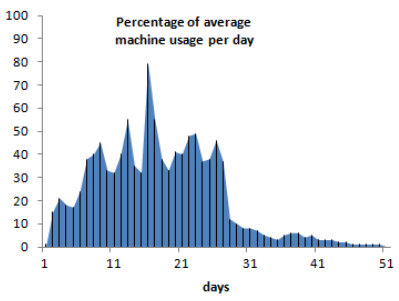


Fig. 4. Percentage of average machine usage per day

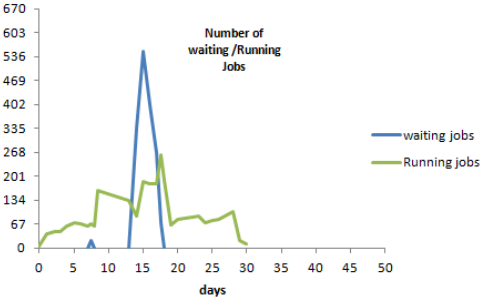


Fig. 5. Average job execution per day

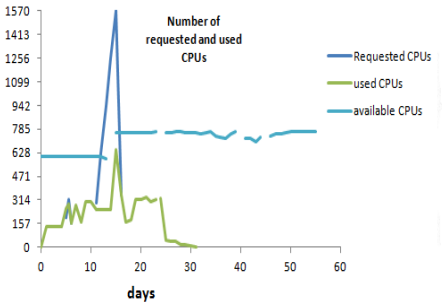


Fig. 6. CPU usage per day

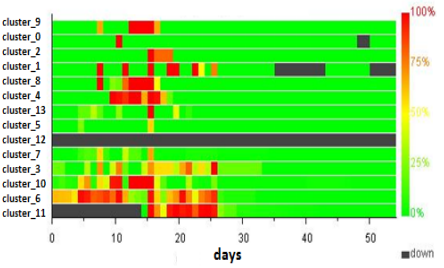


Fig. 7. Percentage of cluster usage

number of waiting jobs through the time and the number of waiting and running jobs per day. The requested and available CPU usage per day is shown in Figure 6. Figure 7 presents the average machine usage per cluster. Simulation results show that there is a minimization of overall response time and waiting time of gridlets.

6 Conclusion

The paper describes a new approach to schedule tasks efficiently in a grid environment. We proposed a Multilevel Feedback Queue Scheduling (MLFQ) for Alea, a GridSim based simulator. The approach is based on processing capability of individual grid resources. Our policy provides a solution by implementing MLFQ scheduler where lower priority gridlets will complete quickly, without migrating to the lower levels of the hierarchy. So that we can achieve high times. The transportation cost and overall communication delay is considered for future work.

References

[1] Foster, I., Kesselman, C.: The Grid 2: Blueprint for a New Computing Infrastructure, IInd edn. Elsevier and Morgan Kaufmann Press (2004)
[2] Hoschek, W., Jaen-Martinez, J., Samar, A., Stockinger, H., Stockinger, K.: Data Management in an International Data Grid Project. In: Buyya, R., Baker, M. (eds.) GRID 2000. LNCS, vol. 1971, pp. 77–90. Springer, Heidelberg (2000)
[3] Buyya, R., Steve Chapin, S., DiNucci, D.: Architectural Models for Resource Management in the Grid. In: IEEE/ACM International Workshop on Grid Computing (2000)
[4] Mohammad Khanli, L., Analoui, M.: Resource Scheduling in Desktop Grid by Grid-JQA. In: The IEEE 3rd International Conference on Grid and Pervasive Computing (2008)
[5] Mohammad Khanli, L., Analoui, M.: Grid_JQA: A QoS Guided Scheduling Algorithm for Grid Computing. In: The 6th IEEE International Symp. on Parallel and Distributed Computing (2007)
[6] Dong, F., et al.: A Grid Task Scheduling Algorithm Based on QoS Priority Grouping. In: Proc. of the 5th IEEE International Conf. on Grid and Cooperative Computing (2006)
[7] Etminani, K., Naghibzadeh, M.: A Min-min Max-min Selective Algorithm for Grid Task Scheduling. In: The 3rd IEEE/IFIP International Conf. on Internet, Uzbekistan (2007)

[8] Hoganson, K.: In: Reducing MLFQ Scheduling Starvation with Feedback and Exponential Averaging Consortium for Computing Sciences in Colleges, Southeastern Conference, Georgia (2009)
[9] He, X., Sun, X.-H., Laszewski, G.V.: QoS Guided Min-min Heuristic for Grid Task Scheduling. J. Computer Science and Technology 18, 442–451 (2003)
[10] Mohammad Khanli, L., Analoui, M.: Resource Scheduling in Desktop Grid by Grid-JQA. In: The 3rd IEEE International Conf. on Grid and Pervasive Computing (2008)
[11] Mohammad Khanli, L., Analoui, M.: Grid_JQA: A QoS Guided Scheduling Algorithm for Grid Computing. In: The 6th IEEE International Symp. on Parallel and Distributed Computing (2007)
[12] Dong, F., Luo, J., et al.: A Grid Task Scheduling Algorithm Based on QoS Priority Grouping. In: Proc. of 5th IEEE International Conf. on Grid and Cooperative Computing (2006)
[13] Ullah Munir, E., Li, J., Shi, S.: QoS Sufferage Heuristic for Independent Task Scheduling. Grid J. Information Technology 6(8), 1166–1170 (2007)
[14] Etminani, K., Naghibzadeh, M.: A Min-min Max-min Selective Algorithm for Grid Task Scheduling. In: 3rd IEEE/IFIP International Conf. on Internet, Uzbekistan (2007)

Location and Detection of a Text in a Video

T.N.R. Kumar, S.K. Srivatsa, and S. Murali

Vels University
St. Joseph's College of Engg.,
Jeppiaar Nagar,
Chennai - 600 119
tnrkumar24785@yahoo.com

Abstract. Recognizing and identifying the objects present in a video is a challenging task in the area of computer vision. The paper presents a method through image processing activities, to recognize a vehicle passing through a highway. The video of the vehicle is recorded and a frame, which contains the backside or front side of the vehicle, is separated for further processing. This single frame is an image, might have the vehicle object any where in it. A prior knowledge of the structure of the vehicle is used to locate the vehicle in the frame. The number plates of the vehicles normally will have significant features like with the specific background (White or Yellow) and the vehicle numbers are written in a rectangular area with these features is adopted to locate the number plate in the frame. Further segmentation is carried at character level using contour traversal method. A fourteen segment projection method is followed to recognize the characters of the segmented characters. The proposed method is able to segment and recognize the vehicles in the most of the ideal situations.

Keywords: Video to image conversion, Edge detection, Contour traversal, Character segmentation.

1 Introduction

Interest in the potential of digital images has increased enormously over the last few years, fuelled at least in part by the rapid growth of computer imaging. Users in many professional fields are exploiting the opportunities offered to access and manipulate remotely sensed images in all kinds of new and exciting ways[1]. However they are also discovering that the process of locating a desired image in a large and varied ambiance can be a source of considerable frustration[1].

Detecting the position of the text in digital images is of paramount importance in document analysis[1][2]. There are some research attempt made towards recognizing text in videos[3][4]. Image processing algorithms often consider the identification of text characters as trivial problem since those images of text characters. However the scenario is complicates when text is part of image. Image segmentation is required

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 9–17.
springerlink.com

before the characters are being recognized. The presence and interpretation of text in these images can provide visual information in addition to possible text in the form of captions, subtitles or image objects. Input scenes are decomposed in to set of binary images where connected components are analyzed for the possible presence of text characters[3]. The process of locating text in a given image is the first step in the problem of text reading. The problem of text detection gets complex with the variations in fonts, sizes and textures. However in this case law enforcement makes the numbers written on the vehicle with normal font makes the recognition much easier.

Input is taken from a stationary camera, which continuously takes the video of the vehicles passing in front if it. To increase the performance of recognition, preprocessing activities like normalization, skew detection and correction and segmentation are performed. Quality of the video produced by the camera is not always consistent; hence it is required to preprocess the video.

The paper is organized as follows: In section 2 details about the preprocessing adopted for this application is presented. Section 3 discusses about contour traversal technique adopted to segment the rectangles present in the image. In section 4 an illustration is given about the details of segmenting the characters in a number plate. Experimental results and conclusions are presented in section 5 & 6 respectively.

2 Preprocessing

Generally all images contain some noise in the image. This noise should be removed before any further processing is done[9]. There are several standard algorithms available, which can be used for noise for removal and smoothening.

2.1 Gaussian Convolution

Gaussian convolution is typically used for image smoothing, in which large changes in intensities between adjacent pixels are diminished by weighted averaging. It uses a symmetric normalized 2-D Gaussian smoothing operator G (x, y) for its convolution kernel:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

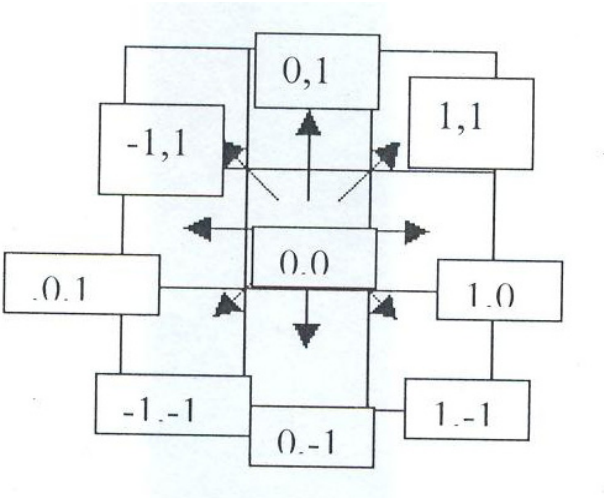


Fig. 1. Gaussian Convolution

In Gaussian convolution method, substitution of the values of x and y as given in figure 1 and logically placing on the image to produce smoothened image. Normally, the center value of the mask is dominating as shown in table 1. The mask shown in table 1 is derived from equation (1) for =0.4.

Table 1. The Guassian Mask for $\sigma = 0.4$ before normalization (Sum of the weights =1.7722)

0.00192026	0.0437049	0.00192026
0.0437049	0.994718	0.0437049
0.00192026	0.0437049	0.00192026

Table 2. The Guassian Mask for $\sigma = 0.4$ before normalization (Sum of the weights =1)

0.00163118	0.0371255	0.00163118
0.0371255	0.844973	0.0371255
0.00163118	0.0371255	0.00163118

Before performing any operation the picture is converted to binary as it reduces complexity and decision is only to know the area of text in the image.

Edge Detection

Edges of the objects give more details about the boundary and shape of the intermediate component. After scanning picture horizontally, if there is a change in then it is considered that point as edge point. Scanning is performed on same picture vertically and if there is any change in color, then it is considered that point as edge point. Figure 2a shows the image before edge detection and 2.b shows the edges of the image. Robert cross operator [6] is used to detect the edges

First form of Roberts Operator

$$\sqrt{[I(r,c)-I(r-1,c-1)]^2+[I(r,c-1)-I(r-1,c)]^2}$$

Second form of Roberts Operator

$$\nabla f \approx |z_5 - z_9| + |z_6 - z_8|$$

$$|I(r,c)-I(r-1,c-1)|+|I(r,c-1)-I(r-1,c)|$$

Where Vf is the gradient operator

$$h_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad h_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

$w_1(x,y)=$

1	0
0	-1

$w_2(x,y)=$

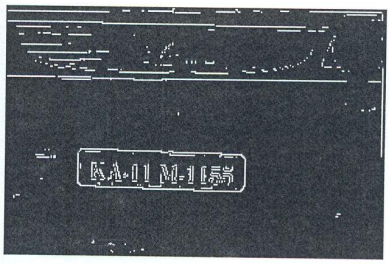
0	1
-1	0

Z1	Z2	Z3
Z4	Z5	Z6
Z7	Z8	Z9

3*3 Image region

Before edge detection

Fig. 2a.



After Edge detection

Fig. 2b.

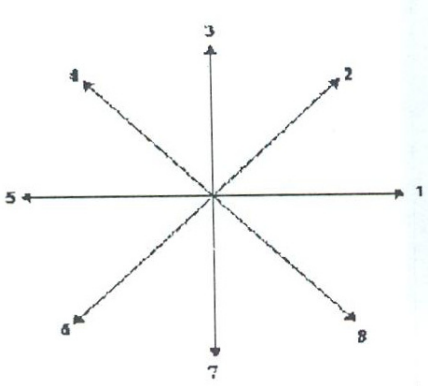
3 Segmentation of Number Plate

Number plates in any vehicle will have significant features, like with specific background and written on a rectangular portion. This knowledge enables the segmentation of number plates using contour traversal method, a rectangle which contains some characters. The method is illustrated below.

3.1 Contour Traversal

Contour traversal is widely used in image processing application. The chain codes are used to characterize a pixel on the contour. Chain code is a local feature, which gives the spatial association between two successive points. An arbitrary point is considered as starting point. Traversing along a contour of the object and reaching the starting point may produce a rectangle. Situation may arise where the contour traversal encounters two branches. In such situations the contour traversal is terminated.

This technique is used to recognize the closed loop present in the image. If there is more than one rectangle present in the image only bottom most rectangle where generally the number plate is present is only considered and the remaining rectangles present in the image is ignored.



P8	P1	P2
P7	P0	P3
P6	P5	P4

Fig. 3.

4 Character Segmentation and Recognition

The segmented number plate has two distinct intensities, the foreground that is generally black and the background that is generally white/yellow. Segmentation is initiated by scanning the image horizontally and vertically. When a black pixel is encountered, scanning is continued until a white pixel encountered and all the co-ordinates of the black pixel as well as the white pixel as stored in an array and check for continuity. If there is continuity in black pixel then it is a character and if there is continuity in white pixel then it is considered as gap between two characters and these characters are segmented separately.

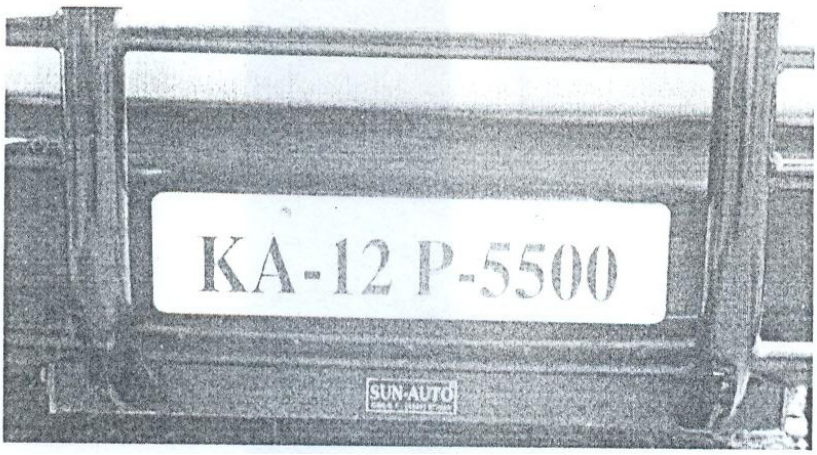


Fig. 4.

Fig. 5.

5 Character Using Projection Method

Segmented characters are taken as input to this stage of recognition[10]. A logical box with horizontal line in the middle enclosing the number is imagined. Logically divide the number in to vertically two halves. Horizontally project all left half bright pixels of the numbers on to the segments 'f' and 'e' and project all right half bright pixels of the number on to the segment 'b' and 'c'.

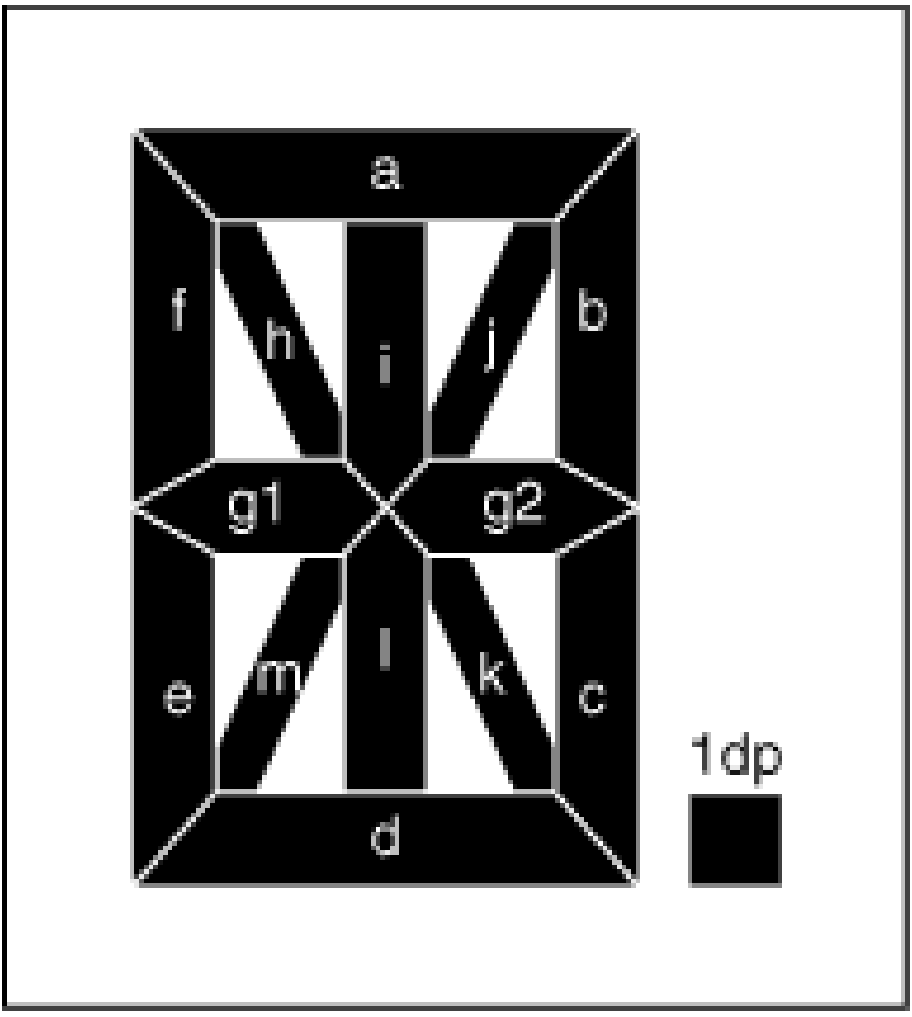


Fig. 6. Segmentation of the logical box in to 14 lines

Similarly divide the numbers in to three parts horizontally. Vertically project all the pixels in the first part on to segment 'a', project all the pixels in the second part on to the segment 'g' and project all the pixels in the third part on to segment 'd'. Determine the dynamic thresholds of each segment. These thresholds are used to find whether sufficiently large number of pixels is projected on to the corresponding segments. Considering the string segments that cross the thresholds identify the number. Figure 7 shows how the number '3' is divided horizontally and vertically.

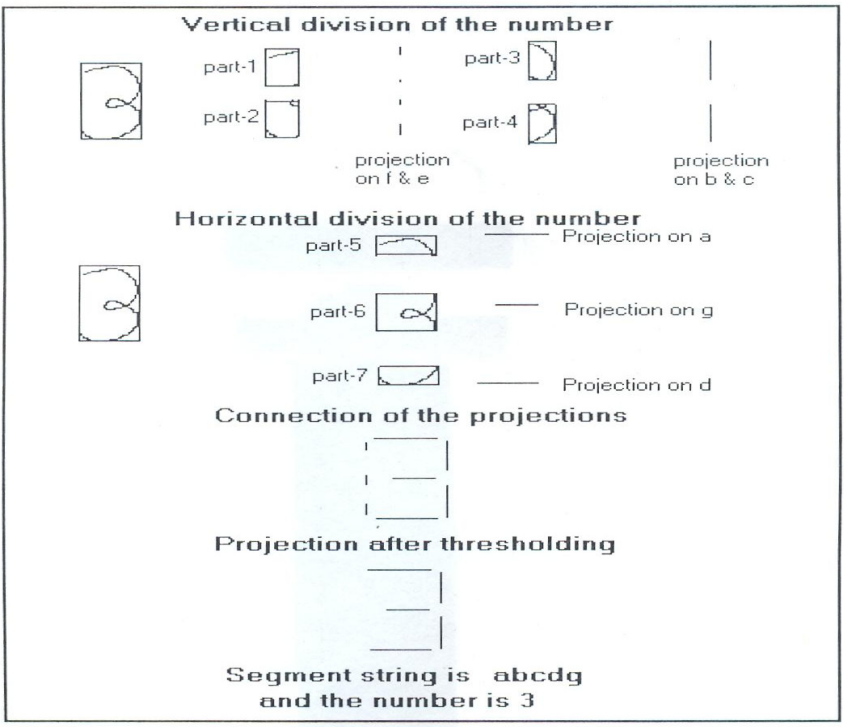


Fig. 7. Projection of Number 3 to segments and formulation of segments

Experimental Results

Video taken by the stationary camera is to be converted to frames or picture frames. Software video works is used to convert the video in to frames. A best frame is selected out of the number of frames generated by the software and it can be saved as a bitmap file.

In experimentation a sample set of 25 different vehicle images are taken from a camcorder. The recognition of number plates and recognition of characters varies from 70% to 80% for different images. In this experiment it will only recognize English characters and as well as numbers. It is assumed that the number plate is written in normal font, however some of the number plates are written in fancy styles, is difficult to recognize.

6 Conclusion

Segmenting the number plate in an image and recognizing the vehicle through its number is one the major task highlighted in the paper. The various activities involved in this process are taken from some existing algorithms. The simple recognition algorithm proposed here is suitable for recognizing the characters in a limited set.

The performance of the algorithm as a whole has shown satisfactory results, which could be used to record the vehicle numbers in database. However it is a real time activity, further enhancement specially in video processing and handling partially visible number plates are required to make it complete.

References

1. Zhong, Y., Karu, K., Jain, A.K.: Locating Text in Complex Color Images. Journal of Pattern Recognition 1.28(10), 1523 (1995)
2. Qi, W., Gu, L., Jiang, H., Chen, X.– R., Zhang, H.-J.: Integrating Visual, Audio and Text analysis for News video. Microsoft Research Publications
3. Jain, A.K., Yu, B.: Automatic Text Detection and Tracking in Digital Video. publications (2000)
4. Chanda, B., Datta Majumder, D.: Digital Image processing. Mc Graw Hill publications
5. Gonzales, R.C., Woods, R.E.: Digital Image Processing. Mc Graw Hill Publications
6. Qi, W., Gu, L., Jiang, H., Chen, X.R., Zhang, H.J.: Integrating visual, audio and text analysis for news video. Microsoft Research Publication
7. Eakins, J., Graham, M.: Content – based Image. University of Northumbria Publication, Newcastle
8. Vasudev, T., Hemathkumr, G., Guru, D.S., Nagabhushan, P.: Extension of 7-segment display concept for Numeral Recognition: A simple projection method Approach. In: Proceedings of the National Conference on Document Analysis and Recognition, Mandya, p. 57 (2001)
9. Ledenbaum, M., Sergey, V., Alexander, R., Roman, S.: Moving car license plate recognition. Israel Institute of technology
10. Mike constant The principle and practices of CCTV. the benehmark for CCTV United Kingdom

Hidden Markov Model with Computational Intelligence for Dynamic Clustering in Wireless Sensor Networks

Veena K.N.[1] and Vijaya Kumar B.P.[2]

[1] Dept. of Telecommunication Engg,
JNN College of Engg, Shimoga, Karnataka, India
Veena_k_n@yahoo.co.in
[2] Dept. of Computer Science Engg,
MS Ramaiah Institute of Technology, Banglore, Karnataka, India

Abstract. Most important challenge in Wireless Sensor Networks is to improve the operational efficiency in highly resource constrained environment based on dynamic and unpredictable behavior of network parameters and applications requirement. In this paper we have proposed a method of clustering and analysis, to study the system behavior with respect to network parameters and applications requirement. The method involves in the adoption of Fuzzy logic and Hidden Markov Model for the analysis of sensor node parameters and Computational intelligence for clustering. The simulations are carried out to evaluate the performance of the proposed method with respect to different parameters of sensor networks and applications requirement.

Keywords: Wireless Sensor Networks, Clustering, Hidden Markov Model, Neural networks, Fuzzy logic.

1 Introduction

In the recent years, Wireless Sensor Networks (WSNs) has found a wide variety of applications and systems with vastly varying requirements and characteristics. As a result, the complexities of the system design issues and their requirement have increased with respect to hardware and software. Integrating sensor nodes into sophisticated sensing, computational and communication infrastructures to form wireless sensor networks will have a significant impact on a wide array of applications ranging from military, scientific, industrial, health-care and domestic services. Various applications of WSN are Telemedicine, Habitat monitoring, Structure health monitoring, Active Volcano, Seismic monitoring, and Avalanche Victims. Wireless sensor networks are usually a large number of sensor nodes, which are tiny, compact and low cost embedded devices, which can be readily deployed in various types of unstructured environments within predefined and specified area or sometimes an approximate area of interest, either inside the phenomenon or very close to it. Efficiency of such systems would improve by providing distributed, localized and energy efficient techniques. Here the sensor nodes have native capabilities to detect the nearest neighbors and help to develop an ad-hoc network through a set of well-defined protocols. This leads to the existence of clusters of nodes, which can enhance the network

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 19–30.
springerlink.com

operations under network management, processing and aggregation of sensor data. Therefore there are many advantages with clustering and their analysis based on the system parameters and applications requirement. The systematic clustering by choosing the key parameters and response to their dynamic behavior in resource constrained sensor networking environment will be a challenging and complex issue.

In this work, we have proposed a method of clustering and analysis to study the dynamic behavior of the system that includes sensor network parameters and applications requirement. The method involves the Fuzzy Logic in fuzzifyng the sensor node parameters. Hidden Markov Model (HMM) is used in estimating a sensor node suitable for a particular application. We have used Kohonen Self Organization Map Neural Network (KSOM-NN) [1] as Computational Intelligence for clustering.

The organization of the paper is as follows. A brief discussion on some of the related works is explained in section 2. Section 3 explains the proposed Dynamic clustering model. Concept of clustering is explained in section 4. The description of the proposed clustering algorithm is narrated in section 5. Simulation and results are described in section 6 and concluding remarks are given in section 7.

2 Related Works

In this section a brief discussion on the related works pertaining to Fuzzy logic, Hidden Markov Model and clustering in WSNs are discussed. In [2] fuzzy logic systems to analyze the lifetime of a wireless sensor network are presented. It uses a type-2 fuzzy membership function (MF), where a Gaussian MF with uncertain standard deviation is used to model a single node lifetime in wireless sensor networks. Clustered WSN implemented in [3], it implements a protocol for trade-off between loads and increase in life expectancy of the network. In [4] a two-level fuzzy logic is utilized to evaluate the qualification of sensors to become a cluster head. In the first level the qualified nodes are selected based on their energy and number of neighbors.. Then, in the second level nodes overall cooperation is considered in the whole network with fuzzy parameters are discussed. A human movement monitoring, which also adapts to the new movement and new sensors using hidden markov model is implemented in [5]. In [6] an action recognition framework based on an HMM, which is capable of both segmenting and classifying continuous movements for the distributed architecture of Body Sensor Networks is presented. In [7] wearable sensor network that monitors relative proximity using Radio Signal Strength indication (RSSI), and then construct a HMM system for posture identification in the presence of sensing are discussed.

A distributed, randomized clustering algorithm [8] to organize the sensors in a wireless sensor network into clusters is discussed. The algorithms generate a hierarchy of cluster heads and observe that the energy savings increase with the number of levels in the hierarchy. The Linked Cluster Algorithm [9] where a node becomes the cluster head, if it has the highest identity among all the nodes within one hop of itself or among its neighbors is presented. Budget-based clustering approaches [10], message efficient distributed clustering [11] are proposed, involving autonomous sensor network clusters of bounded size by keeping lower overall message complexity. In [12] determination of the optimal number of clusters in an observation area is implemented.

3 Dynamic Clustering Model

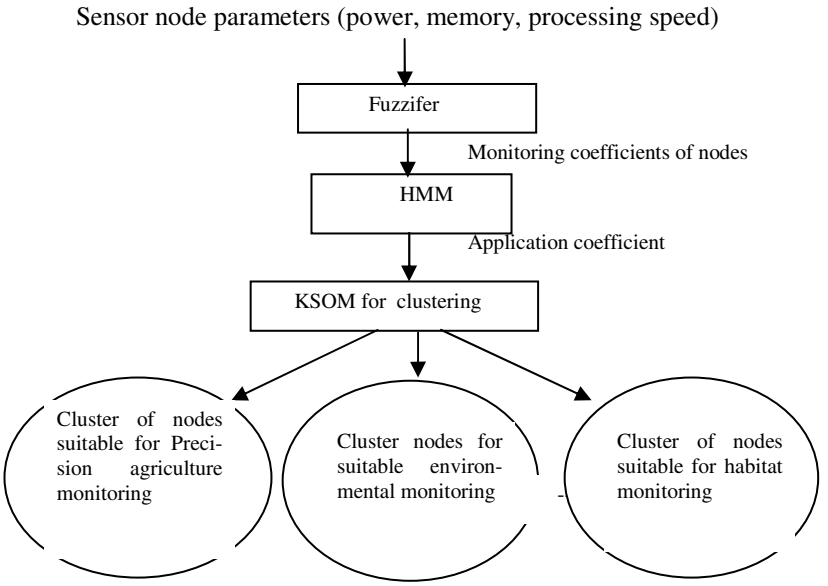


Fig. 1. Dynamic Clustering model

Dynamic clustering model using Neuro-Fuzzy Technique and Hidden Markov Model is shown in figure 1. The sensor data readings are fuzzified. Ex: Power present in sensor nodes, memory available and processing speed of sensor nodes are given to fuzzifier. The fuzzifier assigns monitoring coefficients for each sensor nodes according to predefined fuzzy rules. Ex: Monitoring coefficients for sensor node which can be used for critical monitoring, used for event monitoring, continuous and cannot be used for monitoring because of very low power availability. These coefficients are then given HMM to estimate the application coefficient for each sensor node that is suitable for various applications. Ex: A senor node can be suitable for Precision agriculture, Environmental monitoring, Habitat monitoring, Disaster monitoring etc. These coefficients of sensor node are given to KSOM neural network for clustering. Depending on the application requirement, cluster of sensor nodes suitable for Precision agriculture, environmental monitoring, seismic monitoring, habitat monitoring and disaster monitoring are obtained.

4 Clustering in WSN

Clustering is the architecture of choice as it keeps the traffic local and sensor nodes would send information only to the nearby cluster head within a fixed radius.

Clustering sensor nodes are advantageous because they conserve limited energy resources (power) and improve energy efficiency. In a cluster, each cluster will contain a cluster head. Each cluster head gathers information from its group of sensors, perform data aggregation and relay only relevant information to the sink as shown in figure 2. By aggregating the information from individual sensors, can abstract the characteristics of network topology along with applications requirement and it also reduces the bandwidth. This provides scalability and robustness for the network and increases the lifetime of the system.

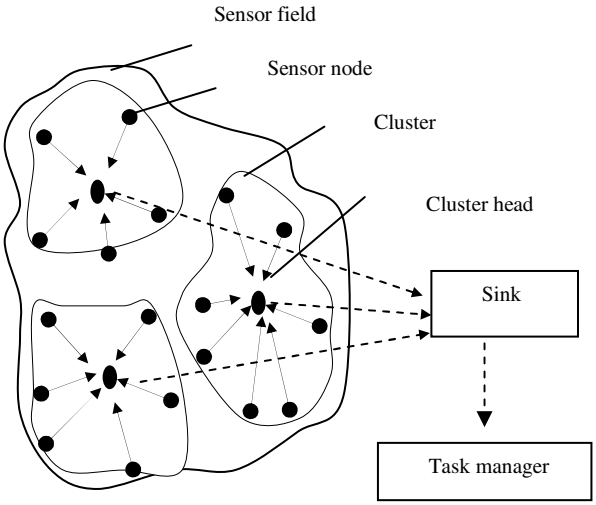


Fig. 2. Clustered Wireless Sensor Network

5 Proposed Clustering Algorithm

We propose a method for clustering and cluster analysis to study the behavior and operation of WSN with respect to network parameters and applications requirement. There are application parameters, which will have the influence over the resources available in WSN. Hence there is a need to understand how WSN behaves keeping its resource-constrained environment with respect to applications requirement. For clustering we have adopted the computational intelligence technique, which deals with the numerical data and does not use knowledge in the AI sense and exhibits computational adaptivity with fault tolerance [13]. This study is essential to enhance the clarity in understanding the behavior of WSN parameters pertaining to the specific applications to improve the operational efficiency. The proposed clustering algorithm is given below.

Algorithm 1. Proposed Dynamic Clustering Algorithm

Begin

1. *The sensor node parameters such as memory, power available, processing speed of each sensor nodes is considered.*
2. *These sensor node parameters are fuzzified using fuzzy logic with fuzzy rules as given in Table1.*
3. *The fuzzifier assigns monitoring coefficient for each node, depending on its power availability, memory availability, processing speed.*
4. *These monitoring coefficients are given to HMM to estimate application coeffi cient for each sensor node.*
5. *These application coefficients are given to KSOM neural network to obtain clusters depending on the application requirement.*

End

5.1 Fuzzy Logic

The model using fuzzy logic consists of a fuzzifier, fuzzy rules, fuzzy inference engine, and a defuzzifier. Input variables used are sensor node parameters such as power available in sensor node, memory and processing speed. Membership function and Fuzzy rules are applied to the sensor node parameters. The membership functions developed and their corresponding linguistic states are represented in Table 1 and figures 3 through figure 6. The Fuzzifier assigns the monitoring coefficients for each sensor nodes.

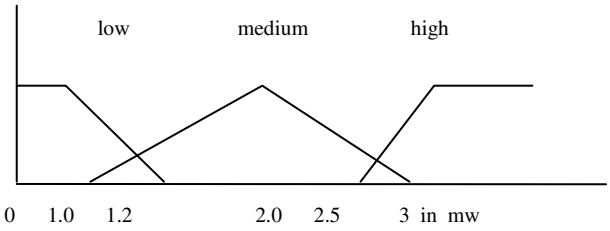


Fig. 3. Membership function for power present in sensor node

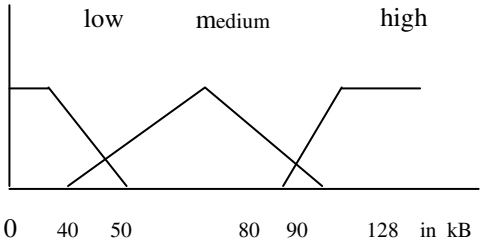


Fig. 4. Membership function for memory available in sensor node

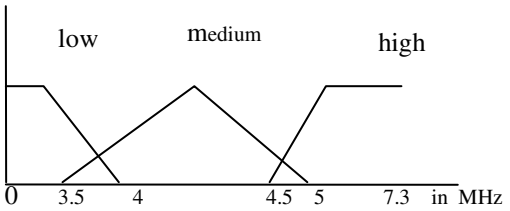


Fig. 5. Membership function for processing speed in sensor node

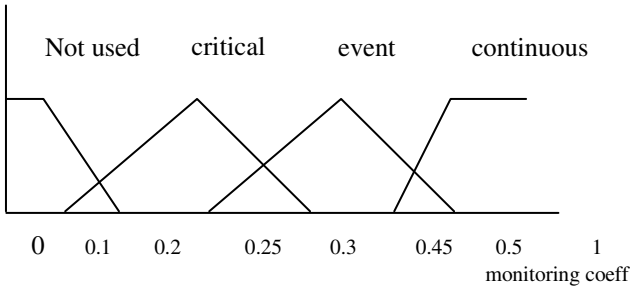


Fig. 6 Membership function for monitoring coefficient of sensor node

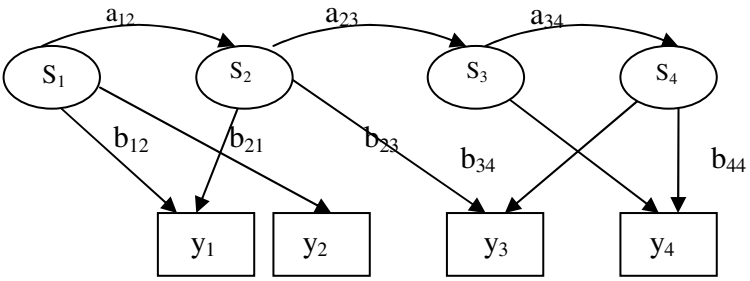
Table 1. Fuzzy rules

	Power	Memory	Processing speed	Monitoring coefficient
1	low	low	low	Not used
2	medium	medium	medium	critical
3	medium	high	medium	event
4	medium	medium	high	event
5	medium	high	high	event
6	high	high	medium	continuous
7	high	medium	high	continuous
8	high	high	high	continuous

Sensor nodes which have high power, memory and processing speed can be used for continuous monitoring. Sensor nodes with medium power, high memory and high processing speed can be used for event monitoring, sensor nodes with medium power and low memory and processing speed can be used for critical monitoring. Sensor nodes with low power and low memory and low processing speed cannot be used for monitoring.

5.2 Hidden Markov Model

A Hidden Markov Model (HMM)[14] is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (*hidden*) states.



s- states (application coefficients); y-observations (monitoring coefficients);
a- State transition probabilities; b-output probabilities;

Fig. 7. General Architecture of a Hidden Markov Model

A Hidden Markov Model is shown in the figure 7 and it is denoted by equation 1.

$$\lambda=(\pi,A,B) \tag{1}$$

N applications of WSN are modeled as N hidden states, with $S = \{S_0, S_1, \ldots, S_{N-1}\}$ the set of hidden states, in our case N=4 and q_t the hidden state at time t.

M is the number of observable parameters, M monitoring coefficients of sensor nodes are considered, with $y = \{ y_0, y_1, \ldots, y_{M-1}\}$ the set of observable parameters and O_t the observable state at time t.

$A=\{a_{ij}\}$, the transition probabilities between the hidden states S_i and S_j,
where $a_{ij} = P[q_{i+1} = S_j \mid q_t = S_i]$, $0 \le i, j \le N-1$.

$B = \{b_j(k)\}$ - the probabilities of the observable states y_k in hidden states S_j,
Where $b_j(k) = P[O_t = y_k \mid q_t = S_j]$, $0 \le j \le N-1$, $0 \le k \le M-1$

The following transition state matrix probabilities and emission probabilities are considered for HMM computation.

$$A=\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0.8 & 0.2 & 0 \\ 0 & 0 & 0.8 & 0.2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad B=\begin{bmatrix} 07 & 0.3 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\pi = \{\pi_i\}$ - the initial hidden state probabilities, where $\pi_i = P[q_1 = S_i], 0 \le i \le N-1$.

Learning

The parameter learning task in HMM is to compute, given an output sequence or a set of such sequences, the best set of state transition and output probabilities. The task is usually to derive the maximum likelihood estimate of the parameters of the HMM given the set of output sequences. To solve this Baum–Welch algorithm is used. The Baum–Welch algorithm is a forward-backward algorithm, and is a special case of the expectation-maximization algorithm. The forward algorithm is used to compute the

probability of an output sequence given the model parameters. The Viterbi algorithm is used to find the most likely sequence of hidden states to have generated a particular output sequence, given the model parameters.

5.3 Kohonen Self-Organizing Map Neural Network (KNN)

Clustering computation is carried out using Kohonen Self–Organizing Map Neural Networks (KSOM-NN) algorithm to cluster the sensor nodes depending on the parameters of interest. The KSOM-NN is competitive, feed forward type and unsupervised training neural network. They have the capability of unsupervised learning and self-organizing properties, is able to infer relationships and learn more as more inputs are presented to it. The Application coefficient of each sensor nodes is considered to cluster sensor nodes using Kohonen Self-Organizing Map Neural Networks. Cluster of sensor nodes suitable for various applications are formed. The figure 8 shows a typical example of a KSOM-NN [15] with 4 input and 20 output neurons.

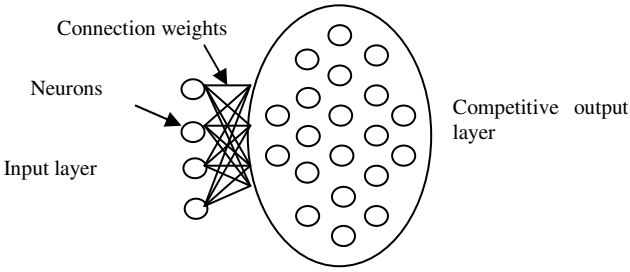


Fig. 8. A typical KSOM-NN model with 4 input and 20 output neurons

Let the input pattern be: $I = (I_1, I_2,.,I_{|V|})$, where there are |V| input neurons in the input layer. Now suppose there are Q neurons in the output (competitive) layer, let O_j be the j^{th} output neuron. So the whole of the output layer will be: $O= (O_1,O_2,.,O_Q)$ for each output neuron O_j there are |V| incoming connections from the |V| input neurons. Each connection has a weight value W. So, for any neuron O_j on the output layer the set of incoming connection weights are: $w_j=(w_{j1},w_{j2},.,w_{j|v|})$. The Euclidean distance value D_j of a neuron O_j in the output layer, whenever an input pattern I is presented at the input layer, is:

$$D_j = \sqrt{\sum_{i=1}^{|V|.} (I_i - W_{ji})^2} \; . \tag{2}$$

The competitive output neuron with the lowest Euclidean distance at this stage is the closest to the current input pattern, called as winning neuron. The neighborhood size and weight updating computation are

$$h_t=ho\ (1-t/T) \qquad \text{and} \qquad w_j new = w_j old + \alpha (I - w_j old) \tag{3}$$

The size of the neighborhood h, starts with a big enough size and decreases with respect to learning iterations. Where, h_t denotes the actual neighborhood size, ho denotes the initial neighborhood size, t denotes the current learning epoch and T denotes the total number of epochs to be done. Where, α is the learning rate parameter, typical choices are in the range [0.2 . . 0.5].

6 Simulation and Results

In order to evaluate the performance of the proposed method simulation is carried out based on sensor node characteristic features such as power present in sensor nodes, memory available and processing speed. Fuzzy logic is used to fuzzify the sensor node parameters. The fuzzifier gives the membership function coefficients for monitoring. Then the monitoring coefficients are given HMM to estimate the application coefficient for each sensor node. These application coefficients are given to KSOM to obtain dynamic clusters depending on the application requirement.

Clustering of sensor node using KSOM-NN is computed for various numbers of nodes by taking monitoring coefficients of sensor node. The clustering implementation is carried out using C++ programming language by defining suitable classes and structures. Simulation experiments are carried out rigorously by taking different number of nodes in sensor networks, i.e., 100 to 1000. The simulation program is run for many iterations until they converge to stable clusters. A plot of monitoring coefficients for various parameter of a sensor node such as power, memory available and processing speed is shown in figure 9, and figure 10. The result shows that as power, memory and processing speed is high for a sensor node, then monitoring coefficients generated is also high. Such type of sensor node can used for continuous monitoring. If power present in a sensor node is very low, then monitoring coefficient generated is low such sensor nodes cannot be used for monitoring purpose. These results help us in analyzing, whether a node is capable of monitoring a phenomenon or not and how well it can monitor an area of interest.

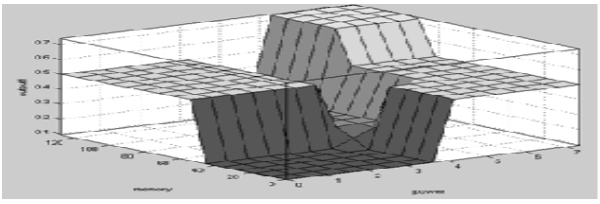


Fig. 9. Monitoring coefficients of sensor node with respect to power and memory

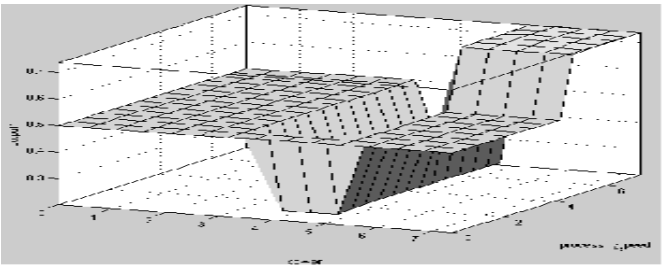


Fig. 10. Monitoring coefficients for sensor node with respect to power and processing speed

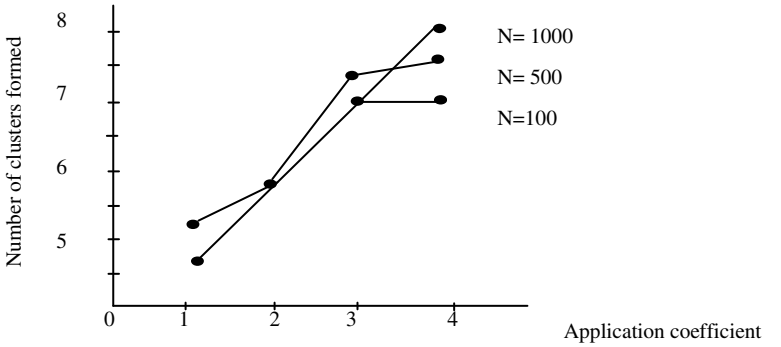


Fig. 11. Cluster formed with respect to application requirement

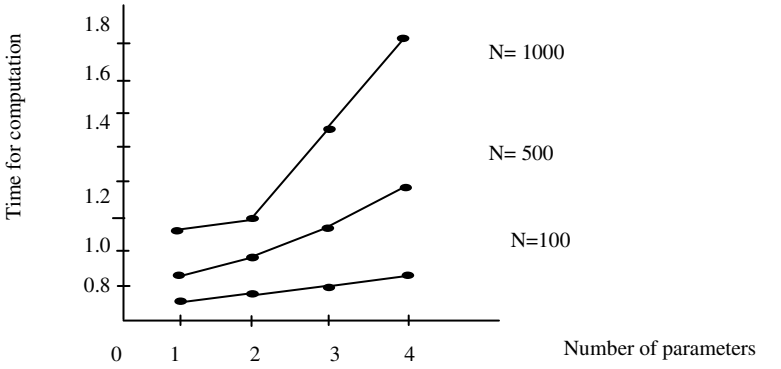


Fig. 12. Time for computation to form clusters with respect to parameters of sensor node

The figure 11 shows the number of clusters formed with respect to application coefficient of sensor nodes in WSN. The results show that the number of clusters formed is dependent of the number of parameters of sensor nodes chosen and the application requirement. These results are essential and can be easily used in various applications depending on the critical region of interest. For example, nodes which can be used for continuous monitoring are suitable for precision Agriculture, and nodes which can be used for event monitoring are suitable for monitoring of seismic area. The computation time for clustering is measured with respect to application and number of sensor nodes as shown in figure 12. These results show on how the individual parameters affect the formation of cluster and its size. These results would be used in real time decision making applications and in adaptive systems.

7 Conclusion

In this work we have proposed a method for clustering and their analysis to study the dynamic behavior of the cluster formation based on the system parameters and application requirements. The technique involves the adoption of computational intelligence to form clustering and analyzing sensor node parameters in WSN. We have used computational intelligence, Neuro-Fuzzy techniques and Hidden Markov Model for the above computations. The simulations are carried out to evaluate the performance of the proposed method with respect to different parameters of sensor networks and applications requirement. Since we are analyzing and estimating sensor nodes depending on the application requirement, proactive decisions can be taken. The decisions can be redeployment of sensor nodes , when we estimates some of the sensor cannot be used for monitoring because of very low power, or sensor nodes are less in number for a particular application requirement. This method of study would give scope for more understanding the chaotic behavior of environment and system parameters of WSN to enhance the operational efficiency.

References

1. Haykin, S.: Neural Networks: A Comprehensive foundation, 2nd edn. Macmillan college publishing, Newyork (1995)
2. Shu, H.: Analysis Using Interval Type-2 Fuzzy Logic System. IEEE Transactions on Fuzzy Systems 16(2), 416–427 (2008)
3. Azad, I., Mashhad, I.: CFGA: Clustering Wireless Sensor Network Using Fuzzy Logic and Genetic Algorithm, Wireless Communications. In: 7th International Conference on Networking and Mobile Computing (WiCOM), September 23-25, pp. 1–4 (2011)
4. Torghabeh, N.A.: Cluster head selection using a two-level fuzzy logic in wireless sensor networks. In: 2nd International Conference on Computer Engineering and Technology (ICCET), April 16-18, vol. 2, pp. V2-357 – V2-361 (2010)
5. Guenterberg, E., Ghasemzadeh, H., Bajcsy, R.: A Method for Extracting Temporal Parameters Based on Hidden Markov Models in Body Sensor Networks With Inertial Sensors. IEEE Transactions on Information Technology in Biomedicine 13(6), 1019–1030 (2009)
6. Guenterberg, E., Ghasemzadeh, H., Loseu, V., Jafari, R.: Distributed Continuous Action Recognition Using a Hidden Markov Model in Body Sensor Networks. In: BodyNets 2008, Proceedings of the ICST 3rd International Conference on Body Area Networks (2008) ISBN: 978-963-9799-17-2
7. Quwaider, M., Biswas, S.: Body posture identification using hidden Markov model with a wearable sensor network. In: Proceedings of the ICST 3rd International Conference on Body Area Networks ICST, BodyNets 2008, Brussels, Belgium, pp. 978–963 (2008) ISBN: 978-963-9799-17-2
8. Bandyopadhyay, S., Coyle, E.J.: An Energy Efficient Hierarchical Clustering Algorithm for Wireless Sensor Networks. In: IEEE INFOCOM (2003)
9. Baker, D.J., Ephremides, A.: The Architectural Organization of a Mobile Radio Network via a Distributed Algorithm. IEEE Transactions on Communications 29(11), 1694–1701 (1981)
10. Tzevelekas, L., Stavrakakis, I.: Directed Budget-Based Clustering for Wireless Sensor Networks. In: IEEE International Conference, MASS (October 2006)

11. Krishnan, R., Starobinski, D.: Efficient clustering algorithms for self organizing wireless sensor networks. Ad Hoc Networks 4(1), 36–59 (2006)
12. Wang, L.-C., Liu, C.-M., Wang, C.-W.: Optimizing the Number of Clusters in a Wireless Sensor Network Using Cross-layer Analysis. In: IEEE International Conference, MASS (October 2004)
13. Witold, P., Vasitakos, A.: Computational Intelligence in Telecommunication Networks. CRC press LLC, USA (2001)
14. Karray, F.O., De Silva, C.: Soft Computing and Intelligent System Design Theory, Tools and Applications. Pearson Education (2004)
15. Vijay Kumar, B.P., Venkataram, P.: Reliable Multicast routing in Mobile networks: a Neural Network approach. IEE Proc.-Communication 150(5), 377–384 (2003)

Assessment of Workload Using Shapely Value in Distributed Database

S. Jagannatha, T.V. Suresh Kumar, D.E. Geetha, and K. Rajani Kanth

M.S. Ramaiah Institute of Technology
{Jagannatha,tvsureshkumar,rajankanth}@msrit.edu
devangelin@gmail.com

Abstract. Performance of the software system to be achieved when the data distribution and load balancing takes place properly. Performance prediction and workload assessment in early design stages is important factor to be considered in distributed database system. Design level fragmentation and allocation helps to assesses workload of applications. One of the application of the shapely value is used in the domain of distributed data base system is to measure the relative importance of individual server contribution. We consider the individual server cooperation. Our goal is to study database distribution issues that, besides workload. We propose an algorithm to measure the individual server contribution for transaction processing system during the early design stages. We develop model using UML2.0.with the case study.

1 Introduction

Developments distributed database systems (DDS) is a collection of sites connected by a communication network, in which each site is a database system in its own right, but the sites have agreed to work together, so that a user at any site can access data anywhere in the network exactly as if the data were all stored at the user's own site. Distributed Design involves making decisions on the fragmentation based on the requirement and placement of data across the sites of a computer network [10]. In a distributed design has two phases: fragmentation and allocation.

The Unified Modeling Language (UML) is widely used as modeling language. UML is a graphical modeling language that is used for visualizing, specifying, constructing and documenting software systems [11]. Performance models can be generated from use cases during requirements phase.

Game theory aims to help us in understanding situations in which decision makers interact. An important aspect associated with the solution concepts of Cooperation Game Theory (CGT) is the equitable and fair sharing of fragments which are distributed in system.

The remaining part of this paper is organized as fallows. In Section 2 we proposed to study the related work. Then we proposed a methodology and algorithm using game theory concepts of distributed database design in section 3. Application and methodology proposed. in Section 4. Numerical results and analysis proposed in section 5. Finally a conclusion and future work is given in Section 6.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 31–40.
springerlink.com

2 Related Work

Various researchers made significant contribution to distributed databases and shapely value and its applications. In [15] author address forming procurement networks for items in supply chain management. Procurement networks involves a bottom-up assembly of complex production, assembly, and exchange relationships through supplier selection and contracting decisions based on cooperative game theory concepts. In [17][18] author describes cooperative profits during the cooperation, which seriously affects the efficiency and effectiveness. So it is necessary to allocate cooperative profit reasonably among members to improve the cooperative relationship between members in stability of supply chain. In [16] this paper the author address marginal contribution of each sensor the observation of spatially correlated sensor field, and can be used to allocate the probability of each sensor's being measured in proportion to its contribution using shapely value. In [6] authors describes the coalitional games were used in the domain networks and it is used in measure the relative importance of individual nodes. Author develops exact analytical formulas for computing.

In [8][16][6] author describes the Supply chain Profit allocation issue based on the game theory concepts. In [4] [3] the author considers how to design coordination strategies to achieve the coordination in decentralized supply chain, considering different types of pricing, and they propose the algorithm for sharing of revenue by supply chain coordination. In [16] author proposed two model ie Stackelberg model and Cournot model for a multistage cooperation. A multi-stage cooperation and competition model based on flexibility in port service supply chain. In [15] this paper author addresses the interruptible load management by allocation and quantification of wholesale price spike. The Shapley value in cooperative game theory is applied to allocate the benefit among the participants. In [10] author investigate the performance of Marginal cost (MC) and Shapley value (SH) mechanisms for the sharing the cost of multicast transmissions. The supply chain coordination mechanism between supply and demand are addressed in [9].

In all the approaches, the researchers use application of shapley value in supply chain management system. Some authors address the cooperation game theory concept by allocation of cost sharing by cooperative mechanism and some address the node centrality, domain networks and it is used in measure the relative importance of individual nodes. Authors do not consider the application of shapley value in server cooperation in transaction processing system. Using Shapley value concept authors do not consider the performance analysis, workload assessment of server in distributed database system during early design stages. Keeping these in view we propose an algorithm to predict the performance of distributed database system during early stages of software system and assess the workload of the server. We use the shapely value for computing the individual server workload.

3 Methodology

3.1 Proposed Methodology

The algorithm for the proposed methodology is given below. The algorithm uses the procedure to compute the server cooperation using the shapely value for assesses the workload of the server and also used to measure the relative importance of individual nodes.

Step 1. We propose the network infrastructure by considering the number of servers which are connected in high speed LAN

Step 2. We propose the applications and identify the functional requirements, data requirements, and identify important scenarios of the system.

Step 3. Based on the function requirements, develop use case diagram by using UML.

Step 4. Identify the entity, relation ship among the entities, types of attributes in each entity. Draw E-R diagram and develop global conceptual schema using E-R diagram

Step 5. In a given application identify the set of query that arise, data requirement of the server and propose deployment environment.

Step 6. Attribute usage matrix determines the exact number of attributes is associated with other attributes and identify the possible predicates and develop predicate usage matrix. These matrix which helps for fragmentation.

Step 7. Divide the global schema into a set of horizontal, vertical and mixed fragments.

Step 8. Fragment in step seven are distributed into the proposed system architecture. ie allocate these fragments into servers using allocation algorithm .

Step 9. Identify the probability of server interacting with the other server for computing the transaction. The results are input to the step ten in computing the shapley value.

Step 10. The details of Shapley value computation is given in the section 3.3.

3.2 Preliminaries

A cooperative game with transferable utility [1] is defined as the pair (N; v) where N = {1,2,…..n} the set of players and v : 2N → R is a real-valued mapping with v(φ) = 0. The concept of Shapley value, which was developed axiomatically by Shapley [2], takes into account the relative importance of each player to the game in deciding the payoff to φ(N, v) = (φ1(N, v); φ2(N, v) ; φn(N, v)) the Shapley value of the transferable unit game (N, v). Mathematically, the Shapley value, φi(N, v), of a player i is given by,

$$\varphi i(N; v) = \sum_{C \subseteq N \setminus i} \frac{|C|!(n-|C|-1)}{n!}\{v(C \cup \{i\}) - v(C)\}$$

where φi(N; v) is the expected payoff to player i.

3.3 Algorithm for Calculating Shapley Value :(Step 10 of Methodology)

```
Input the number of servers N
for all scenarios S do
        Develop Use case model U
        Develop global schema
          Fragment and allocate in proposed system architecture.
        for all  use case U do
                Input coalition function V ie the probability
                     each server     contribution with other server
                   Input V(1,2,….. N) ←100
                        (All combination of servers)
                 for  i    ← 1 to  N    do
                             v(i) ←   0
                end for
                 for  i ←  1  to  N     do
                     for j ←  i + 1  to  N   do
                                    input V(i, j)
                         end for;   end for
                {calculate the shapley value }
                 for i ←  1   to  N do
                   Generate the permutation of servers
                for each server do
                  Compute shapley value ie individual server  workload
                   end  for
                end for
end for.
```

4 Application and Methodology

4.1 Description of the Problem

The banking system we have considered is highly distributed in nature. The database of the application deployed in these servers where application resides. We consider three servers customer accounts, Loans, and the Employees for illustrative purpose and we assume that data base index exist in three servers. We focus on transaction of the customer, loan processing, employee salary, leaves monitoring, attendance, queries etc. The data base fragments are allocate in to these three servers using allocation algorithm given in[5].

4.2 Illustration

The important scenario as represented in use case diagram in figure 1. Identify the major entity and attributes and draw ER diagram, develop global conceptual schema for example Customer Accounts, Employee, Loan, Leave, Customer Transaction, Salary are in the global conceptual schema. The major transaction is deposit, withdrawn, loan processing, balance enquiry; know the status of salary, Status of leave etc. The attribute usage and predicate usage matrix are shown in the table 1 and

table 2 respectively. The deployment environment as shown in figure 2 which contains three servers which we proposed.

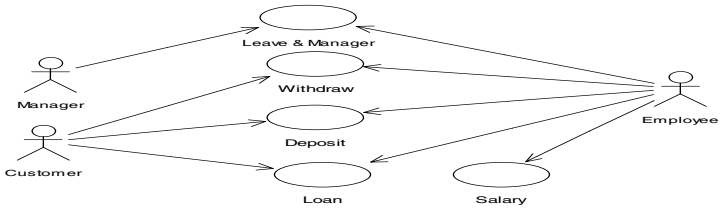


Fig. 1. Use case diagram for Banking system

Table 1. Attribute usage matrix

Method/Att.	Cust name	ACC No	Cust Address	Bal	Type of Trans	signature	Access Freq.
Deposit	1	1	0	1	1	0	40
With draw	1	1	0	1	1	1	60
Bal inquiry	1	1	0	1	0	0	40

The set of transaction as T1: Deposit (Acc no = (P1), T2: Withdraw : (P1) , T3: Loan Enquiry : P1,P3 T4: Status of Accounts (P4)(P1), Display (P1)(P2)(P3)(P4)

Table 2. Predicate usage matrix

Transaction	P1	P2	P3	P4	Acc Feq
T1	1	0	0	0	40
T2	1	0	0	0	50
T3	1	0	1	1	60
T4	1	1	1	1	70

Based on the attribute usage, application requirement divide the global conceptual schema into set of fragments and allocate these into server. F1 ← CUST (No, Name, Accono, Loan No,Adds),F2 ← SAL(Eno, Ename, Bp, Adds, Desc, DOJ).

F3 ← LOAN (Lno, Cno, Amt,Type) etc Allocate these fragments in to Customer, Salary, Loan server respectively.

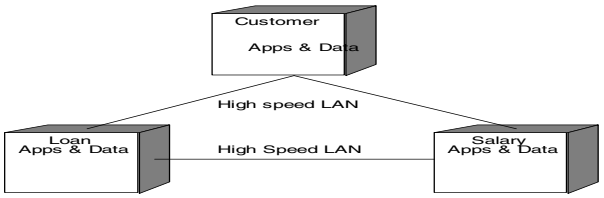


Fig. 2. Deployment diagram for case study

Determine the probability of server interacting with other server by analysis. For example the loan server S1 and Customer server S2 must exchange of data. The probability of cooperation with S1, S2 is represented in value function ie V(S1, S2) = 40. This value represents the Percentage of server coordinating with S1, S2. The computational results are represented in section 5.

5 Numerical Results and Analysis

The term V is the probability of server cooperation with the other server in query processing. Let N be the number of servers in the system architecture. The servers are involved in executing a query by using different fragment in the distributed system.

Table 3. Characteristic function and Shapley value of three servers

V(1,2)	V(1,3)	V(2,3)	Shapley value(Φ1)S1	Shapley Value(Φ2)S2	Shapley value (Φ3)S3
20	30	30	31.6	31.6	36.6
20	30	70	18.3	38.3	43.3
20	80	30	40	15	45
20	70	80	21.6	26.6	51.6
80	20	70	26.6	51.6	21.6
80	20	30	40	45	15
80	80	30	50	25	25
80	80	90	30	35	35
50	50	90	20	40	40
50	90	50	40	20	40
50	80	90	25	30	45
90	50	50	40	40	20
90	50	90	26.6	46.6	26.6
90	50	50	40	40	20
20	20	50	23.3	38.3	38.3
20	50	20	38.3	23.3	38.3
20	50	50	28.3	28.3	43.3
50	20	20	38.3	38.3	23.3
50	22	50	28.6	42.6	28.6
50	50	15	45.	27.5	27.5
50	50	50	33.3	33.3	33.3
22	55	90	16.6	33.6	50.16
19	90	55	33.16	15.6	51.6
45	19	90	14	49.5	36.50
50	90	22	49.3	15.3	35.33
90	25	53	34.83	48.83	16.33
96	60	26	50.6	33.3	15.66

Let V(S1, S2, S3) = 100 indicates the 100% cooperation of all servers in transaction processing under distributed database system. Let V(S1, S2) = 40 is the probability of sever cooperation in query processing using two servers S1 and S2. and V(S1)=V(S2)= V(S3) = 0 indicates that the probability of cooperation is zero ie no transaction processing with the single server alone. The computation of shapely value by varying combination of servers (S1,S2),(S2,S3),(S1,S3) is given in the table 3. The probability of possible coordination of (S1,S2), (S2,S3) and (S1,S3) of servers are takes as low, medium and maximum. The computational results of shapely values are mentioned in the table 3.

In table 3 the probability of servers coordination S1, S2 represented in column 1 as V (S1, S2). Similarly V (S1, S3) and V (S2, S3). We have considered low, medium and maximum likelihood of contribution of servers. The respective shapely value which essentially provides workload of each server for processing of query. Earlier the database are fragmented and distributed on to different servers . For illustration we have considered three servers . For example if V(S1, S2) = 20 similarly is V(S1, S3) = 30 and V(S2, S3) = 30. The computed value of shapely value of the server S1 , S2 and S3 are 31.6, 31.6, and 36.6 respectively. The results represents workload of the individual sever..

Table 4. Statistical values

Parameters	V(1,2)	V(1,3)	V(2,3)	Shapley value(Φ1)S1	Shapley Value (Φ2)S2	Shapley value (Φ3)S3
AVERAGE	50.23077	50.80769	54.42308	32.02654	33.81269	34.12385
STD DEV	26.88391	24.9624	25.48281	9.830821	10.73288	10.88328
MODE	50	50	50	40	38.3	43.3
MEDIAN	50	50	50	32.38	34.3	35.915

From the table 4 we observed that the average probability of server contribution in various combinations is almost is 50 percentage correspondingly the wok load handling by the server S1, S2, and S3 average is 32.02, 33.81 and 34.12. respectively. Hence the results shows that the servers are equally bearing the workload . Similarly the standard deviation, mode and median shows the load sharing by the server is consistent.

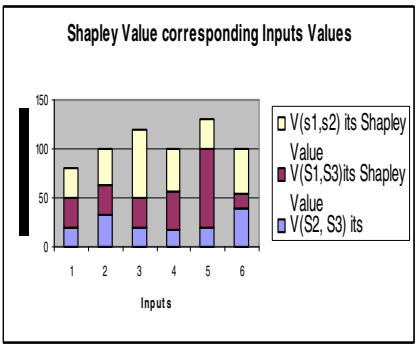


Fig. 3. Load sharing of servers with corresponding value function of the servers

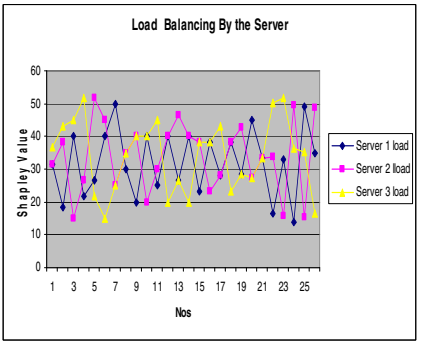


Fig. 4. The shapley Value of three Server vs the server coordination

We observed from the graph in figure 3 indicates the column 1, 3 and 5 are the probability of server contribution with the combination of (S2, S3), (S1,S3) and (S1, S2) values the corresponding shapley are represented by the server S1, S2, and S3 in column 2, 4 and 6 respectively. The computed results show the workload managed by server with different probable input values by the server coordination. The graph in figure 4 indicates the server cooperation in different values of server combination. These represents the all possible combination of input values, the servers are sharing the workload .

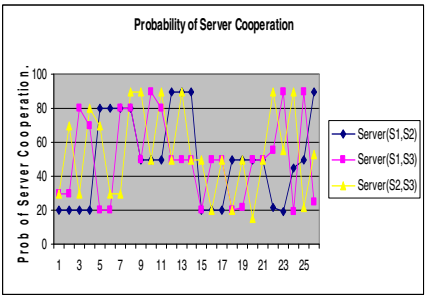


Fig. 5. Probability of Server cooperation vs Server S1,S2 and S3 in different inputs

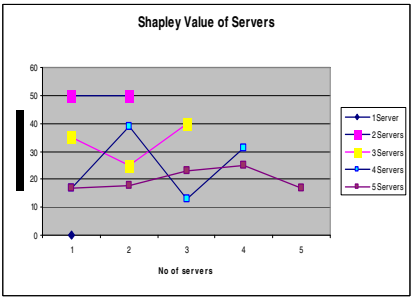


Fig. 6. Work load sharing by the different servers

The graph in figure 5 indicates server contribution in distributed system in sharing the workload. When the number of servers is increases the load is mutually sharing by transaction processing system. In figure 7 shows the server behavior in load sharing in transaction processing system.

6 Conclusion

In this paper we proposed an algorithm for assessing the workload of the sever. in distributed database system in early design stages. We address the relative importance of each server by computing SV in the proposed architecture. The key requirements are to understand the importance of each server which are involved in processing a query. We model the application with UML and simulate using C language and report the results. Further the system can be simulate and find the various performance parameters as response time, system utilization and queue length average waiting time, idle time of the server, average service time, average time between arrival, and result can be reported.

References

1. Anda, B., Dreiem, H., Sjoberg, D.I.K., Jorgensen, M.: Estimating Software development Effort based on Use Cases – Experiences from Industry, http://www.idi.ntnu.no/emner/tdt4290/docs/faglig/uml2001-anda.pdf
2. Petriu, D., Woodside, M.: Analysing Software Requirements Specifications for Performance. In: Proceedings of the 3rd International Workshop on Software and Performance, Rome, Italy, July 24 - 26, pp. 1–9 (2002)
3. Cruijssena, F., Borma, P., Fleurena, H., Hamers, H.: Insinking: a methodology to exploit synergy in transportation (2004)
4. Liu, G.-Q., Zhang, F.-H.: Research on Supply Chain Coordination with the Consideration of Pricing and Transportation Cost. In: Proceedings of the IEEE International Conference on Automation and Logistics, Qingdao, China (September 2008)
5. Huang, Y.F., Chen, J.H.: Fragment Allocation in Distributed Database Design. Journal of Information Science and Engineering 17, 491–506 (2001)
6. Yi, C.-H.: Using Modified Shapley Value to Determine Revenue Allocation within Supply Chain. In: 2009 International Conference on Information Management, Innovation Management and Industrial Engineering (2009), doi:10.1109/ICIII.2009.26, Jkl;h 978-0-7695-3876-1/09 $25.00 © 2009 IEEE
7. Shapley, L.S.: A Value for N-Person Games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games, vol. I. Princeton University Press, Princeton (1950)
8. Chen, L., Shen, M., Chen, C.: A Research in Supply Chain Profit Allocation Based on Cooperation Game Theor. In: 2010 International Conference on System Science, Engineering Design and Manufacturing Informatization (2010), doi:10.1109/ICSEM.2010.152, 978-0-7695-4223-2/10 $26.00 © 2011 IEEE
9. Hu, L.-Y., Chungu, Y.-Q., Jiang, Z.-S.: Research on the Coordination Mechanism Model of the Three-level Supply. In: Chain HU 2007, 14th International Conference on Management Science & Engineering, Harbin, P.R.ChinaKjl, August 20-22 (2007)
10. Garg, N., Grosu, D.: Performance Evaluation of Multicast Cost Sharing Mechanisms. In: 21st International Conference on Advanced Networking and Applications (AINA 2007) (2007) 0-7695-2846-5/07 $20.00 © 2007
11. Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge (1997)

12. Narayanam, R., Narahari, Y.: A Shapley Value Based Approach to Discover Influential Nodes in Social Networks. IEEE Transactions on Automation Science and Engineering, IEEE TASE (2010)
13. Navathe, S., Karlapalem, K., Ra, M.: A mixed fragmentation methodology for initial distributed database design. Journal of Computer and Software Engineering 3(4), 395–426 (1995)
14. Byun, S.-S., Moussavinik, H., Balasingham, I.: Fair Allocation of Sensor Measurements Using Shapley Value. In: 2009 IEEE 34th Conference on Local Computer Networks (LCN 2009), Zürich, Switzerland, October 20-23 (2009)
15. Zhang, S., LiSs, Q.: Application of Cooperative Game Theory to Benefit Allocation of Interruptible Load Management (2011) 978-1-4244-6255-1/11/$26.00 © 2011 IEEE
16. Sun, S., Zhang, J.: Tai'anEvolutionary Game Analysis on the Effective Coopetition Mechanism of Partners within High Quality Pork Supply Chain Support by National Science Foundation in China (No. 70572100) (2008) 978-1-4244-2013-1/08/$25.00 © 2008 IEEE
17. Chandrashekar, T.S., Narahari, Y.: A Shapley Value Analysis to Coordinate the Formation of Procurement Networks. In: Proceedings of the 3rd Annual IEEE Conference on Automation Science and Engineering, Scottsdale, AZ, USA, September 22-25 (2007)
18. Yan, J., Zhang, Y.: Supply Chain Business Process Reengineering Directed by the Project. In: Proceedings of the IEEE International Conference on Automation and Logistics, Jinan, China, August 18 - 21 (2007)
19. Wang, L., Zhou, Y.: Research on Cooperation Profit Allocation in Three-stage Supply Chain Based on Distribution Channel (2008) 978-1-4244-2108-4/08/$25.00 © 2008 IEEE
20. http://www.omg.org
21. Zheng, Y., Zhang, S., Chen, X., Liu, F.: Application of Modified Shapley Value in Gains Allocation of Closed-loop Supply Chain under. Third-Party Reclaim Energy Procedia 5, 980–984 (2011)

Modeling and Estimation of Cooperative Index for Multi-Agent Systems Using Execution Graph

S. Ajitha, T.V. Suresh Kumar, D. Evangelin Geetha, and K. Rajani Kanth

M.S. Ramaiah Institute of Technology
{ajithasankar,rajanikanth.kundurthi}@gmail.com
{sureshkumar,degeetha}@msrit.edu

Abstract. A multi-agent system (MAS) is a system composed of multiple interacting intelligent agents. MAS can be used to solve problems that are difficult or impossible for an individual agent to solve. The different characteristics of MAS help in solving highly complex distributed problems. One of the important characteristics of MAS is its cooperative nature. This character helps different agents to interact with each other by exchanging messages. One of the major challenges in MAS is quantifying the cooperation between agents. In this paper, we propose a new methodology to compute the cooperative index of MAS in the early stages of its development. For this calculation, execution graph is used to model the software specifications. The proposed methodology is illustrated with a case study and the results are compared with the cooperative index obtained from the Unified Modeling Language (UML) sequence diagram.

Keywords: Multi-Agent Systems, Cooperative Index, Execution Graph.

1 Introduction

Multi Agent System has grown in popularity as a feasible solution to complex distributed information system where it is assumed that the computational components are autonomous in nature. The main characteristic of agent includes autonomy, co-operation, pro-activeness, social ability, mobility. Against this background a wide range of software engineering paradigms have been devised. Several agent oriented software engineering methodologies and different architectures have been proposed by different authors [11-12] [15 - 16]. Cooperation is the primary characteristic of multi agent system where the overall system exhibits significantly greater functionality than the individual components. The agents in MAS cooperate with each other in the system in some way to accomplish some goal. Such cooperation can be communicative in that the agents communicate by sending and receiving of signals/messages with each other in order to cooperate. The agent cooperation has received a considerable amount of attention in the MAS literature.

Performance is one of the important quality attributes that to be addressed along with the development of any software system. As the communication between the

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 41–48.
springerlink.com

agents plays a vital role in the performance, it is necessary to address the performance issues of cooperating agents. There are various approaches for predicting the performance of software systems in the early stages of software development. Software performance engineering (SPE) coined by Connie U Smith is a methodical quantitative approach for constructing software systems to meet the performance requirements. The advantages of SPE includes increased productivity, improved quality and effectiveness of the resulting software product, controlled cost of the sustaining hardware and software, enhanced productivity during implementation and testing. SPE methodology covers performance data gathering, quantitative analysis, prediction strategies, management of uncertainties, data presentation and tracking, model verification and validation, critical successes factors and performance design principles. SPE coined by Smith consists of two models called software execution model and system execution model. The software execution model is represented by the Execution Graph (EG). The data required for performance prediction are annotated in the execution graph. Solving the software execution model provides the data required for the machine model (system execution model) [7].

In this paper, we are proposing a methodology to quantify the cooperation among the agents in MAS with the help of execution graph by considering the cooperative characteristics of agents at the early stages of Software Development Life Cycle (SDLC). The remaining part of the paper is organized as follows: Section 2 presents a brief related work in the area of cooperation in MAS and SPE; the methodology for calculating Co-operative index using execution graph is explained in section 3; the proposed methodology is illustrated with a case study in section 4; the conclusion and future work are discussed in Section 5.

2 Related Work

Nicholos R Jennings, Katia Sycara and Michael Wooldridge discussed an overview of research and development activities in the field of autonomous agents and multi-agent systems in [19]. They aim to identify key concepts and applications, and to indicate how they relate to one-another. In their paper, some historical context to the field of agent-based computing is given, contemporary research directions are presented and a range of open issues and future challenges are highlighted. Agent cooperation is one of the well studied areas of MAS. Many classic theories are applied in the research of cooperation: logic theory, game and economic theory and Petri-net etc. A series of cooperation models are put forward based on these theories [10] [14] [18] [20]. As cooperation has attracted so many researchers, the current state of the art is captured in [4]. A cooperative game theory (CGT) for coalition formation in multi-agent systems is proposed in [3], where a novel model for the cooperative game has been used. The implementation of cooperation is enforced in terms of coalition formation and algorithms for their formation are discussed. A metric suite for the communication of MAS is suggested and a demonstration is provided to prove that a well balanced communication is related to high levels of Quality of Service measured

by response times in [13]. Software Performance Engineering (SPE) has evolved over the past years and has been demonstrated to be effective during the development of many large systems [8]. The extensions to SPE process and its associated models for assessing distributed object-systems are discussed in [6]. [3] Describes the use of SPE-ED, a performance-modeling tool that supports SPE process, for early life cycle performance evaluation of object–oriented systems. The transformation of Unified Modeling Language (UML) models into execution graphs and predicting the performance at the early stages of the software system is illustrated with a case study in [9]. Software performance engineering (SPE) has been established as a strong area of research in software engineering. Huge literature is available on SPE for distributed systems. A very little amount of research is carried out to assess the performance for MAS in the early stages of SDLC. Cooperative index can be used in the assessment of software performance [1]. Sequence diagram is used as a model to calculate the cooperative index in [1]. Various representations are used for modeling performance characteristics of software systems in the literature. However, initially, most of the researchers have used execution graph as the performance model for assessing the performance in the early stages of the software development. Hence, we propose to use the execution graph as the performance model to quantify the cooperative index between the agents. In this paper, a methodology is proposed to augment the communication between the agents in the execution graph and to compute the cooperative index from the execution graph.

3 Proposed Methodology

Cooperation between agents is defined as how effectively agents respond to the request of its neighbor agent. We have modeled the scenario of cooperation of agents with its neighbor by message passing. We have defined a term called Cooperative Index (CI) which quantifies the cooperation between the agents by considering the number of messages an agent received from the neighbor, the number of messages forwarded by the agent to the neighbor, the total number of messages generated by the agent for accomplishing its own task and the number of messages sent out of the total number of messages generated to accomplish the agents own task. We have formulated this model by considering the models developed in [21]. In the proposed methodology, the annotated Execution Graph is used to predict the Cooperative Index of agents in MAS. An execution graph is a graph, whose nodes represent one (or more than one) sets of actions (actors) and edges represent control transfer between them. According to [8], an EG node can be of *basic node*, *expanded node*, *repetition node*, *case node*, *pardo node* and *split node*. Each node is weighted by the demand-vector representing the resource usage by the node. We have devised a vector called as *communication vector* which is represented by 4-tuple to represent the message passing type between the agents. The 4-tuple is of the form (F, R, Sr, Sp) where “F” represents the number of messages forwarded by an agent to another agent, “R” represents the number of messages received by an agent, “Sr” represents the number of messages send by an agent to other agents for its own work to be processed, “Sp”

represents the messages for its self process. Depending upon the type of the request, values are assigned to the 4-tuple. If there is no action for any of the element in the 4-tuple a value "0" is assigned for the corresponding element, otherwise the name of the agent involved in that action is assigned to the corresponding element. A sample 4-tuple vector is given below:

RU (U, A1, A1, A1)
"RU" → Request from User
"U " → The message was forwarded by user
"A1" → A1 (agent) received a message for activity RU
"A1" → A1 (agent) generated a message for its own process and send to other agent
"A1" → A1 (agent) has self process.

We propose an algorithm to model the software specification with execution graph and annotate with communication vector.

```
while (Agent exists)
loop
  Initialize the elements of the 4-tuple
  {F = 0; R = 0; Sr = 0; Sp = 0}
  Get all Execution Graph
  while (Execution Graph Exits) loop
      do (for all the nodes in the Execution Graph)
          if (the value of the element in the 4-tuple is matched with agent)
then
            increment the corresponding element count by 1
        endif
       enddo
endwhile
```

4 Numerical Results

The MAS we have considered is based on supply chain management (SCM) systems [17]. Agents are autonomous and can operate in open electronic environments that are now becoming very popular which is the case with SCM. The proposed MAS include five agents namely Manager Agent (MA), Production Agent (PA), Inventory Agent (IA), Supply Agent (SA) and Delivery agent (DA) in a Supply Chain Management (SCM) system. While following their own goals, the agents work in cooperation in order to achieve the common ultimate goal to maximize the overall profit.

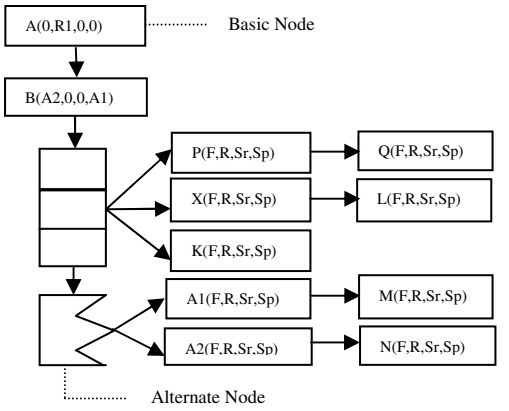


Fig. 1. Execution Graph with communication vector

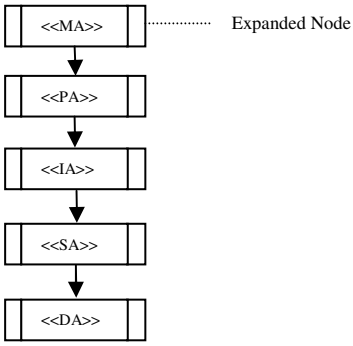


Fig. 2. Execution graph for Case Study

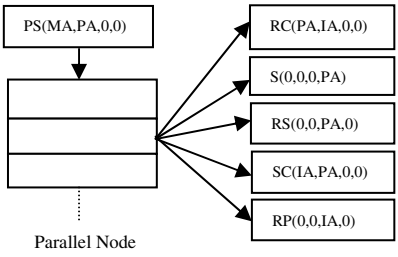


Fig. 3. Execution Graph for Production Agent

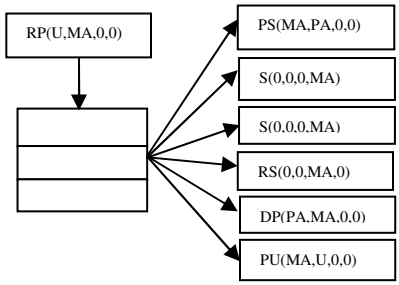


Fig. 4. Execution Graph for Manager Agent

Fig 1 represents a general model for execution graph with the communication vector for the agents in MAS. Fig 2 represents the execution graph in the high level view for the SCM system. Fig 3 represents the execution graph for the particular agent Manager and Fig 4 represents the execution graph for the production agent for a request from the user. Similarly we can construct the execution graph for other agents also. From the execution graph, the Co-operative index of each agent is calculated with the help of the equation 1.

$$\text{Co (i)} = \text{Ru} * \text{Wai} + (1\text{- Ru})* \text{Gai} \tag{1}$$

$$\text{Wa}_i = \begin{cases} w(i) & \text{if } Sa_i + Fa_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where} \quad w(i) = \frac{\sum_{j \in N_i} Fa_j^i + Ra_j^i}{Sa_i + Fa_i}$$

$$\text{Ga}_i = \begin{cases} g(i) & \text{if } \sum_{j \in N_i} Ta_i^j > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where} \quad g(i) = \frac{\sum_{j \in N_i} Sa_i^j}{\sum_{j \in N_i} Ta_i^j}$$

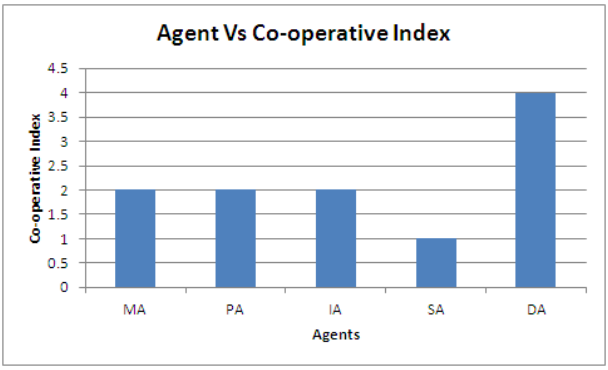


Fig. 5. Agent Vs Co-operative Index

Fig 5 represents the illustration of obtained co-operative index value using execution graph. The result shows that the Delivery agent (DA) is having the highest Co-operative index and Supplier agent (SA) has lowest Co-operative index. From the result we can infer that an agent who receives more messages on behalf of other agents is having high Co-operative Index.

The proposed methodology discusses the calculation of cooperative index from the execution graph. In [2] we have calculated the cooperative index of MAS with the help of UML sequence diagram for the same case study .The calculated Cooperative index value in that paper is same as the value obtained using the proposed methodology. But if we consider the construction of the software models, sequence diagram is easier to construct than execution graph. Moreover, it can be noted that execution graph is not a standard graph; but UML Sequence diagram is widely used by the industry people for modeling the software systems. However, the execution graph is used as the performance model in assessing the performance early in the SDLC [22] by that research group. Hence, we suggest depending on the familiarity with the model and purpose, the corresponding model can be used for calculating the cooperative index.

5 Conclusion and Future Work

One of the vital characteristics of MAS is Cooperation. The Cooperative index quantifies the cooperation between the interacting agents. In the performance prediction of MAS the Cooperative index plays an important role. Hence, in this paper we have proposed a methodology to calculate the cooperative index from the execution graph, which is a representation of performance model. The methodology is illustrated with a case study on MAS and the Cooperative index is obtained from the execution graph. The cooperative index value is compared with the one that is obtained using sequence diagram for the same case study. As the results are same, the cooperative index can be obtained from any of these models depending on the familiarity of the analyst. As the future direction, while assessing the performance of MAS the Cooperative index can be considered as one of the performance parameters.

References

1. Ajitha, S., Suresh Kumar, T.V., Geetha, D.E., Rajanikanth, K.: Early Performance Prediction of Co-operative Multi-Agent Systems. In: International Conference on Modeling Optimization and Computing, April 9-10. Elsevier Publications (2012)
2. Ajitha, S., Suresh Kumar, T.V., Rajanikanth, K.: A Quantitative Frame Work For Early Prediction of Cooperation in Multi agent systems. Technical Report, Department of MCA, MSRIT (2011)
3. Adel, G., Habib, R., Reza, M.: A Novel Algorithm for Coalition Formation in Multi-agent Systems using Cooperative Game Theory. In: Proceedings of the ICEE, May 11-13, IEEE (2010), doi:978-1-4244-6760-0
4. Abdellah, B., Ranjeev, M., Boukhtouta, A., Berger, J.: Distributed Intelligent Systems. Spring US, US (2009)
5. Smith, C.U., Williams, L.G.: Performance Engineering Evaluation of Object Oriented Systems with SPE-ED. In: Marie, R., Plateau, B., Calzarossa, M.C., Rubino, G.J. (eds.) TOOLS 1997. LNCS, vol. 1245, pp. 135–153. Springer, Heidelberg (1997)
6. Smith, C.U., Williams, L.G.: Performance Engineering Models of CORBA-based distributed-object systems. Performance Engineering Services and Software Engineering Research (1998)
7. Smith, C.U.: Performance Engineering of Software Systems. Addison-Wesley, Reading (1990)
8. Smith, C.U., Williams, L.G.: Performance Solutions: A Practical Guide to Creating Responsive, Scalable Software. Addison-Wesley, Boston (2002)
9. Geetha, D.E., Reddy, R.M., Suresh Kumar, T.V., Rajanikanth, K.: Performance Modelling and Evaluation of e-commerce Systems Using UML 2.0. In: Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (2007)
10. Hashel, E.A.: A Conceptual agent Cooperation Model for Multi-agent Systems' Team Formation Process. In: Third 2008 International Conference on Convergence and Hybrid Information Technology, pp. 12–20 (2008)
11. Giunchiglia, F., Mylopoulos, J., Perini, A.: The Tropos software development methodology: Processes, Models and Diagrams. In: Third International Workshop on Agent-oriented Software Engineering (July 2002)
12. Caire, G., Leal, F., Chainho, P., Evans, R., Garijo, F., Gomez, J., Pavon, J., Kearney, P., Stark, J., Massonet, P.: Agent oriented analysis using MESSAGE/UML
13. Gutierrez, C., Garcia, M.I.: A Metric Suite for the Communication of Multi-agent Systems. J. Phys. Agents 3(2), 7–15 (2009)
14. Kotb, Y.T., Beauchemin, S.S., Barron, J.L.: Petri Net-Based Cooperation in Multi-Agent Systems. In: Fourth Canadian Conference on Computer and Robot Vision (2007)
15. Padgham, L., Winikoff, M.: Prometheus: A methodology for developing intelligent agents. In: Third International Workshop on Agent-Oriented Software Engineering (July 2002)
16. Wooldridge, M.J., Weiß, G., Ciancarini, P. (eds.): AOSE 2001. LNCS, vol. 2222, pp. 101–108. Springer, Heidelberg (2002)
17. Moyaux, T., Chaib-draa, B., D'Amours, S.: Multiagent based Supply Chain Management. Springer, Heidelberg (2006)
18. Michael, W., Nicholas, R.J.: The Cooperative Problem-Solving Process. Journal of Logic and Computation 9(4), 563–592 (1999)
19. Nicholas, R.J., Katia, S., Michael, W.: A Roadmap of Agent Research and Development. Autonomous Agents and Multi-agent Systems 1(1), 7–38 (1998)

20. Semsar-Kazerooni, E., Khorasani, K.: A Game Theory Approach to Multi-Agent Team Cooperation. In: American Control Conference, USA, pp. 4512–4518 (June 2009)
21. Urpi, A., Bonucceli, M., Giordano, S.: Modelling Cooperation in Mobile Adhoc Networks:a Formal Description of Selfishness. In: Proceedings of the WiOpt 2003 (2003)
22. http://www.pergeng.com

QoS Multicast Routing Using Teaching Learning Based Optimization

Anima Naik[1], K. Parvathi[2], Suresh Chandra Satapathy[3], Ramanuja Nayak[1], and B.S. Panda[1]

[1] MITS, Rayagada, India
{animanaik,ramanuja.nayak}@gmail.com
bspanda@sify.com
[2] CUTM, Paralakhemundi
Kparvati16@gmail.com
[3] ANITS, Vishakapatnam
sureshsatapathy@ieee.org

Abstract. The QoS multicast routing problem is to find a multicast routing tree with minimal cost that can satisfy constraints such as bandwidth, delay. This problem is NP Complete. Hence, the problem is usually solved by heuristic or intelligence optimization. In this paper, we present a Teaching learning based optimization method to optimize the multicast tree. A fitness function is used to implement the constraints specified by the Quality of Service conditions. The experimental results dealt with relations between the number of nodes, edges in the input graph and convergence time, the optimal solution quality comparison with other evolutionary techniques. The results reveal that our algorithm performs better than the existing algorithms.

Keywords: Multicast, QoS, routing, teaching learning based optimization, fitness function.

1 Introduction

The rapid development in network multimedia technology enables more and more real-time multimedia services such as video conferencing, on-line games, distance education etc. to become mainstream internet activities. These services often require the network to provide multicast capabilities. Multicast refers to the delivery of packets from a single source to multiple destinations. The central problem of QoS routing is to set up a multicast tree that can satisfy certain QoS parameters. However, the problem of constructing a multicast tree under multiple constraints is NP-Complete [1]. Hence, the problem is usually solved by heuristic or intelligence optimization.

In QoS multicast routing, each node or link has some parameters associated with it. These parameters are used to determine the most efficient path from the source to the destinations. Thus, these network resources must be handled and shared in such a way that the most optimal solution can be obtained for the QoS multicast routing problem with minimal cost. This cost is determined by the parameter values associated with

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 49–55.
springerlink.com

each link which may be present in a chosen path from a source to a destination. Genetic Algorithms(GA) and Ant Colony Optimization(ACO)[2] have also been used to solve this problem. Particle Swarm Optimization(PSO)[3][4] technique is applied to solve the QoS multicast routing. Besides this other algorithm based on quantum mechanics named as Quantum-Behaved Particle Swarm Optimization (QPSO) was proposed [5]. Later on PSO along with Genetic Algorithm (GA) was introduced which become hybrid genetic algorithm and particle swarm Optimization (HGAPSO) [6] to solve multicast QoS routing. A tree based PSO has been proposed in [7] for optimizing the multicast tree directly. However, the performance depends on the number of particles generated. Another drawback of the algorithm is merging the multicast trees, eliminating directed circles and nested directed circles are very complex. In this paper we have use Teaching Learning based Optimization techniques for optimization of QoS multicast routing problems.

The TLBO method is based on the philosophy of teaching learning processes effect of the influence of a teacher on the output of learners in a class. Here, output is considered in terms of results or grades. The teacher is generally considered as a highly learned person who shares his or her knowledge with the learners. The quality of a teacher affects the outcome of the learners. It is obvious that a good teacher trains learners such that they can have better results in terms of their marks or grades. We have used this concept of algorithm for optimization of QoS multicast routing problems.

The rest of this paper is structured as follows. Section 2 discusses the Multicast Routing problem formulation in computer network. Section 3 presents the Concepts of Teaching learning based Optimization technique. Section 4 presents the TLBO algorithm for multicast routing. Section 5 details the experimentation carried out and presents the discovered results. The paper concludes with a discussion on the observations.

2 Multicast Routing Problem Formulation

A QoS multicast routing problem is usually involved in several constraints. In this paper, we simplify the QoS constraints and mainly focus on the band-width-delay constrained least cost multicast routing problem.

Communication network can be modeled as an undirected graph $G =< V, E >$, where V is the set of all nodes representing routers or switches, E is the set of all edges representing physical or logical connection between nodes. Each link $(x, y) \in E$ in G has three weights $(B(x, y), D(x, y), C(x, y))$ associated with it, in which positive real values $B(x, y)$, $D(x, y)$, $C(x, y)$denote the available bandwidth, the delay and the cost of the link respectively. Given a path $P(x, y)$ connected any two nodes *x*, *y* in *G*, it can be presumed that:

The delay of a path is the sum of the delays of the links (x , y) :

$$Delay\big(P(x,y)\big) = \sum_{(a,b)\in P(x,y)} D(a,b) \tag{1}$$

The available bandwidth of $P(x, y)$ is considered as the bottle neck bandwidth of $P(x, y)$:

$$Width\big(P(x,y)\big) = \min_{(a,b) \in P(x,y)} (B(a,b)) \tag{2}$$

In QoS transmission of real time multimedia service, the optimal cost routing problem with delay and bandwidth constrained can be described as follows: Given $G =< V, E >$, a source node s, and a multicast member set $M \subseteq V - \{s\}$, the problem is to find the multicast tree $T = (V_T, E_T)$ from source s to all destinations $v \in M$, where $T \subseteq G$, and T must satisfy the following conditions:

$$Cost(T) = \min \left(\sum_{(x,y)\in E_T} C(x,y)\right) \tag{3}$$

$$\sum_{(x,y)\in P_T(s,v)} D(x,y) \leq D_{max} \qquad \forall v \in M \tag{4}$$

$$Width(P_T(s,v)) \geq W_{min,} \qquad \forall v \in M \tag{5}$$

Where $P_T(s,v)$ is the set of links in the path from source nodes s to destination v in the multicast tree. Relation (3) means that the cost of multicast routing tree should be minimum. Relation (4) means that the delay requirement of QoS, in which D_{max} is the permitted maximum delay value of real time services. And relation (5) guarantees the bandwidth of communication traffic, in which W_{min} is the required minimum bandwidth of all applications.

3 Teaching Learning Based Optimization

This optimization method is based on the effect of the influence of a teacher on the output of learners in a class. Like other nature-inspired algorithms, TLBO [11] is also a population based method that uses a population of solutions to proceed to the global solution. A group of learners are considered as the population. In TLBO, different subjects offered to learners are considered as different design variables for the TLBO. The learning results of a learner is analogous to the 'fitness', as in other population-based optimization techniques. The teacher is considered as the best solution obtained so far.

There are two parts in TLBO: 'Teacher Phase' and 'Learner Phase'. The 'Teacher Phase' means learning from the teacher and the 'Learner Phase' means learning through the interaction between learners.

3.1 Teacher Phase

In our society the best learner is mimicked as a teacher. The teacher tries to disseminate knowledge among learners, which will in turn increase the knowledge level of the whole class and help learners to get good marks or grades. So a teacher increases the mean learning value of the class according to his or her capability i.e. say the teacher T_1 will try to move mean M_1 towards their own level according to his or her capability, thereby increasing the learners' level to a new mean M_2. Teacher T_1 will put maximum effort into teaching his or her students, but students will gain knowledge according to the quality of teaching delivered by a teacher and the quality

of students present in the class. The quality of the students is judged from the mean value of the population. Teacher T_1 puts effort in so as to increase the quality of the students from M_1 to M_2, at which stage the students require a new teacher, of superior quality than themselves, i.e. in this case the new teacher is T_2.

Let M_i be the mean and T_i be the teacher at any iteration$i$. T_i will try to move mean M_i towards its own level, so now the new mean will be T_i designated as M_{new}. The solution is updated according to the difference between the existing and the new mean given by

$$Difference\_mean_i = r_i(M_{new} - T_F M_i) \tag{6}$$

where T_F is a teaching factor that decides the value of mean to be changed, and r_i is a random number in the range [0, 1]. The value of T_F can be either 1 or 2, which is again a heuristic step and decided randomly with equal probability as

$$T_F = round[1 + rand(0,1) * (2 - 1)] \tag{7}$$

This difference modifies the existing solution according to the following expression

$$X_{new,i} = X_{old,i} + Difference\_mean_i \tag{8}$$

3.2 Learner Phase

Learners increase their knowledge by two different means: one through input from the teacher and the other through interaction between themselves. A learner interacts randomly with other learners with the help of group discussions, presentations, formal communications, etc. A learner learns something new if the other learner has more knowledge than him or her. Learner modification is expressed as

$For\ i = 1{:}P_n$
  Randomly select two learners X_i and X_j, where $i \neq j$
  $If\ f(X_i) < f(X_j)$ $X_{new,i} = X_{old,i} + r_i\ (X_i - X_j)$
  $Else$ $X_{new,i} = X_{old,i} + r_i\ (X_j - X_i)$
  $End\ If$
$End\ For$

Accept X_{new} if it gives a better function value.

4 The TLBO Algorithm for Multicast Routing

The TLBO-based algorithm for solving multicast routing are overviewed as follows. The process is initialized with a group of random learners (solutions). Then determine the best teacher among the learners which have the best fitness value. Subsequently in each time interval learners improved based upon teacher and themselves.

Initialization of individuals

To apply the TLBO method for multicast routing problems, we use the "individual" to replace the "learner". In the initialization process, a set of individuals is created at random. Individual $i's$ position at iteration 0 can be represented as the vector

$$x_i(0) = \{x_{i1}(0), x_{i2}(0), \ldots\ldots\ldots, x_{in}(0)\} \tag{9}$$

where n is the number of network nodes of individual i corresponds to the generation update quantity covering all network nodes.

Before starting the TLBO algorithm, we can remove all the links, which their bandwidth are less than the minimum of all required thresholds W_{min}. If in the refined graph, the source node and all the destination nodes are not in a connected sub-graph, this topology does not meet the bandwidth constraint. In this case, the source should negotiate with the related application to relax the bandwidth bound. On the other hand, if the source node and all the destination nodes are in a connected sub-graph, we will use this sub-graph as the network topology in our TLBO algorithm.

Evaluation Function Definition

We defined the evaluation function *f* given in Equation (10) as the evaluation value of each individual in population. The evaluation function f is a reciprocal of the performance criterion $Cost\ (T)$ in Equation (3).It implies the smaller $Cost(T)$ the value of individual *T*, the higher its evaluation value.

$$f(T) = \frac{\prod_{m\in M} \emptyset(s) \times \prod_{m\in M} \varphi(s)}{\sum_{e\in T} cost(e)} \tag{10}$$

$$\emptyset(s) = Delay(P(s,m) - D_{max}) \tag{11}$$

$$\varphi(s) = \begin{cases} 1, & W_{min} \le Width(P(s,m)) \\ \gamma, & W_{min} > Width(P(s,m)) \end{cases}. \tag{12}$$

Where $\varphi(s)$ is the penalty function, and the value of the $\gamma = 0.5$ in the paper.

5 Experimental Results

We have performed simulation to investigate the performances of multicast routing algorithms based on TLBO algorithm. The source and the destination are randomly generated. The bandwidth and delay of each link are uniformly distributed in range [10,50] and [0,50ms] respectively. The cost of each link is uniformly distributed in range [0,200].

To analyze the performance of the proposed algorithm, different sets of input were generated and the algorithm was tested for varying number of nodes and edges. In Table 1, we have compared the running times between genetic algorithm (GA), immune algorithm(IA), ant colony algorithm (ACO), Particle Swarm Optimization (PSO) and (Teaching learning based Optimization)TLBO algorithm for different combinations of node and edge.

Table 1. Comparison of running time (in s)

Nodes	Edges	TLBO	PSO	ACO	IA	GA
20	32	0.08	0.21	0.27	0.35	0.48
40	89	0.19	0.38	0.45	0.51	0.63
80	172	0.80	1.36	1.42	1.48	1.59
120	239	1.32	1.97	2.68	3.12	3.16
160	336	3.46	4.18	5.37	5.82	7.72
180	371	5.12	6.46	8.19	8.93	9.85
200	427	6.89	8.36	9.73	10.26	11.25

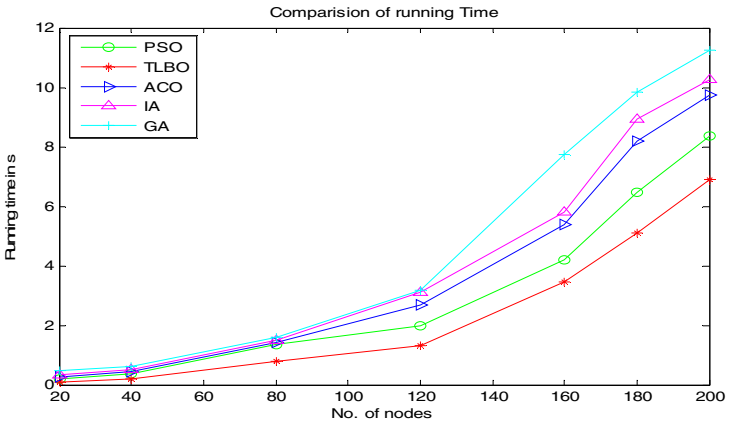


Fig. 1. Convergence behavior

Table 2. The optimal solution quality comparison

Algorithm	Optimal	Sub-optimal	Invalid
TLBO	89.4%	9.7%	0.9%
PSO	81.2%	17.5%	1.3%
GA	78.4%	19.4%	2.2%
IA	78.9%	19.6%	1.5%
ACO	79.9%	18.2%	1.9%

Table 1 results clearly show that the running time of TLBO algorithm grows very slowly with the size of the network and the running times of TLBO is smaller than GA's ,IA's and ACO's and PSO's . This behavior has been shown in the Fig. 1. Therefore, the proposed TLBO algorithm is very effective. Furthermore, for the same multicast routing, we made 300 simulations by TLBO algorithm against GA[8],immune algorithm(IA)[9], ACO algorithm [10] and PSO algorithm. The computation results are shown in Table 2.We can find that TLBO algorithm performances better than GA, IA, ACO, PSO. So our proposed PSO algorithm has good performance.

6 Conclusion

Multicast routing problem arises in many multimedia communication applications, computing the band-width-delay constrained least-cost multicast routing tree is a NP-complete problem. In this paper, a novel multicast routing algorithm based on the TLBO algorithms is proposed. The experimental results show that this algorithm has better performance and efficiency.

References

[1] Wang, Z., Crowcroft, J.: Quality of service for supporting multimedia application. IEEE Journal on Selected Areas in Communication 14, 1228–1234 (1996)
[2] Lhotská, L., Macaš, M., Burša, M.: PSO and ACO in Optimization Problems. In: Corchado, E., Yin, H., Botti, V., Fyfe, C. (eds.) IDEAL 2006. LNCS, vol. 4224, pp. 1390–1398. Springer, Heidelberg (2006)
[3] Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: Proceeding of the IEEE International Conference Neural Networks, pp. 1942–1948 (1995)
[4] Jin, X., Bai, L., Ji, Y., Sun, Y.: Probability Convergence based Particle Swarm Optimization for Multiple Constrained QoS Multicast Routing. Proceeding of the IEEE, 412–415 (2008)
[5] Sun, J., Liu, J., Xu, W.-b.: QPSO-Based QoS Multicast Routing Algorithm. In: Wang, T.-D., Li, X., Chen, S.-H., Wang, X., Abbass, H.A., Iba, H., Chen, G.-L., Yao, X. (eds.) SEAL 2006. LNCS, vol. 4247, pp. 261–268. Springer, Heidelberg (2006)
[6] Li, C., Cao, C., Li, Y., Yu, Y.: Hybrid of genetic algorithm and particle swarm optimization for multicast QoS routing. In: IEEE International Conference on Control and Automation, pp. 2355–2359 (2007)
[7] Wang, H., Meng, X., Li, S., Xu, H.: A tree-based particle swarm optimization for Multicast routing. Computer Networks 54, 2775–2786 (2010)
[8] Wang, Z., Shi, B.: Solution to Qos multicast routing problem based on heuristic genetic algorithm. Journal of Computer, China Computer Federation, 55–61 (January 2001)
[9] Liu, F., Feng, X.J.: Immune algorithm for multicast routing. Chinese Journal of Computer, 676–681 (June 2003)
[10] Carrillo, L., Marzo, J.-L., Fabrega, L., Vila, P., Guadall, C.: Ant colony behaviour as routing mechanism to provide quality of service. LNCS, pp. 418–419. Springer, Berlin (2004)
[11] Wang, Z., Sun, X., Zhang, D.: A PSO-Based Multicast Routing Algorithm. In: Third International Conference on Natural Computation, ICNC 2007 (2007), doi: 0-7695-2875-9/07 $25.00 © 2007
[12] Rao, R.V., Savsani, V.J., Vakharia, D.P.: Teaching–learning-based optimization: A novelmethod for constrained mechanical design optimization problems. Computer-Aided Design 43, 303–315 (2011)
[13] Satapathy, S.C., Naik, A.: Data Clustering Based on Teaching-Learning-Based Optimization. In: Panigrahi, B.K., Suganthan, P.N., Das, S., Satapathy, S.C. (eds.) SEMCCO 2011, Part II. LNCS, vol. 7077, pp. 148–156. Springer, Heidelberg (2011)

A New Privacy Preserving Measure: *p*-Sensitive, t-Closeness

Sowmyarani C.N.[1], G.N. Srinivasan[2], and Sukanya K.[1]

[1] Department of Computer Science and Engineering,
MSRIT, Bangalore
[2] Department of Information Science and Engineering,
RVCE, Bangalore
sowmyarani.cn@msrit.edu, gnsri@yahoo.com,
sukanya.k.prakash@gmail.com

Abstract. Preserving a sensitive data has become a great challenge in the area of research under data privacy. There are popular approaches such as k-anonymity, t-closeness [1] and l-diversity which are effective measures for preserving privacy. These techniques lead to solving many of the privacy issues. But all these measures suffer from one or the other types of attacks. To minimize these attacks, a new measure called p-sensitive, t-closeness is introduced. This measure will preserve the sensitive data by distributing different values of sensitive attribute according to t-closeness approach by introducing p-sensitivity, by minimizing attacks and improving the efficiency and utility of the data. This technique is termed as p-sensitive, t-closeness which satisfy p-sensitivity and t-closeness for a table by relaxing the threshold value t, so that; it will satisfy the p sensitivity to overcome many of limitations of previous approaches.

Keywords: k-anonymity, t-closeness, *p*-sensitivity, p-sensitive, t-closeness.

1 Introduction

There are popular approaches to preserve privacy[5] in the field of data privacy. But, many of the techniques have their own limitations. The proposed technique is *p*-sensitive, t-closeness, will overcome many of limitations of other popular techniques such as k-anonymity [2] and t-closeness. Our technique will overcome skewness attack and similarity attack, since there is sensitivity concept applied on t-closeness. This would be a great advantage to preserve sensitive attribute values. There is a need of better techniques which will overcome many types of attacks from which the present techniques are suffering and increase the utility of data. Increase in utility of data [4] is very important to achieve which prevents information loss, as researchers need the data to utilize for the analysis or different purpose in the field of research. *P*-sensitive, t-closeness will overcome limitations of many of techniques and minimize the data loss. It satisfies the condition for table in which the minimum p number of distinct values will be there for sensitive attribute under the specific threshold value t.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 57–62.
springerlink.com

2 Proposed Approach

The new measure states that, the table which satisfies t-closeness should also satisfies *p*-sensitivity. The records of table are distributed in such a way that, it should satisfy the p-sensitive property and threshold level should not go beyond relaxation limit t_{relax}. The threshold t can be relaxed up to t_{relax} under which it satisfies the *p*-sensitivity. In such a way that, utility of data should not decrease and correlation among the *Qi* group.

Definition1. Quasi-identifier [2] (Qi): Qi is a set of attributes in a table, which cannot identify individual by itself, but can identify individual by linking with external table.

Definition2. (*k-anonymity property*): The *k-anonymity property* for a anonymized data is satisfied if every combination of key attribute values in *anonymized data* occurs *k* or more times.

Definition3. (*p-sensitive, k-anonymity property*): The anonymized data satisfies *p-sensitive k-anonymity property* [3] if it satisfies *k*-anonymity, and for tuples in each *Qi* with the identical combination of key attribute values that exists in anonymized data, the number of distinct values for each sensitive attribute occurs at least *p* times within the same group.

Definition4. (*p-sensitive, t-closeness):* The table *T* satisfies *p*-sensitive, t-closeness property, if it satisfies t-closeness with the threshold value ranging from t to t_{relax}, and each *Qi*-group has at least *p* distinct categories of the sensitive attribute.

3 *P*-Sensitive, t-Closeness

The table1 shows the original data which can have 4 records in one equivalence class.

Table 1. Original data

Zip_code	Age	Salary	Disease
47977	21	360000	Heart Attack
47901	57	430000	Heart Attack
47982	47	380000	Diabetes
47904	45	590000	Diabetes
47609	34	143000	Brain Tumour
47605	21	600000	Bladder Cancer
47654	23	360000	Brain Tumour
47609	30	650000	Brain Tumour
47604	10	230000	Flu
47602	45	230000	Gastritis
47678	50	160000	Neck Pain
47903	21	467000	Neck Pain

Table 2 satisfies the *p*-sensitive, t-closeness for p=2 and t=0.2. The Zip_code, Age and Salary belong to set of Quasi-identifiers, *Qi.* Disease is considered as a sensitive attribute.

Table 2. Table satisfying 2-sensitive, 0.2-closeness

Zip_code	Age	Salary	Disease
47***	>20	6LPA	Brain Tumour
47***	>20	4LPA	Heart Attack
47***	>20	2LPA	Gastritis
47***	>20	3LPA	Heart Attack
47***	>20	6LPA	Bladder Cancer
47***	>20	5LPA	Diabetes
47***	>20	1LPA	Brain Tumour
47***	>20	3LPA	Diabetes
47***	>10	1LPA	Neck Pain
47***	>10	2LPA	Flu
47***	>10	4LPA	Neck Pain
47***	>10	3LPA	Brain Tumour

Table T2 satisfies 2-sensitive, 0.2-closeness by distributing salary values and satisfying 2 distinct values for sensitive attribute disease.

Algorithm 1. (Basic Algorithm to test the *p*-sensitive t-closeness property)

Input: anonymized data – a masked microdata. *p*, k ($p \leq k$) natural numbers greater than or equal to 2. Threshold value t is minimum 0.1 to maximum relaxed value 0.65

Output: Condition is true (*p*-sensitive k-anonymity [7] and t, 0.1<=t<= t_{relax}, where t_{relax} =0.65 is satisfied)
Condition is false (*p*-sensitive k-anonymity and t, 0.1<=t<= t_{relax}, where t_{relax} =0.65 is not satisfied)

```
if anonymized data  has k-anonymity property then
{
Condition = true;
for each combination of key attribute values and each confidential attribute do
{
Let x be the number of distinct values for that confidential attribute.
If ((x < p) && (t<=0.1&&t>=0.65)) then
{
Condition = false;
Break loop;
}
}
}
else Condition = false;
```

Fig. 1. Algorithm for *p*-sensitive, t-closeness

The main concept behind this technique is to combine the advantages of t-closeness and *p*-sensitive, k-anonymity. So that, it will overcome the skewness attack [4] and similarity attack [4] without losing the correlation among *Qi* attributes. The algorithm in figure1 tests that, if the table satisfies the *p*-sensitive, t-closeness, then it should satisfy the t-closeness and p-sensitive, k-anonymity for threshold value range between t to $t_{relax.}$

The quasi-identifier attribute values are generalized and suppressed within the same equivalence class to anonymize the data. In Table 2, the first equivalence class, the values of the sensitive attribute disease is having more than 2 distinct values satisfying 2-sensitivity as value of p>=2. There is always trade-off between the utility and threshold value. As the threshold value decreases, utility will also decreases.

As the t value relaxed to higher levels, utility will increase. But at the same level, the privacy [6] should also be preserved.

4 Experimental Set-Up and Results

A set of experiments were conducted for an test data consisting of 10000 tuples randomly selected from the Adult dataset from the UC Irvine Machine Learning Repository [10]. In all the experiments, we considered age, education_num, work class and sex as the set of quasi-identifier attributes; and occupation as the sensitive attribute. Micro data p-sensitive t-closeness was enforced in respect to the quasi-identifier consisting of all 6 quasi-identifier attributes and 1 occupation sensitive attribute. Although many values of k and p were considered, due to space limitations, we present in this paper only a small subset of the results.

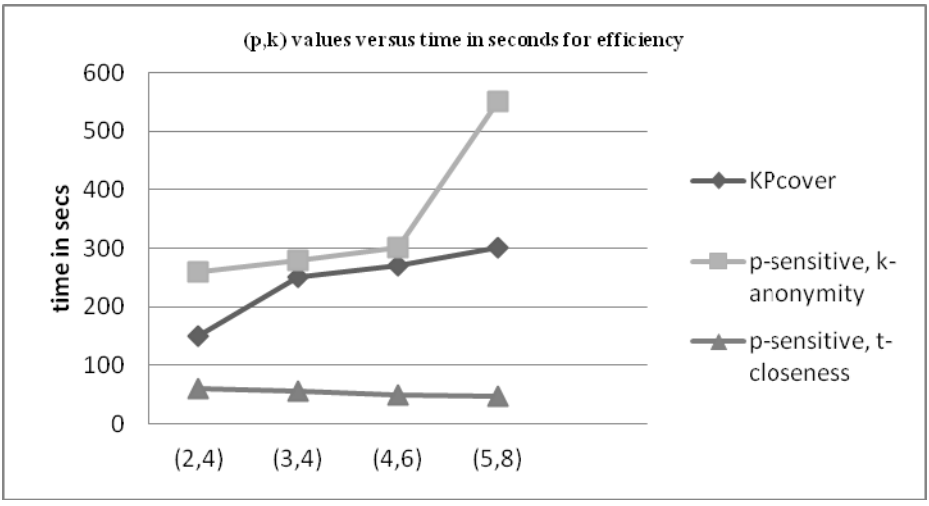


Fig. 2. Time efficiency for different to (p,k) values in terms of seconds

Figure 2 shows the resulting values for t value ranging from 0.45 to 0.65. Data sample is randomly chosen as 10000 tuples from adult dataset.

5 Comparison

Table 3. Comparison of privacy models v/s attack models

Privacy Models	Attack Models				
	Attribute disclosure	Skewness attack	Membership disclosure attack	Similarity attack	Data utility loss
k-anonymity	✓				
p-Sensitive, k-anonymity			✓		✓
t-closeness					✓

Table 3 shows comparison of privacy models with respect to their attacks. K-anonymity fails to protect attribute disclosure [4]. *p*-sensitive, k-anonymity undergoes membership disclosure and t-closeness will include data utility loss. Our approach is compared with other privacy models such as, k-anonymity, *p*-Sensitive k-anonymity, and t-closeness. The results shows there are no attacks and the data quality is improved by relaxing the threshold level up to $t_{\text{relax.}}$ to improve the data quality.

6 Conclusion

In this paper, the anonymization [8] is done using generalization [8] and suppression method. Most of the popular techniques will employ these methods. But, generalization and suppression will cause the numerical attributes generalizes to become categorical which is one of the disadvantage. The closeness cause the records distributed evenly throughout the table by impairing correlation between quasi-identifiers and sensitive attributes [9] by preserving privacy. The open challenge for many of the popular techniques such as, k-anonymity, t-closeness is, there is no specific computational approach to reach them for a specific data to be anonymized. The combined feature of applying p-sensitivity and threshold for achieving closeness, causes overcoming the skewness attack and similarity attack which will be an added advantage. The further work will be extending the concept of sensitivity to notice the trade-off between threshold and sensitivity in detail.

References

[1] Li, N., Li, T., Venkatasubramanian, S.: Closeness: A New Privacy Measure for Data Publishing. IEEE Trans. Knowl. Data Eng. 22(7), 943–956 (2010)
[2] Sweeney, L.: k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge based Systems, 557–570 (2002)
[3] Truta, T.M., Vinay, B.: Privacy protection: p-Sensitive k-anonymity property. In: Proceedings of the 22nd on Data Engineering Workshops. IEEE Computer Society, Washington, DC (2006)
[4] Domingo-Ferrer, J., Torra, V.: A Critique of k-Anonymity and Some of Its Enhancements, pp. 990–993 (2008)
[5] Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proc. SIGMOD 2000, pp. 439–450 (2000)
[6] Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., Theodoridis, Y.: State-of-the-art in privacy preserving dat mining. In: Proc. of ACM SIGMOD, pp. 50–57 (2004)
[7] Truta, T.M., Vinay, B.: Privacy protection: p-sensitive k-anonymity property. In: 2nd International Workshop on Privacy Data Management PDM 2006, p. 94. IEEE Computer Society, Berlin (2006)
[8] Vijayarani, S., Tamilarasi, A., Sampoorna, M.: Analysis of Privacy Preserving K-Anonymity Methods and Techniques. In: Proceedings of the International Conference on Communication and Computational Intelligence 2010, Kongu Engineering College, Perundurai, Erode, T.N.,India, December 27-29, pp. 540–545 (2010)
[9] Wu, Y., Ruan, X., Liao, S., Wang, X.: P-Cover K-anonymity model for Protecting Multiple Sensitive Attributes. In: The 5th International Conference on Computer Science & Education, Hefei, China, August 24-27 (2010)
[10] Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: UCI Repository of Machine Learning Databases, UC Irvine (1998), http://www.ics.uci.edu/~mlearn/MLRepository.html

Indic Language Translation in CLIR Using Virtual Keyboard

Mallamma V. Reddy and M. Hanumanthappa

Bangalore University,
Psychology Building,
Jnanabharathi Campus Bangalore
Mallamma_vreddy@bub.ernet.in
hanu5672@bub.ernet.in

Abstract. Natural Language is the primary medium for human communication. Function of a Natural Language (NL) is to communicate semantic content of its expressions directly to unravel this task we used the notion of Virtual Keyboards. Virtual Keyboards are commonly used as an on-screen input method in devices with no physical keyboard. This paper presents the usage of Virtual Keyboards the aim is to build a machine that could communicate in Natural Language particularly for Kannada to Telugu and vice-versa. It is designed using Unicode Character Set "UTF-8", achieves Translation of characters, words and sentences from source language text to target language. We have built the Bilingual Dictionary for Kannada-Telugu which consists around 1000 words. The experimental results and their performance are analyzed using Precision and Recall as described in this paper.

1 Introduction

There is a dramatic increase in the quantum of knowledge and information resulting in increase in the production of books and other multimedia communication materials including Compact Discs - Read Only Memory (CD-ROM). These repositories of knowledge are the bridges between information generators and the information users. The success of such a repository is completely dependent upon how tactfully the recorded knowledge is well organized and retrieved. To do this CLIR [1] is used.

Cross Language Information Retrieval (CLIR) is the retrieval of relevant information for a query expressed in native language. While retrieval of relevant documents is slightly easier, analyzing the relevance of the retrieved documents and the presentation of the results to the users are non-trivial tasks. To accomplish this task, we present our Kannada Telugu and Telugu Kannada CLIR systems as part of Ad-Hoc Bilingual task. We take a query translation based approach using bi-lingual dictionaries [2] shown in Fig 1. Both Kannada and Telugu use the "UTF-8" / western windows encode and draw their vocabulary mainly from Sanskrit.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 63–67.
springerlink.com

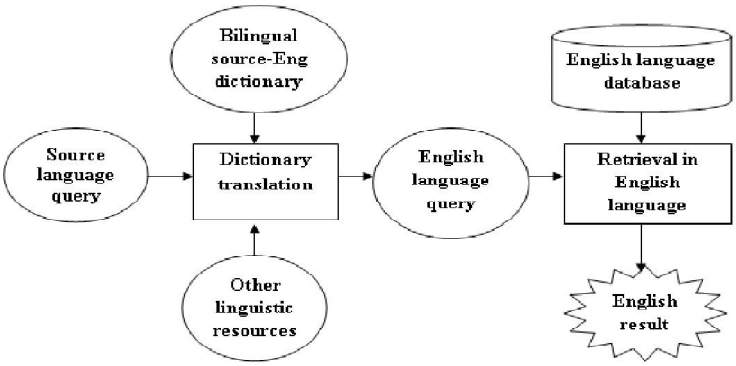


Fig. 1. Dictionary Based method for Query Translation

1.1 Machine Transliteration

The Language transliteration is one of the important area in natural language processing. Transliteration is mapping of pronunciation and articulation of words written in one script into another script. Transliteration should not be confused with translation, which involves a change in language while preserving meaning. There are many standard formats possible for Devanagari English transliteration viz. ITRANS, IAST, ISO 15919, etc. but they all use small and capital letters, and diacritic characters to distinguish letters uniquely and do not give the actual English word found in the corpus.

1.2 Machine Translation

Machine translation[3], sometimes referred to by the abbreviation MT (not to be confused with computer-aided translation, machine-aided human translation MAHT and interactive translation) is a sub-field of computational linguistics that investigates the use of computer software. Machine translation is the process of translating from source language text into the target language. There are many challenges to face when attempt to do machine translation.

- Not all the words in one language have equivalent words in another language
- Two given languages may have completely different structures.
- Translation requires not only vocabulary and grammar but also past knowledge. These transliteration and Translation can be done by using virtual keyboards which is more easy to use and provides graphical user Interface.

2 Virtual Keyboards for Kannada/Telugu

Kannada or Canarese is one of the 1652 mother tongues spoken in India. Forty three million people use it as their mother tongue. Kannada has 44 speech sounds. Among them 35 are consonants and 9 are vowels. The vowels are further classified into short vowels, long vowels and diphthongs. It is also one of the 18 Scheduled

Languages included in the VIII Schedule of the Constitution of India is recognized as the Official and Administrative language of the state of Karnataka [4]. It belongs to the Dravidian family of languages. Within Dravidian, it belongs to the South Dravidian group. The Dravidian languages stand apart from other family of Indian languages like Indo Aryan, Sino Tibetan and Austro Asiatic by having distinctive structural differences at phonological, morphological, lexical, syntactic and semantic levels.

Telugu (తెలుగు) is a Central Dravidian language primarily spoken in the state of Andhra Pradesh, India, where it is an official language [5]. According to the 2001 Census of India, Telugu is the language with the third largest number of native speakers in India (74 million), 13th in the Ethnologue list of most-spoken languages worldwide, and most spoken Dravidian language. We have designed The Virtual keyboard for Kannada/Telugu as shown in Fig 2. (a) and (b).

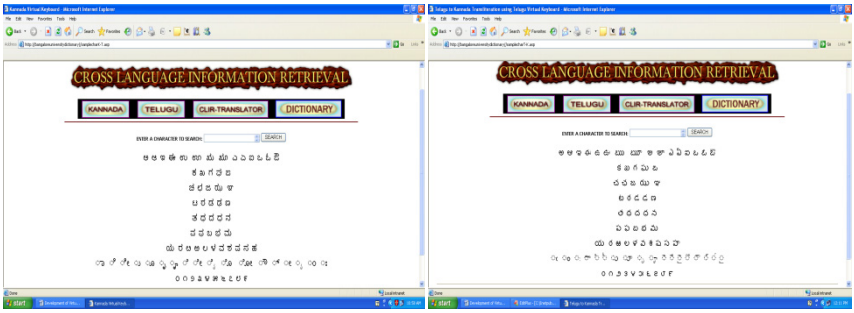


Fig. 2. (a). Kannada Virtual Keyboard (b). Telugu Virtual Keyboard

3 Experimental Setup

Several corpora were collected to estimate the parameters of the proposed models and to evaluate the performance of the proposed approach. The corpus BUBShabdaSagara-2011 for training consisted of 1000 words in dictionary for Kannada/Telugu. The training corpus composed of a bilingual word list. In the experiment, the performance of word translation extraction was evaluated based on precision and recall rates at the word. Since, we considered exactly one word in the source language and one translation in the target language at a time.

The word level recall and precision rates were defined as follows:

$$Word\ Precision(WP) = \frac{number\ of\ correctly\ extracted\ word}{number\ of\ extracted\ \ words} \tag{1}$$

$$Word\ Recall(WR) = \frac{number\ of\ correctly\ extracted\ Words}{number\ of\ correct\ words} \tag{2}$$

4 Results and Performance

Cross Language Information Retrieval Tool [6] is built by using the ASP.NET as front end and Database as back end, the Kannada/Telugu is encrypted by using the "UTF-8" [7] /Encoding system. Telugu and Kannada and vice-versa are the source language and the target language, respectively, in our query translation. All the experiments carried out here involve the same set of Kannada/Telugu queries and the same query expansion [8], translation and retrieval method. The only difference between the experimental conditions is in what dictionaries are used in the query translation. We have trained the systems with corpus size of 200, 500 and 1000 lexicons and sentences respectively. Performances of the systems were evaluated with the same set of 500 distinguished sentences/Phases that were out of corpus. The experiment results as shown in Fig 3 and Fig 4.

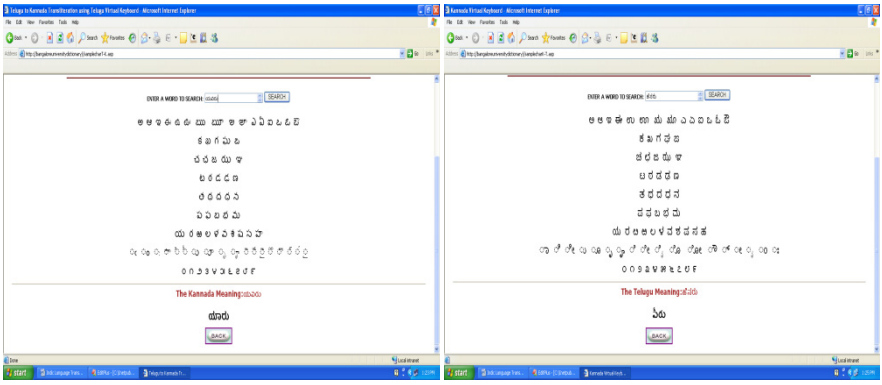


Fig. 3. Sample Results for Word

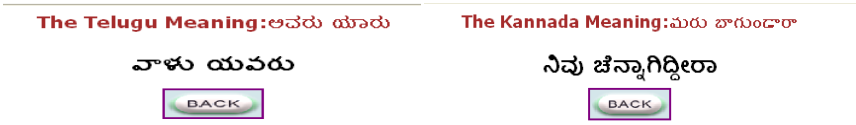


Fig. 4. Sample Results for Sentences

From the experiment we found that the performances of our systems are significantly well and achieves very competitive accuracy by increasing the corpus size as shown in Fig 5.

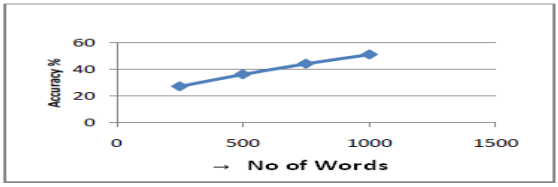


Fig. 5. Performance Graph

5 Conclusion

Dictionary-based query translation has been widely used in CLIR because of its simplicity and the increasing availability of machine-readable bilingual lexicons. We have presented our Kannada→Telugu and Telugu→Kannada CLIR system developed for the Ad-Hoc bilingual Task. Our approach is based on query Translation using bilingual dictionaries. We focused on translation of lexis and sentences, which has been demonstrated to be one of most effective ways to obtain more accurate translations. The implementation results are shown above with evaluation metric. The proposed method can be easily extended to other language pairs that have different sound systems without the assistance of pronunciation dictionaries. Further paragraph translation is to be carried out as part of Ad-Hoc bilingual task for Kannada to Telugu, Telugu to Kannada.

Acknowledgments. The support of the University Grant Commission (UGC), India, is gratefully acknowledged. The author would also like to acknowledge the help and support from Dr. M. Hanumanthappa, for his valuable guidance and suggestions which helped me a lot to write this paper. This paper is in continuation of the major research project entitled ***Cross-Language Information Retrieval*** sanctioned to Dr. M. Hanumanthappa, PI-UGC-MH, Department of Computer Science and Applications by UGC carried out at the Bangalore University, Bangalore, India.

References

1. Daelemans, W., Sima'an, K., Veenstra, J., Zavrel, J.(eds): Different Approaches to Cross Language information Retrieval, number 37 in Language and Computers: Studies in Practical Linguistics, Amsterdam, Rodopi (2001)
2. Reddy, M.V., Hanumanthappa, M.: Dictionary Based Word Translation in CLIR Us-ing Cohesion Method. In: Proceedings of the 6th National Conference; INDIACom 2012 Computing For Nation Development, February 23-24, Bharati Vidyapeeth's Institute of Computer Applications and Management, New Delhi (2012)
3. Homiedan, A.H.: Machine Translation
4. The Karnataka Official Language Act". Official website of Department of Parliamentary Affairs and Legislation. Government of Karnataka (2007) (retrieved June 29, 2007)
5. http://en.wikipedia.org/wiki/Telugu_language
6. Reddy, M.V., Hanumanthappa, M.: CLIR Project (English to Kannada and Telugu), http://bangaloreuniversitydictionary-menu.asp
7. http://www.ssec.wisc.edu/~tomw/java/unicode.html#x0C80
8. Pingali, P., Varma, V.: Hindi and Telugu to English Cross Language Information Retrieval at CLEF 2006. In: Working Notes for the CLEF 2006 Workshop (Cross Language Adhoc Task), Alicante, Spain, September 20-22 (2006)

Energy Efficient Clustering and Grid Based Routing in Wireless Sensor Networks

Amrutha K.M., Ashwini P., Divyashree K. Raj,
Kavitha Rani G., and Monica R. Mundada

Department of CSE,
MSRIT, Bangalore
monica_mundada@yahoo.co.in

Abstract. A wireless sensor network is a deployment of massive numbers of small, inexpensive, self powered devices that can sense, compute and communicate with other devices for the purpose of gathering local information to make global decisions about a physical environment. Wireless Sensor Networks are tightly constrained in terms of transmission power, onboard energy, processing capacity and storage, and thus require careful resource management. We mainly have LEACH(Low Energy Adaptive Clustering Hierarchy) and PEGASIS(Power Efficient Gathering in Sensor Information Systems) protocols. But LEACH doesn't support distribution of cluster heads optimally and PEGASIS has the overhead of data delay. This proposed approach emphasizes on increasing network lifetime via efficient routing path ands , efficient energy consumption.

Keywords: Wireless Sensor Networks- WSN, Cluster Head- CH, Base Station- BS

1 Introduction

This approach emphasizes on increasing the life-span of the network by optimum selection and distribution of the Cluster Heads (CH), thus increasing energy efficiency of the network**.**

2 Assumptions and Dependencies

- Base Station (BS) is immobile.
- All sensor nodes in the network are energy constrained.
- Nodes in the network are not dynamic while the CHs are being selected.
- All the nodes are present in the X-Y co-ordinate system with left bottom corner as the origin.
- All sensor nodes are scattered within the range.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 69–74.
springerlink.com

3 General Description

Wireless Sensor Networks are tightly constrained in terms of transmission power, onboard energy, processing capacity and storage, and thus require careful resource management.

The proposed approach emphasizes on energy efficiency, network life-time, optimum selection of the cluster heads and providing security.

Initially the sensor nodes send a Hello packet to the BS, which comprises of the position of the node, transmission time and the residual energy using CSMA/CA protocol.fig 1 shows the initial network considered.

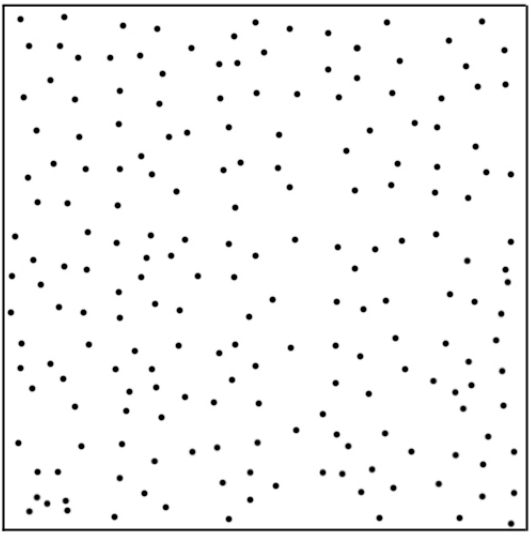


Fig. 1.

The BS gets the maximum and minimum values of X and Y according to the position of the sensor nodes using which a virtual square is drawn. The virtual square is partitioned into grids in 5X5 order as shown in fig 2. In response to the Hello packet sent by the sensor nodes, the BS will provide the grid-ID based on the co-ordinates of the respective grid.

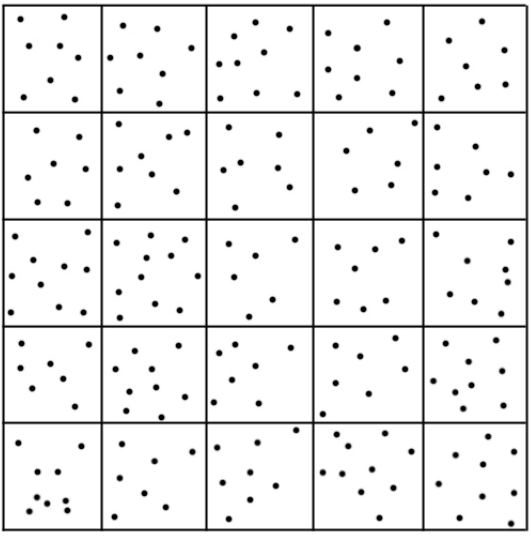


Fig. 2.

And only one CH is selected for every grid, i.e., the one which has the least transmission time.

The BS sends a message to the CH which includes the following fields, Source, Destination, MAC addresses of all its subscribed nodes and the neighboring CH-ID. The CH multicasts an advertising message to all its subscribed nodes .The subscribed nodes will save the MAC address of its CH for secure data transmissions.

Compute the value 'k', which is equal to total number of nodes divided by number of grids formed. If the number of nodes, say p in each zone <<k, then assign the sensor nodes to the cluster in neighboring grids. As each node maintains the address of all its neighboring nodes, it checks for all its neighbors one by one according to the grid IDs', and if found averagely densed, then the nodes are added to it until it reaches value k. Repeat this until all the p nodes are distributed to its neighbors.

Now find the position (side) of the BS with respect to the virtual square, such that the CHs in the row nearest to that, communicates with it.

When the BS receives query from the user, it will forward to the CHs' in the nearest row/column based on its position, and from which it sends to its neighboring CHs' along the same line as shown in fig(c) forming 5 virtual linear chains in the network.

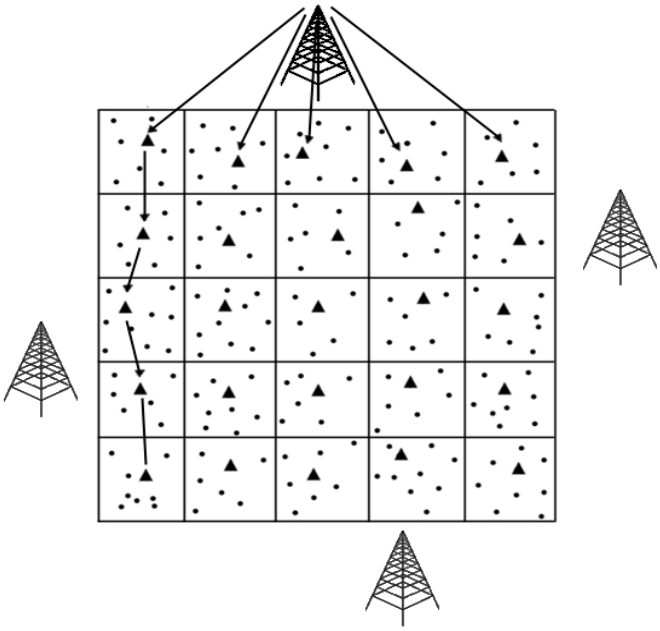


Fig. 3.

Now the CH transmits the query to its subscribed nodes using the combination of FDMA and TDMA protocols. In each allotted slot, FDMA is used for transmission and reception of data.

Based on the query the data is aggregated in each cluster, and is forwarded along the same virtual chain via CHs' to the CH which actually communicates with the BS.

i.e, The last CH in each zone refers the MAC address in the routing table for the CH from which it received the query, to send the aggregated data. This is repeated until all the aggregated data from different CH nodes reaches the CH which communicates with BS and hence sends the data to BS.

And finally the CHs' nearest to the BS will send the aggregated from all the CHs' in its row (if BS is on left or right side of the square (network)) or column (if BS is above or below the square (network)) as shown in fig 4 to the BS. From which the data is processed and sent back to the user.

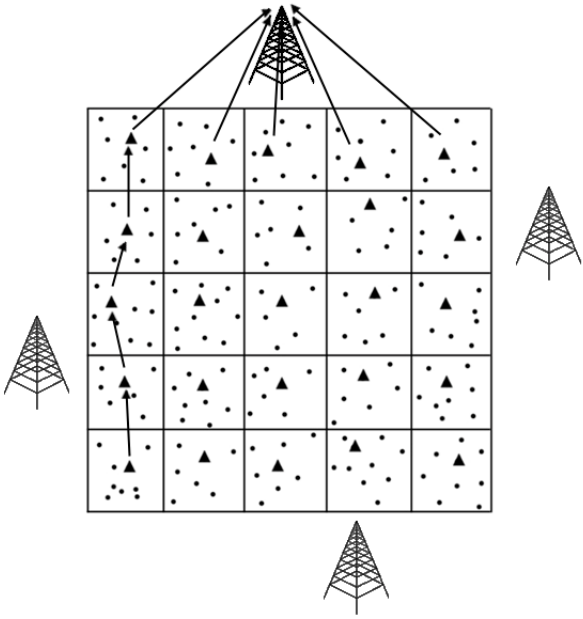


Fig. 4.

In the second round, the node within the same CH with the next least transmission time and highest residual energy is selected as CH and notified. This step is continued until all the subscribed nodes in the cluster become CH.

4 Proposed Algorithm

Step 1: Initially, send the HELLO packets to the BS, containing position of all the nodes in the network along with their energy level. Determine the minimum and maximum values of X and Y based on the position of the sensor nodes and plot the virtual square, which is being partitioned into 5X5 grids.

Step 2: A node with least transmission time in each grid is selected as CH and then a virtual chain is formed along the CHs' in the same line based on the position of BS.

Step 3: Once the network is divided into 5X5 grids, transmit the query from BS to all the CHs' in the nearest row/column, from which it is propagated to all other CHs' along the virtual chain. Now the data is aggregated in each grid based on the query and are sent in the reverse order of its arrival until the initial CH collects all the aggregated data from other CHs' along the virtual chain, which communicates and hence sends data to the BS.

Step 4: For the next coming rounds, select the CH from same cluster which has the next least transmission time and highest residual energy.

Goto Step 3.

5 Future Enhancements

For very large networks, data accuracy can be provided by using a cache memory in the nodes. When the BS receives the data, it checks the accuracy using CRC. If the data is accurate the BS sends an ACK, else NAK is sent to the CH to resend the data present in the cache memory.

References

1. Li, Q., Aslam, J., Rus, D.: Hierarchical Power-Aware Routing in Sensor Networks. In: Proc. DIMACS Wksp. Pervasive Net. (May 2001)
2. Xu, Y., Heidemann, J., Estrin, D.: Geographyinformed Energy Conservation for Ad-hoc Routing. In: Proc. 7th Annual ACM/IEEE Int'l. Conf. Mobile Comp. and Net., pp. 70–84 (2001)
3. Lindsey, S., Raghavendra, C.: Power-Efficient Gathering in Sensor Information Systems. In: Lindsey, S., Raghavendra, C. (eds.) IEEE Aerospace Conf. Proc. 3(9-16), pp. 1125–1130 (2002)
4. Manjeshwar, A., Agarwal, D.P.: Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks. In: 1st Int'l. Wksp. on Parallel and Distrib. Comp. Issues in Wireless Networks and Mobile Comp. (April 2001)
5. Park, S., Savvides, A., Srivastava, M.B.: SensorSim: A Simulation Framework for Sensor Networks. In: Proceedings of the 3rd ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems, Boston, MA (2000)
6. Park, S., Savvides, A., Srivastava, M.B.: Simulating Networks of Wireless Sensors. In: Proceedings of the 2001 Winter Simulation Conference (2001)
7. The Network Simulator – ns-2, `http://www.isi.edu/nsnam/ns/Hung-ying` Tyan, Design, Realization and Evaluation of a Component-based Compositional Software Architecture for Network Simulation, PhD thesis, Department of Electrical Engineering, The Ohio State University (2002)

A Comparative User-Centric Study of Digital Library Software Systems

Samaneh Ahmadi[1], Shiva Shirdavani[1], and Srini Ramaswamy[2]

[1] Esfahan University, Iran
samanehahmadi71@yahoo.com
Shiva.shirdavani@gmail.com
[2] ABB India Corporate Research, Bangalore
srini@ieee.org

Abstract. User-centric services and service provisioning has been a strong motivator for modern internet technologies, and, in particular, Web 2.0 technologies. Of particular interest to this study is its ability to have widespread societal impact. Hence in this work we assess the impact of such advancements in how one widespread societal function works, i.e. libraries; and we study digital libraries and their associated software systems that are being critically impacted by such technologies. Thus the objective of this paper is to derive several overarching standards and criteria for building user-centric digital libraries from both a technical (structure) and utilitarianism dimensions and explore two different Digital Libraries (Dspace Software - Khazar University and Green stone Software - New Zealand Digital library) with their corresponding software systems. Using quantitative methods of comparison we conclude, that while both Digital libraries and their associated software systems have some advantages for their users, they also have several disadvantages.

Keywords: Digital Library, Digital Library Software, Green Stone, Dspace.

1 Introduction

The two principal building blocks that are used in the improvement of digital libraries (DLs) are the Internet and the World Wide Web. According to statistical reports the World Wide Web (or the web) has been one of the great successes in the history of computing and this connected collection of information across the Internet around the world is a significant step forward for mankind. The web and its associated technologies have led to fundamental and rapid development of DLs. In order to manage and control huge collections of artefacts over long periods of time, libraries are often early adaptors new technologies; this includes microfilm, through online information services, and CD-ROMs, etc. The internet and web are thus pervasive technologies used by existing and new libraries to make immense societal impact on the current and future generations of humankind; for example, social search engines such as Aardvark [1,2] are an attempt to establish trust with their clients by concentrating on question and answering techniques. This study focuses on defining

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 75–83.
springerlink.com

characteristics for such a user-centric digital library; as a case study we explore a quantitative evaluation method for comparing two digital library efforts with associated software systems.

2 Background

Several publishers and libraries became involved in building online collections of scientific journals in the late 1980s. According to mercury technical reports series [3] the Mercury Electronic Library project at Carnegie Mellon University between 1987 and 1993 was one of the first steps to create a campus digital library. The advanced computing infrastructure at Carnegie Mellon was built by Mercury that consisted of a high-performance network, a fine computer science department, and the norm of innovation for university libraries. Mercury and CORE cooperated further on numerous other projects to use scanned images of journal articles.

Projects like Mercury, CORE, and Tulip were not envisioned as a long production system.[4] Each of them had technically limitations, for example, they suffered from the small size of the collections provided to researchers. The arrival of the web and the common availability of web browsers went a long way towards solving the problem of user interface development. Though web browsers were a good start, they were not ideal for a digital library; however, they have the huge advantage that they can be on all standard computers and operating systems, thereby addressing the availability issue from a user-centric perspective. In the mid-1990s, publishers' began establishing avenues to provide scientific journals online. When commercial publishing on CD-ROM had increased as a vital part of the industry, few journals were accessible online. Thus users of online information services were expanding faster than printed material and most of the major publishers of scientific journals have got along rapidly to electronic publishing, such as the approach taken by the Association for Computing Machinery (ACM) and the IEEE; as well as the large commercial publishers including Elsevier, John Wiley, and Academic Press, and by societies such as the American Chemical Society, and by university presses, including M.I.T. Press. [5]

3 Digital Library Definition

In literature once can fund find different definitions for DLs and these are named in a variety of ways by its various stakeholders. According to Adams and Blandford [6], librarians refer to them as databases and people in arts and humanities name them electronic archives. In the UK and Western Europe, DLs are referred to as digital surrogates, being regarded as substitutes for traditional libraries in their role of collections of validated and structured information[7]. Leiner [8] stated a DL is a collection of services and information objects that are available digitally. Information objects can be defined as anything in a digital format such as books, journal articles and sounds, since DLs organize and present information objects to users, and support them in dealing with these objects. In Brown [9] the authors treat DLs as an archive, and explore some of the challenges associated in developing smaller digital archives

for internal use, specifically from an education and pedagogical perspective. Lang [10] described that the goal of a DL is to improve access for all users. However, DL is a widely accepted term describing the use of digital technologies to acquire, store, preserve and provide access to information and material originally published in digital form or digitized from existing print, audio-visual and other forms.

Additional to above goal for DLs, new DLs should address another important purpose, which is related to their users (customers). Today DLs are typically centred around specific user communities who are mainly interested in these services; hence they have been built to meet the discipline-specific needs, without much rationale for fixed concepts on how these collections are typically managed. While creators and users improve and grow; the systems need to address overarching issues that relate to organizing information, retrieving it, establishing quality control, standards, and services. Such an approach will allow us to find new and creative answers to these age-old questions in providing meaningful information to the wider society.

4 Digital Library Evaluation Criteria

In literature, researchers have identified various criteria to evaluate DLs. Such evaluation criteria have been adopted to address needs arising from characteristics such as library type, library purpose [9], library mission, library patrons, etc. For example across the various literatures we surveyed, the following have been used as criteria in order to evaluate DLs.

4.1 Collection

Wealth and diversity of resources, how much the collection of library are capable to meet its patrons' needs and fulfill information needs of target users.

4.2 Information Organization and Representation

Which metadata, subject scheme or classification system the library uses. How do these tools help to facilitate of information search and retrieval library materials?

4.3 Technology

What is the fundamental technology? Is it a database, commercial system or open source software? Does the library offer any documentation on the technology or architectural form they have working to build the library?

4.4 Interface Features

Consists of such issues as browsing, searching, navigation, results display etc. Some of the main considerations in this connection are:

- Option of basic and advanced search modes
- easy to use is the interface
- easy to learn is the interface

- What retrieval performances have been provided?
- How customizable are consequence displays? (e.g. brief vs full description of the item)
- Does the interface offer any explanation or customization facilities such as book marking, saving search history, creating or modifying objects inside the library?
- How are fonts, images, icons and button labels designed?
- Document declaration condition (regarding to images and digitized recourses)
- Audio and video information search and retrieval situation (regarding to multimedia DLs)
- Does the interface track usability standards and principles?

4.5 Standards

Does the library use some standards relating to metadata, interoperability, accessibility and knowledge organization?

4.6 Documentation

Does the library offer any information on its policies and guidelines considering target users, collection improvement, digitization and preservation, standards and tools used to support and preserve the library?

Libraries are one of main information centers in societies. According to some authors the rapid growth of World Wide Web and its technologies and tools have lead to a reduced role of libraries to some extent. The internet and related informational websites have thus created a competitive environment for libraries. To sustain, libraries should try to play their fundamental role as a useful information center for their patrons. One of the means by which libraries can still differentiate and maintain their competitive advantage for their patrons is by making available reliable and accurate information than what is available on the internet. Leveraging and utilizing 2.0 technologies is a current and recent trend by libraries. Given the public's acceptance of the internet, implementing DLs from a user-centric perspective can still enhance and revitalize the role libraries in modern society.

In general, library performs three main tasks, which consist of collecting, cataloging, and retrieving resources that are of importance to their users. Certainly there are other additional tasks that were mentioned in our literature survey, such as library management, library preservation, and so on. Our primary consideration in selecting the following special categories and sub-categories discussed below is to cater to the primary library tasks. From this perspective, we select particular elements that are of importance to a user-centric library system. Using these insights, we then analyze some recent DLs and associated software system (Greenstone and Dspace). Hence through this study we have derived certain categories and sub- categories that can enhance the user centric nature of library and compare and contrast two competing DLs and associated software systems to draw some insights.

To conduct this effort, we adopted a combined framework to determine various evaluation criteria (some identified by pervious research works) and evaluate their applicability to open-source library software systems. Using Greenstone and Dspace as examples we then explore the purpose of these libraries and in particular the main mission for any digital library is to be user-centric for its users. Our framework is categorized into four sub- characteristic categories, and each sub-category includes certain characteristics that are deemed important for successfully developing such user-centric DL systems. The major categories we propose here include the following:

4.6.1 General Information

The purpose of this category is to offer general information about the software. In fact, instead of long descriptions about the software and their history / background, readers need to be able to gain such information quickly. The general information category is divided to four sub-categories such as About, URL, membership, license, language support and use. Actually general information can enumerate a kind of library policy and development strategy in the case of DLs software systems.

4.6.2 Design and Customization

Just like storefronts in a mall /shopping lane, design and customization of DLs play the role of 'buy-in' from their clients. Here we identify the role of design and customization on DLs that enhance user experience; these include: stable release, search modes, developer / development status, type, program language used, and protocol supported. Indeed library building, ordnance of building, classification of materials that is related to search mode, special facilities and services for members and patrons in a classic library and so on are important to attract and satisfy users are also important for DLs.

4.6.3 Security and Access

Security and information privacy are the cornerstone of issues that helps sites gain trust of their users. DLs must endure that their patrons are assured of security and privacy, topics central to their adoption. One of the issues related to security of DLs are copyrights. Other sub-categories of interest there include access, file upload, and privacy policy.

4.6.4 Collaborative Tools

As mentioned earlier, the main difference between web2.0 and web1.0 is related to communication. Web2.0 enhances communication among users. To address this gap, some employ social software tools to support user collaboration. Most websites use at least one or more such tools to connect with their users. So the use of collaborative tools in DLs such as blog, email, FAQ, training materials (inform of catalogue or brochure files and social medias) are all key aspects of a user-centric DL system.

5 Comparative Evaluation of OSS DL Systems (Greenstone and Dspace)

Categor	Sub-category	Dspace	Greenstone
General information	About	DSpace is an open source software package that Provides the tools for management of digital assets, and is commonly used as the basis for an institutional repository.	Greenstone is a suite of software to serve digital library collections and build new collections. It provides a new way of organizing information and publishing it.
	URL	www.dspace.org	www.greenstone.org
	Membersh	Free for every one	Free for every one
	License	BSD license (This means that any organization can use, modify, and even integrate the code into their commercial application without Paying any licensing fees.)	GPL (General Public License)
	Language support	The DSpace web application is available in over twenty languages. So if English is not your local language, you can customize the language which DSpace uses. You can also configure DSpace to support multiple languages, so that the language your user sees is the 'preferred language' set in their web browser	English, French, Spanish, Russian and Kazakh. Greenstone also has interfaces in many other languages
	used	educational, government, private and commercial institutions.	universities, libraries, and other public
Design and customization	Stable	Dspace 1.8.0	Greenstone3: 3.04
	Developer	Dura Space	University of Waikato
	Developm	Active	Active
	Type	Institutional repository software	DLs
	Program language	Java	Java
	Search mode	You can decide what fields you would like to display for browsing, such as author, title, date etc. on your DSpace website and can also select any metadata fields you would like included in the	Various searching and browsing options, and include collections in Arabic, Chinese, French, Maori, and Spanish, as well as English.
	Protocol Supports	OAI-PMH, OAI-ORE, SWORD, WebDAV, OpenSearch, OpenURL, RSS, ATOM	OAI-PMH, METS,
Security and access	Access	Internet	Internet or on CD-ROM
	File upload	books, theses, 3D digital scans of objects, photographs, film, video, research data sets, etc.	text, html, jpg, tiff, MP3, PDF, video, and Word
	Privacy policy	The personal information we receive through DSpace, such as names, emails and phone numbers, is used solely for the purposes of the functioning and assessment of the system.	The main Greenstone download, suitable for most users' General documentation about Greenstone, for users and developers. A work in progress.

Collaborative tools	**Mailing Lists**	The primary way DSpace users discuss issues and communicate with one another is through the various mailing lists . The DSpace community has three very active mailing lists where you can meet other DSpace users and developers.	There are two mailing lists intended primarily for discussions about the Greenstone digital library software. Active users of Greenstone should consider joining the mailing list and contributing to the discussions
	Blogs	The DuraSpace Blog features news and information from the DSpace and Fedora communities. You may subscribe to the RSS from this blog, and/or receive a monthly DuraSpace Blog Digest by subscribing to dspace- general@lists.sourceforge.net	Contain a blog category and weekly the updates of this section will report with the plans and changes for next week will be mention.
	FAQ	Check out the EndUserFAQ for answers to the most common questions about DSpace. For technical questions about customizing DSpace or contributing to the codebase, see the development/technical FAQs at the TechnicalFAQ Wiki	The FAQ (Frequently Asked Questions) pages consist of a set of commonly asked questions and their answers. These cover issues such as installing, building collections, formatting etc.
	Training Materials	There are a variety of training materials and resources that have been developed by the DSpace user community. Many of these can be used as self- guided tutorials or adapted and used as the basis to lead training sessions.	Installer's Guide describes in detail the installation process.
	Social medias	RSS, Facebook	RSS, Facebook

6 Vu DL Software System

Given our analysis in the previous section, it can be seen that some facilities provided by a recent development, Vu DL, is an attempt to address some of the recommendations above for Greenstone and Dspace. It is a simple to use DL administration application completely powered by open source technologies (GPL). Some of the functionality currently provided by Vu DL consists of a built-in METS metadata editor, service image generation tools, an XML database repository, and an OAI server, along with record drivers for easy implementation with Vufind, produce "service" files from scanned originals.

For example Vu DL uses image facility to respect patron's habits such as display of resources like book in form of digital book not just html pages, folio book pages, or zoom in and out preferences in order to allow the patrons experience a customized and familiar virtual environment. Also things such as "issue tracking" can be added on as a special option for patrons. An issue tracker (JIRA) allows patrons to search through an issue archive to see if the problem s/he may be experiencing has been fixed. Another special feature of Vu DL package is agent information that includes information about IP owners and editors. Such information can enhance the security. [11]

7 Conclusions

In this paper we presented the results of a user-centric comparative study of two DLs. First, surveying the DL literature, we derived key factors that can support user-centric design of DLs. Then using these factors we did a comparative survey of two popular DL platforms. We observed that, from a user-centric perspective, both these systems

have few advantage as well as some drawbacks. Significant drawbacks include: the lack of a good resources classification, the lack of creativity in order to attract patrons in form of design and customization of the DL webpage such as weak design of search mode especially on display of search results and the lack of offering personal profile for more security.

It is to be noted that new DLs users are not only consumers but also producers of information. By describing information collected through the DL they can create new information objects that are published in the DL, thus inspiring its content. Also DLs of the future will be capable to work on huge information object types. For example, one can mix text, tables, etc. of scientific data as well as images (such as integrated 3D images) with appropriate explanation and videos; for the patron to better understand the subject of interest. New DLs are therefore required to present services that support the authoring of these new objects and the workflows that lead to their publication. Libraries of the future can provide their patrons such integrated DL environments that help user assimilate and understand information in a much more profound manner. Currently we are observing a large expansion in the demand for DLs (from socio-economic to research projects of interest) and these require intense collaborative efforts by user groups that earlier would not have had much of a collaborative working relationship. Supporting such collaborative working partnerships can further enhance the service DLs may provide to different organizations in a globalized world, that extend far beyond a persons' local neighbourhood and also in [1,2] the authors discuss a social search engine that is designed to guide naive users through search networks for finding answers to their questions. In a similar manner, in libraries of the future, DL support assistants will provide the familariality of 'someone' who can help library clients navigate the vast amounts of information that exists on the web, in a guided manner.

Significant future research is necessary to validate these (sometimes intuitive) observations. For example, we are aware that there are limitations in information sharing between institutes, corporate, organizations and universities – such barriers to effective functioning of DLs need to be overcome.

References

[1] Chi, H.: Technical perspective: who know? Searching for expertise on the social web. Communications of the ACM (55/4), 110 (2012)

[2] Horowitz, D., Kamvar, S.D.: Searching the village models and methods for social search. Communications of the ACM (55/4), 111–118 (2012)

[3] The mercury electronic library and library information system ii the first three years. Mercury technical reports series 6. Carnegie Mellon university (1992), http://www.cs.cornell.edu/wya/papers/Mercury6.doc

[4] Seppälä, K.: Trade Union and University Lifelong Learning in Partnership TULIP Interim External Evaluation Report. University of Turku Centre for Extension Studies, Finland (2008), http://www.tulipnetwork.org.uk/Website%20linked%20docs/TULIP%20Interim%20Evaluation%20Report%20final.pdf

[5] Arms, W.: Manuscripts of digital libraries. MIT Press (2000)

[6] Adams, A., Blandford, A.: Digital Libraries in Academia: Challenges and Changes. In: Lim, E.-p., Foo, S.S.-B., Khoo, C., Chen, H., Fox, E., Urs, S.R., Costantino, T. (eds.) ICADL 2002. LNCS, vol. 2555, pp. 392–403. Springer, Heidelberg (2002)
[7] Blandford, A.: Interacting with information resources: digital libraries for education. International Journal of Learning Technologies 2(2/3), 185–202 (2006)
[8] Brown, E., Velazco, L., Kirksey, G., Ramaswamy, S., Rogers, M.: On Developing a Simple in-house Digital Archive. In: 43rd ACM Southeast Conference, Kennesaw, GA, USA, March 18-20 (2005), http://dl.acm.org/citation.cfm?id=1167412
[9] Lang, B.: Developing the digital library. Towards the Digital Library, pp. 227–233. The British Library, London (1998)
[10] Leiner, B.M.: 2005-last update, the scope of the digital library (2005), http://en.wikipedia.org/wiki/Digital_library
[11] Lacy, D.: home-grown digital library system: built upon open source XML technologies and metadata standards (2011), http://vudl.org/files/8113/0107/3566/VuDL_-_code4lib_-_2011.ppt

Adaptive Hexa-Diamond Search (AHDS) Algorithm for Fast Block Matching Motion Estimation

M.K. Pushpa and S. Sethu Selvi

M.S. Ramaiah Institute of Technology
pushpachandan@rediffmail.com
selvi@msrit.edu

Abstract. In this paper, a simple fast block matching algorithm based on Adaptive Hexa-Diamond Search (AHDS) is proposed to estimate motion vector parameter. This search consists of two sequential search stages: 1) initial search 2) refined local search. For initial search stage, hexagonal pattern is proposed to reduce the computational complexity, in which the least error is determined. The point that has the least error becomes the origin for subsequent refined local search steps, and the search pattern is changed to Small Diamond Search Pattern (SDSP) until the final motion vector (MV) is found. Based on the experimental results, AHDS needs only 4% of the total computations compared to Full Search (FS) algorithm and is very close to Adaptive Rood Pattern Search (ARPS). The result shows that time saved is 93.24% compared to FS algorithm. AHDS gives performance close to ARPS in terms of computational complexity, processing time and similar results as it is in other existing algorithms in terms of image quality.

1 Introduction

Block matching algorithm for motion estimation (ME) has been widely adopted by video coding standards[10, 4] such as H.263, H.264, MPEG-2, and MPEG-4. It is an efficient technique to eliminate temporal redundancy. Motion estimation deals with the fact that the two consecutive frames of a video sequence will be similar except for changes induced by moving objects within the frames. In block matching algorithms (BMA) the current frame is divided into a matrix of macro blocks (MB) that are compared with corresponding block and its adjacent neighbors in the previous frame to produce motion vector (MV). The search area for a best match block is constrained up to search parameter 'P' pixels on all four sides of the corresponding macro block in previous frame. There are several block matching algorithms like Full search (FS)[3], Three step search (TSS)[2], New three step search (NTSS)[11], Simple and efficient three step search (SESTSS)[6], Four step search (FSS)[9], Diamond search (DS)[12] and adaptive rood pattern search (ARPS)[14]. Among these algorithms, FS is most computationally expensive with the best image quality and makes ME the main bottleneck in real-time video coding applications, so the fast block matching algorithms like TSS, NTSS, SESTSS, DS and ARPS are developed to reduce computational complexity by limiting the number of search points. Thus, using fast BMA is indispensable to reduce the computational cost. The existing fast BMAs[1] can be

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 85–93.
springerlink.com

classified into four categories as 1) Fast BMA using a fixed set of search pattern, 2) Fast BMA based on inter-block correlation[7], 3) Fast BMA using hierarchical or multi resolution[8] search framework and 4) Fast BMA using sub sampled pixels on matching-error computations [14].

From the analysis, each class of above said fast algorithms achieve different trade-off between algorithm complexity, search speed and picture quality. Hence the idea is to combine the pattern-based method with the spatial-correlation method. Inter-block correlations dynamically determine the size of the search pattern[5]. Our major concern for the algorithm development is to achieve simple and feasible implementation. The proposed algorithm is compared with the FS, TSS, DS, SETSS, NTSS, FSS and ARPS in terms of complexity and performance.

In Section 2, the proposed AHDS algorithm for motion estimation is described in detail. The proposed AHDS uses search points in hexagon pattern[13] for initial search and then switches to SDSP for repetitive refined local search steps. The performance of the proposed method is demonstrated by experimental results in Section 3. Conclusions are drawn in Section 4. In this paper the AHDS is compared with other algorithms in terms of Peak Signal to Noise Ratio (PSNR), number of computations and elapsed processing time.

2 Adaptive Hexa-Diamond Search (AHDS) Algorithm

A small search pattern which has compactly spaced search points is more suitable than a large search pattern containing sparsely spaced search points to detect small motion. The large search pattern has the advantage of quickly detecting large motion, but results in unnecessary search for small MVs. The optimality of the pattern based search depends on size of the search pattern and the magnitude of the predicted MV. Hence, different search patterns are used in accordance with the computed motion behavior for the current block. Therefore there are two issues to be addressed: 1) to predetermine the motion behavior of the current block for effective motion estimation and 2) the adaptive size and shape of the search pattern.

Regarding the first issue, in most cases adjacent MBs belong to the same moving object and has similar motion. Therefore, the current block's motion behavior can be reasonably predicted by referring to its neighboring block's MVs in the spatial and/or temporal domain. Hence the coherency of the motion in a frame is considered i.e. if the MBs around the current MB moved in a particular direction then the current MB will also have similar motion behavior. For the second issue, two types of search patterns are used: one is the adaptive pattern for initial search while the other one is the fixed size search pattern for repetitive refined local search until the final MV is found.

The search is based on the fact that the motion in a frame is coherent and the blocks are processed in a raster-scan order. The MV of the neighboring blocks on the immediate left, above, above-right, above left are available as region of support (ROS) and is considered as reference to the current block to predict the target MV accurately. ROS is used as only one block that is situated at immediate left to the current block to predict the MV because it minimizes the memory requirement. Calculating the statistical average of MVs in the ROS is a common practice to obtain the predicted MV. The mean and median predictions have also been tested.

Two types of search patterns are used. One is hexagon pattern, which is dynamically determined for each MB according to its prediction motion behavior. The hexagon pattern will be used only once at the beginning of each MB search. The objective is to find a good starting point for the remaining local search so as to avoid unnecessary intermediate search and reduce the risk of being trapped into local minimum in the case of long search path. The new starting point identified is as close to the global minimum as possible. Since the usage of more blocks involves higher computational complexity, the search will be done in a selected area (P). To start with the search, first confirm the initial position, if the macro block is in the left most column then make sure that only once hexagonal pattern search is done with step size = 2, else step size can be found

$$S=\max \{|MV_{predicted}(x)|, (|MV_{predicted}(y)|\} \quad (1)$$

where $MV_{predicted}(x)$ and $MV_{predicted}(y)$ are the x-coordinate and y-coordinate of the predicted MV respectively. Hence the proposed search algorithm is used in an area where there is a probability of finding the best match block limiting the search to the neighboring blocks.

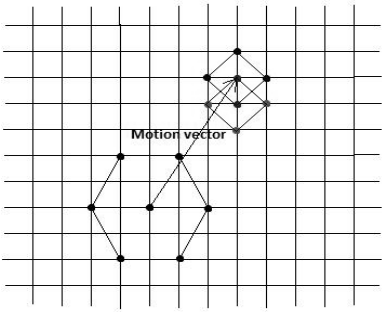


Fig. 1. Adaptive Hexa-Diamond Search Pattern

Fig.1 shows the search pattern used in the algorithm the shape of the initial pattern is symmetrical with six search points located at six vertices to form a hexagon pattern. Including one at the centre point there are total 7 search points. The step size refers to the distance between the centre and the search points. Since the initial shape of the search pattern includes all directions it can quickly detect any motion and the search will directly jump to a region in the direction of the predicted MV. The outcome of the initial stage is to detect the major trend of the moving object. Furthermore, the symmetry in shape of the AHDS not only benefits hardware implementation but also contributes to the robustness.

The hexagon format is used because the boundaries of the hexagon are interlaced horizontally and vertically due to which the AHDS search could search for motion to the very edge of the image. In addition to the hexagon pattern it is desirable to add predicted motion vector into the search as it is likely to be similar to the target MV. Hence, this contributes to increase in probability of accurately detecting the motion in the initial search stage. The magnitude of the MV's component with greater absolute value is nearly close to the length of the MV. The initial idea in deciding the pattern

size is to use only one of the two components of the predicted MV that has the larger magnitude.

$$S=\max \{|MV^2_{predicted}(X)|, |MV^2_{predicted}(Y)|\} \quad (2)$$

In the above step, the adaptive pattern gives the new search centre, which becomes the origin for subsequent search steps i.e. SDSP, similar to which is used in DS, comprising of 5 points of unit step. In refined local search the procedure keeps on doing SDSP until least error is found to be at the centre of the SDSP.

In summary, the search pattern contains a hexagon pattern consisting of either 8 (no overlapping) search points including the search point indicated by the predicted MV or 7 (with overlapping) search points to be searched in the initial search stage when predicted MV is not zero while the refined local search comprises of 5 search points in diamond pattern.

3 Experimental Results and Analysis

A standard mathematical model is used for the analysis of PSNR which measures an objective difference between the two images by mean-square error (MSE). The following Eq.3 and Eq.4 are used to calculate PSNR

$$PSNR \ (dB) = 10 * \log \left(\frac{255^2}{MSE} \right) \quad (3)$$

where

$$MSE = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} | C_{ij} - R_{ij} |^2 \quad (4)$$

C_{ij} and R_{ij} are the pixels being compared in Current block and reference block respectively and NxN is image size.

In terms of block distortion method, the sum of absolute differences (SAD) is commonly used and is defined by Eq.5

$$SAD = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} | C_{ij} - R_{ij} | \quad (5)$$

The AHDS algorithm uses %PSNR to indicate gain in PSNR using the Eq.6

$$\% \ PSNR = 100 - \left(\frac{FS_{psnr} - AHDS_{psnr}}{FS_{psnr}} * 100 \right) \quad (6)$$

To evaluate the performance of the proposed AHDS technique, it is compared against FS, TSS, NTSS, SESTSS, FSS, DS, ARPS in terms of %PSNR, number of computations or number of search points and elapsed processing time. The macroblock size is 16×16 pixels and search window size is 7×7. Six video sequences are used in simulation. The original and compensated images of “Alex”, ”diskus”, ”mom”, ”mom_daughter”, “sflowg”, “stennis” are shown in Figures 2.a to 2.f.

Fig. 2a. Original and Compensated images of "alex"

Fig. 2b. Original and Compensated images of "diskus"

Fig. 2c. Original and Compensated images of "mom"

Fig. 2d. Original and Compensated images of "mom_daughter"

Fig. 2e. Original and Compensated images of "sflowg"

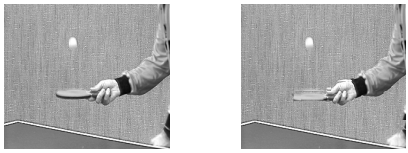


Fig. 2f. Original and Compensated images of "stennis"

The simulation is done using MATLAB7. 30 frames are used. The results are tabulated for AHDS and other algorithms. Table 1 shows the average %PSNR improvement values per frame for six video sequences. FS gives the highest PSNR amongst all block matching algorithms as it calculates the cost function at each possible location in the search window and hence the FS algorithm is used as reference and %PSNR improvement value is calculated. Figures 3.a to 3.f shows the average PSNR for all sequences.

Table 1. Average %PSNR Improvement for 30 frames compared with FS

Algorithms Sequences	TSS	SESTSS	NTSS	FSS	DS	ARPS	AHDS
alex	99.785	99.561	99.995	99.853	99.977	99.958	99.954
diskus	99.629	98.674	99.631	99.426	99.478	99.668	99.694
mom	99.894	99.808	99.938	99.811	99.767	99.904	99.882
mom_daughter	99.738	99.29	99.864	99.753	99.81	99.793	99.796
sflowg	99.369	99.188	99.546	98.857	98.673	99.442	99.424
stennis	99.315	98.895	99.272	99.246	99.475	99.083	99.086

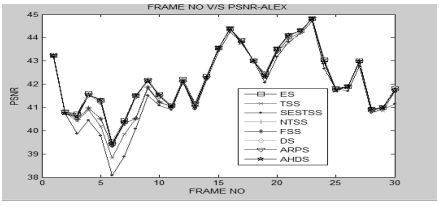


Fig. 3a. Comparison of average PSNR for "alex" video sequence

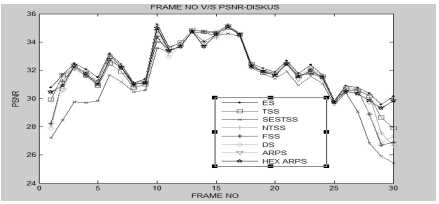


Fig. 3b. Comparison of average PSNR for "diskus" video sequence

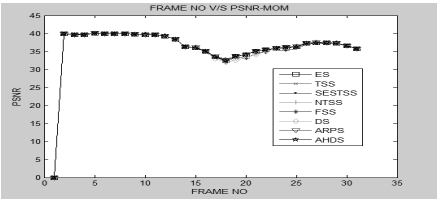


Fig. 3c. Comparison of average PSNR for "mom" video sequence

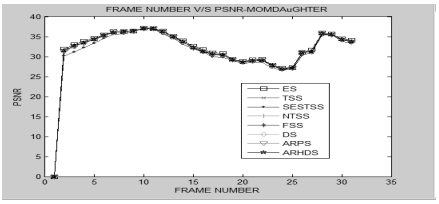


Fig. 3d. Comparison of average PSNR for "mom_daughter" video sequence

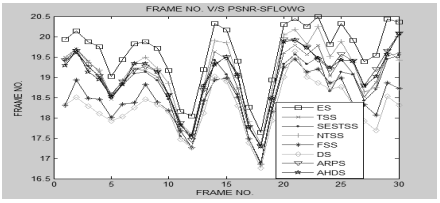


Fig. 3e. Comparison of average PSNR for "sflowg" video sequence

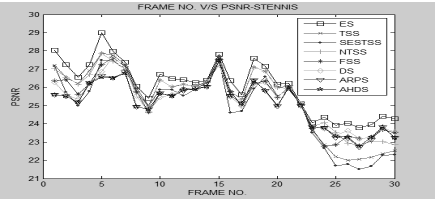


Fig. 3f. Comparison of average PSNR for "stennis" video sequence

Referring to Table 2 AHDS requires less number of computations as compared to other algorithms. AHDS needs only 4% of the total computations compared to FS algorithm and is very close to ARPS. The results are as shown in Figures 4.a to 4.f for all sequences.

Table 2. Average computations for 30 frames

Algorithms Sequences	ES	TSS	SESTSS	NTSS	FSS	DS	ARPS	AHDS
alex	204.28	23.5	16.679	16.847	16.843	13.667	7.3557	8.1046
diskus	204.28	23.426	16.216	21.99	19.226	17.593	9.5987	10.913
mom	206.52	23.416	16.504	20.164	17.768	14.899	7.3592	8.182
mom_daughter	206.52	23.464	16.436	19.951	17.8	15.022	7.7289	8.6652
sflowg	202.05	23.352	15.442	26.07	20.427	19.856	10.619	12.29
stennis	202.05	23.165	16.492	17.421	14.698	18.963	7.0886	7.7284

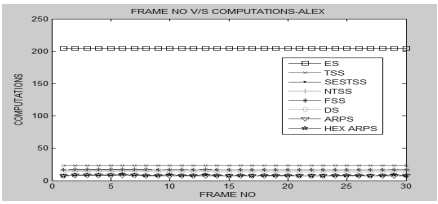


Fig. 4a. Comparison of average computations per frame for "alex" video sequence

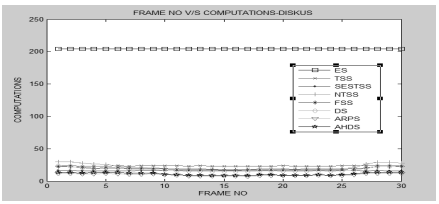


Fig. 4b. Comparison of average computations per frame for "diskus" video sequence

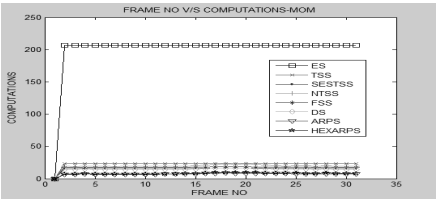


Fig. 4c. Comparison of average computations per frame for "mom" video sequence

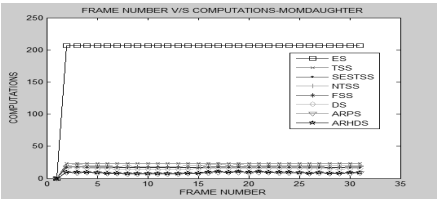


Fig. 4d. Comparison of average computations per frame for "mom_daughter" sequence

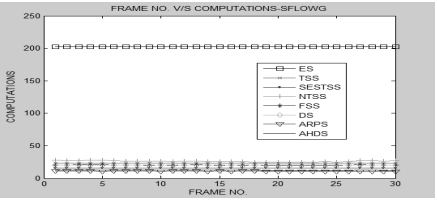


Fig. 4e. Comparison of average computations per frame for "sflowg" video sequence

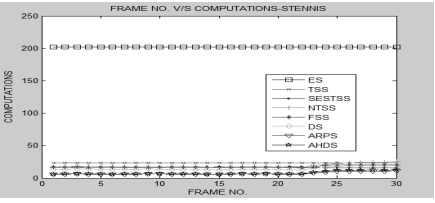


Fig. 4f. Comparison of average computations per frame for "stennis" video sequence

Table 3 shows that the percentage total time saving for 30 frames compared with FS algorithm. The result shows that time saved is 93.24% compared to FS algorithm for "tennis" sequence (fast motion). Total time saving is calculated in Intel Core2Duo processor @1.80GHz. The results are as shown in Figures 5 .a to 5.f for all sequences.

Table 3. Total Percentage time saving for 30 frames compared with FS

Algorithms Sequences	TSS	SESTSS	NTSS	FSS	DS	ARPS	AHDS
alex	87.39	90.07	90.62	90.65	90.49	93.58	93.06
diskus	87.41	90.24	87.97	89.32	88.13	92.49	91.61
mom	87.04	89.61	88.47	89.67	89.39	93.34	92.77
mom_daughter	86.90	89.65	88.57	89.68	89.33	93.13	92.56
sflowg	87.15	90.14	85.74	88.33	86.26	91.57	90.63
stennis	87.43	89.86	90.09	89.71	89.23	93.63	93.24

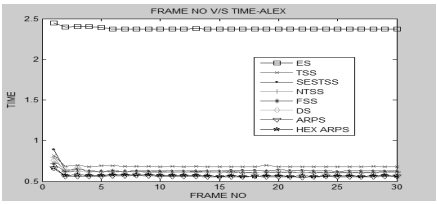


Fig. 5a. Comparison of total percentage time saving for "alex" video sequence

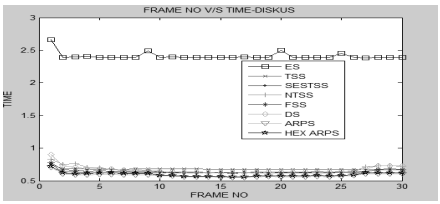


Fig. 5b. Comparison of total percentage time saving for "diskus" video sequence

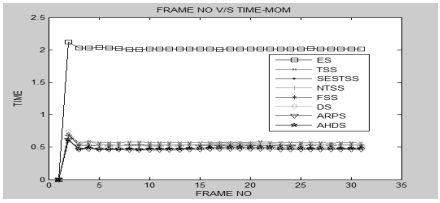


Fig. 5c. Comparison of total percentage time saving for "mom" video sequence

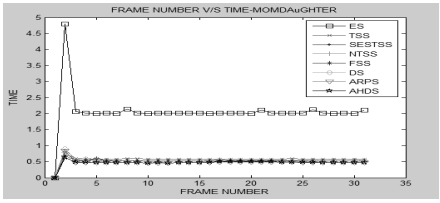


Fig. 5d. Comparison of total percentage time saving for "mom_daughter" sequence

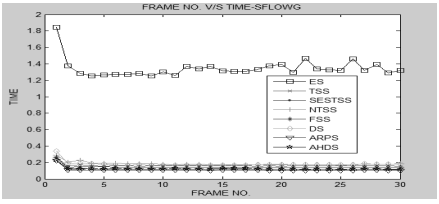


Fig. 5e. Comparison of total percentage saving time for "sflowg" video sequence

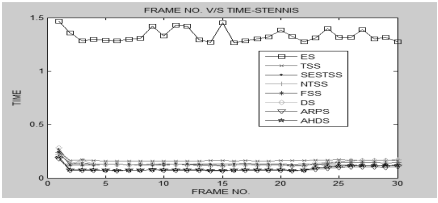


Fig. 5f. Comparison of total percentage time saving for "stennis" video sequence

4 Conclusion

The paper shows that the AHDS technique has less number of search points and as a result less computation is required. It shows that the improvement results from the adaptability of the search pattern. This avoids local minimum matching error points by tracking the trend of the motion at the initial stage. The algorithm is compared with FS algorithm and it shows improved PSNR. AHDS needs only 4% of the total computations compared to FS algorithm and is very close to ARPS. The result shows that time saved is 93.24% compared to FS algorithm for tennis sequence. Thus the proposed algorithm demonstrates superior speed while maintaining similar performance. Since it requires less computation it contributes to less memory usage. The results that are tabulated eliminate temporal redundancy and produces better quality images. Thus AHDS is an efficient and robust motion estimation algorithm for real time video coding application.

References

1. Murat Tekalp, A.: Digital Video Processing. Prentice-Hall, Englewood Cliffs (1995)
2. Barjatya, A.: Block Matching Algorithms for Motion Estimation. DIP 6620 spring final project (2004)
3. Furht, B., Greenberg, J., Westwater, R.: Motion Estimation Algorithms for Video Compressio, ch. 2 & 3. Kluwer Academic Publishers, Massachusetts (1997)
4. Richardson, I.E.G.: Video Codec Design, ch. 4, 5, & 6. John Wiley & Sons Ltd., West Sussex (2002)
5. Chana, J., Agathoklis, P.: Adaptive motion estimating for efficient video compression. In: Conf. Rec. 29th Asilomar Conf. Signals, Systems and Computers, vol. 1, pp. 690–693 (1996)
6. Lu, J., Liou, M.L.: A Simple and Efficient Search Algorithm for Block-Matching Motion Estimation. IEEE Trans. Circuits and Systems For Video Technology 7(2), 429–433 (1997)
7. Jain, J.R., Jain, A.K.: Displacement measurement and its application in inter frame image coding. IEEE Trans. Commun. COM-29, 1799–1808 (1981)
8. Uz, K.M., Vetterli, M., LeGall, D.: Interpolative multiresolution coding of advanced television with compatible subchannels. IEEE Trans. Circuits Syst. Video Technol. 1, 86–99 (1991)
9. Po, L.-M., Ma, W.-C.: A Novel Four-Step Search Algorithm for Fast Block Motion Estimation. IEEE Trans. Circuits and Systems for Video Technology 6(3), 313–317 (1996)
10. Ghanbari, M.: Video Coding-An Introduction to Standard Codecs, ch. 2, 5, 6, 7 & 8. The Institute of Electrical Engineers, London (1999)
11. Li, R., Zeng, B., Liou, M.L.: A New Three-Step Search Algorithm for Block Motion Estimation. IEEE Trans. Circuits and Systems for Video Technology 4(4), 438–442 (1994)
12. Shan, Z., Ma, K.K.: A New diamond search algorithm for fast block matching motion estimation. IEEE. Trans. Image Process 9, 287–290 (2000)
13. Ranjit, S.S.S., Sim, K.S., Besar, R., Md Salim, S.I., Subramaniam, S.K.: Estimation of Motion Vector Parameter using Hexagon-Diamond Search Algorithm (February 2010)
14. Nie, Y., Ma, K.-K.: Adaptive Rood Pattern Search for Fast Block-Matching Motion Estimation. IEEE Trans. Image Processing 11(12), 1442–1448 (2002)

A Probabilistic Solution to Rendezvous Problem

Shivam Agarwal, Arup Kumar Pal, Vihang Gosavi, and Hemant Gangolia

IIT Bombay
{shivamagarwal.iitb,arup0007,vihangtycoon,hemantgenie}@gmail.com

Abstract. Here is a description of a probabilistic algorithm to help two people, lost in a complex maze, find each other in minimum possible time. These problems are broadly classified as 'Rendezvous Problems'[1]. The algorithm can be used in many practical situations like robots finding each other, detectives catching thieves, etc. The important point to note is that this algorithm works when the user is aware of the entire network of roads. He need not know anything else say the initial location or the present location of the other person.

1 Introduction

'Rendezvous Problem' is a classical problem which dates back to the 18th century. One of its instances is say when two people get lost in a complex maze and there is no mode of communication between them. They have to find each other as quickly as possible. Should they wait at their positions hoping the other person to find them or should they roam aimlessly to increase the probability of meeting (according to them). Below is a probabilistic algorithm which might help them in this regard.

2 Key Idea of the Algorithm

The user who is going to use the algorithm needs to know the complete network (a graph with nodes as meeting of two or more roads and edges as the roads). Once this is known, using the speed of user, the algorithm finds all possible ways of going from the initial position and finds the probability of meeting for each such way. This probability is found using the general notion that nodes from where there is more visibility are better than those with less visibility. Once all probabilities are known, the user is directed to go via the maximum probable path.

3 Design of the Algorithm

3.1 Elements Used

3.1.1 Data Structures

1. Graph: Graphs are used to represent the maze i.e. the network of roads with roads as edges and crossings as nodes of the graph.
2. Priority Queue or Heap: Heap is used to implement Dijkshtra's algorithm[2] because of its fast insertion, extraction and modification of data (O(logn) time).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 95–99.
springerlink.com

3.1.2 Behavioral Functions

1. Setting of Probabilities: This function sets up the probabilities of vertices i.e. the nodes of the graph. The entire procedure is explained later.
2. Dijkshtra's Algorithm: This is a standard algorithm used to find the shortest distance between two nodes of the graph.
3. Decision Maker: This function, once given the probabilities of all nodes of the graph, finds the optimum path to be taken by the user.

3.2 Developing the Algorithm

3.2.1 Setting the Initial Probabilities of the Nodes

The nodes of the graph are assigned an initial probability based on the length of the path visible from them. The more is the length of path visible from a particular node, the more is the probability of finding the other person from that node. So, in other words, this probability is a measure of the quality of the node. The way this probability is assigned is quite simple and is as follows:

Let there are n edges originating from a vertex v of lengths l_1, l_2, l_3, ..., l_n. We denote the probability of the vertex v denoted by P_v as:

$$P_v = \frac{\sum_{i=1}^{n} l_i}{2 \times Total\ Length\ of\ all\ edges\ (L)} \tag{1}$$

The use of the factor of 2 in the denominator is because we want the sum of probabilities to be exactly equal to 1 and we have counted every edge twice.

3.2.2 Setting the Weighted Probabilities of the Nodes

Once the initial probabilities of the nodes are set, we need to set the weighted probabilities to these nodes. If the node is situated close to a probable node, such a node should have a high weighted probability since we can easily switch from such a node to the highly probable node and increase the probability of meeting. Let there are in total N vertices in the graph denoted by v_1, v_2, v_3, ..., v_n. The weighted probability of a vertex v denoted by W_v is:

$$W_v = \sum_{j=1}^{n} \frac{P_{v_j}}{t_{i,j}} \tag{2}$$

Where $t_{i,j}$ is the time taken by the user to go from vertex v_i to the vertex v_j. This time is calculated via the following equation:

$$t_{i,j} = \frac{dis_{i,j}}{speed} \tag{3}$$

Here, $dis_{i,j}$ is the shortest distance between the vertices v_i and v_j and is found using the 'Dijkshtra's Algorithm'.

Note:

Everything is in terms of discrete time intervals. There is no notion of fractional or zero time now. So, the time taken to go from any vertex to itself is now 1 unit i.e. $t_{i,i} = 1$ even though $dis_{i,i} = 0$. This is a kind of least count error in our system.

The weighted probabilities Wv_i are not actually probabilities. They may be greater than 1. They are only a measure to denote the quality of the nodes.

3.2.3 Making the Decision

Once we have with us the initial probabilities and the weighted probabilities, almost half the work is done. Given the initial position of the user, the algorithm directs him to the node with maximum weighted probability amongst all which are reachable from the initial position of the user.

3.2.4 Dynamic Probabilities

User waiting over a certain vertex: As the waiting time of the user over a certain vertex increases, the dynamic probability of that vertex continuously decreases (but in discrete time).

If the waiting time over the vertex v is denoted by lag_v, the dynamic probability of the vertex v, D_v is calculated as follows:

$$D_v = \frac{W_v}{1 + lag_v} \tag{4}$$

This inclusion of such a dynamic probability is necessary since if there were no such dynamic probability then the user would have remained static for the rest of the time and the algorithm would have failed.

User visiting a certain vertex repetitively: If the user visits a certain vertex again and again and he is not able to find the other person over there, its dynamic probability should get decreased. This is done by including a '*reluctance ratio*' denoted by α and is a constant between 0 and 1. Usually,

$$\alpha = 0.8$$

So, the dynamic probability becomes:

$$D_{v_{final}} = \alpha \times D_{v_{initial}} \tag{5}$$

Every time user visits a particular vertex, the dynamic probability of that vertex gets multiplied by the constant α. This is important to prevent the user from following a certain path repetitively.

4 Working with an Example

Figure 1 shows an example of a typical network of roads. The distances are written on the edges and nodes are numbered. Let's take two persons, both having speeds of, say 10 m/sec. Initial probabilities can be easily found using the equations given in the previous section. Also, a simulation has been shown when A starts from node 9 and B starts from node 1 in the Table 2.

Table 1. Probabilities and qualities of the vertices

Vertices	Weighted Probabilities	Quality
Vertex 1	0.127	Good
Vertex 2	0.114	Good
Vertex 3	0.115	Good
Vertex 4	0.098	Average
Vertex 5	0.099	Average
Vertex 6	0.138	Good
Vertex 7	0.115	Good
Vertex 8	0.109	Good
Vertex 9	0.07	Bad
Vertex 10	0.155	Best
Vertex 11	0.15	very good
Vertex 12	0.127	Good
Vertex 13	0.096	Average
Vertex 14	0.12	Good
Vertex 15	0.15	very good
Vertex 16	0.094	Average
Vertex 17	0.063	Poor
Vertex 18	0.065	Poor

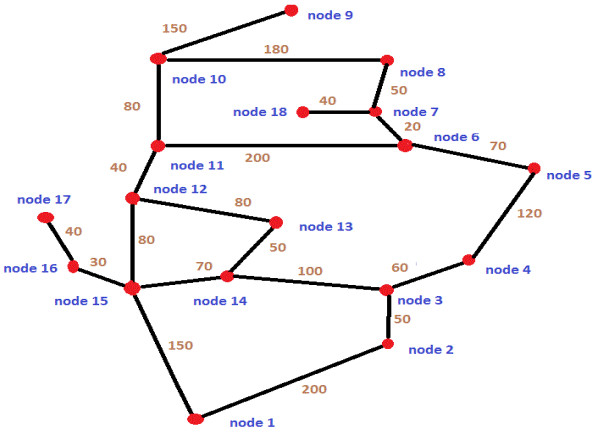


Fig. 1. The network of roads, with roads as edges and crossings as nodes

Table 2. The overall status of A and B trying to find each other

Time	A's Status	B's Status
0 sec	static at node 9	static at node 1
10 sec	moving towards node 10	moving towards node 15
15 sec	reached node 10	reached node 15
18 sec	moving towards node 11	static at node 15 but starts to move
23 sec	reached node 11	moving towards node 12
26 sec	moving back towards 10	reached node 12
30 sec	moving towards node 10	reached node 11

The above table clearly shows the simulation for the particular case we have taken when person A starts from the vertex 9 while B starts from the vertex 1. Notice the use of dynamic probabilities given by the Formulas 4 and 5. By following the algorithm, they are able to meet each other in a little span of time i.e. 30 sec.

5 Major Assumptions

The major assumptions in the algorithm are as follows:

Co-linearity of Nodes: The algorithm assumes that if a node, say B is connected to two nodes, say A and C, then A and C are not visible from each other even though all three nodes might be in a straight line.

Least Count Error: We are working with times which are positive integers. This forces an assumption in our algorithm that the speed of the users must be quite less compared to the distances since then only will the generated time intervals will be much larger than the least count time so as to neglect the least count error.

Linearity of Paths: The paths should be linear, not curved.

References

1. http://en.wikipedia.org/wiki/Rendezvous_problem
2. http://en.wikipedia.org/wiki/Dijkstra%27s_algorithm

Satellite Image Feature Extraction Using Neural Network Technique

T. Karthikeya Sharma, Sarvesh Babu N.S., and Y.N. Mamatha

Dept of Telecommunication Engineering,
RV College of Engineering, Bangalore 560059, India
{t.karthiksharma,sarveshbabu88}@gmail.com,
mamatharaj_76@rediffmail.com

Abstract. There has been a focus on developing image indexing techniques which have the capability to retrieve image based on their contents. The main feature extraction methods are content Based Image Retrieval (CBIR) also known as query by Image content (QBIC). This paper presents a technique to derive the colors, shapes, textures, or any other information that can be derived from a satellite image Using Texture filters and realizing it with artificial neural networks. This image processing technique are been utilized to identify important urban features such as buildings and gardens and rural features such as natural vegetation, water bodies, and fields. Textures are represented by Texel, which are then placed into a number of sets, depending on how many textures are detected in the image.

Index Terms: feed-forward, back-propagation, artificial neural network, SOM (self-Organization Mapping).

1 Introduction

The main objective is to extract the features of a satellite image using ANN tool. An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. Each neuron is assigned a weight to particular node of a texture of an image. SOM (Self Organization Mapping) technique is used by ANN to analyze texture of an image. ANN analyzes the image using a clustering tool and pattern recognition tools.

Procedure for Satellite Image Processing

Spatial Transformation Stage

A spatial transformation (also known as a geometric operation) modifies the spatial relationship between pixels in an image, mapping pixel locations in an input image to new locations in an output image. These functions perform certain specialized spatial transformations, such as resizing and rotating an image. In addition, it includes functions that you can be used to perform many types of 2-d and n-d spatial transformations, including custom transformations.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 101–106.
springerlink.com

2 Working on ANN Tool

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems.

3 ANN Tool Works on the Principle of SOM

There are two initialization (random and linear) and two training (sequential and batch) algorithms implemented in the tool box. By default linear initialization and batch training algorithm are used. The training is done in two phases: rough training with large (initial) neighborhood radius and large (initial) learning rate, and fine tuning with small radius and learning rate. If tighter control over the training parameters is desired, the respective initialization and training functions, e.g. SOM batch train, can be used directly. There is also a graphical user interface tool for initializing and training SOMs.

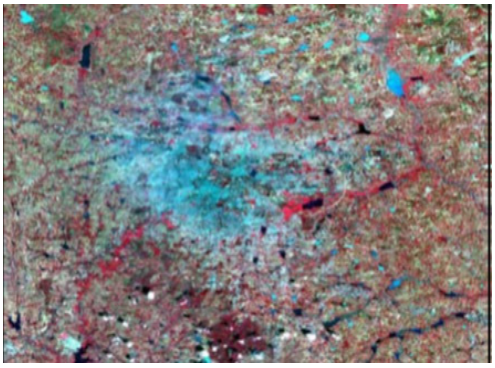


Fig. 1. Satellite Image used

4 Principle behind SOM

Units are connected to one another. Connections correspond to the edges of the underlying directed graph. There is a real number associated with each connection, which is called the weight of the connection. We denote by Wij the weight of the connection from unit ui to unit uj. It is then convenient to represent the pattern of connectivity in the network by a weight matrix W whose elements are the weights Wij. Two types of connection are usually distinguished: excitatory and inhibitory. A positive weight represents an excitatory connection whereas a negative weight represents an inhibitory connection. The pattern of connectivity characterizes the architecture of the network.

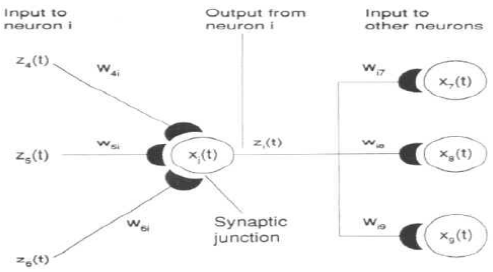


Fig. 2. Back propagation algorithm

5 Equations

A unit in the output layer determines its activity by following a two-step procedure. First, it computes the total weighted input xj, using the formula:

$$X_j = \sum_i y_i W_{ij} \tag{1}$$

Where yi is the activity level of the jth unit in the previous layer and Wij is the weight of the connection between the ith and the jth unit. Next, the unit calculates the activity yj using some function of the total weighted input. Typically we use the sigmoid function:

$$y_j = \frac{1}{1 + e^{-x_j}} \tag{2}$$

Once the activities of all output units have been determined, the network computes the error E, which is defined by the expression:

$$E = \frac{1}{2}\sum_i \left(y_i - d_i\right)^2 \tag{3}$$

Here yj is the activity level of the jth unit in the top layer and dj is the desired output of the jth unit. An important application of neural networks is pattern recognition. Pattern recognition can be implemented by using a feed-forward (figure 2) neural network that has been trained accordingly. During training, the network is trained to associate outputs with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern. The power of neural networks comes to life when a pattern that has no output associated with it, is given as an input.

6 Neural-Network Implementation Using MATLAB

6.1 Neural Network Clustering Tool

NCTOOL launches the neural network clustering wizard and leads the user through solving a clustering problem using a self-organizing map. The map forms a

compressed representation of the inputs space, reflecting both the relative density of input vectors in that space, and a two-dimension compressed representation of the input space topology. The work flow for any given problem has seven steps.

1. Collect data
2. Create the Network
3. Configure the Network
4. Initialize the weights and biases
5. Train the Network
6. Results obtained
7. Validate the Network

7 Results Obtained

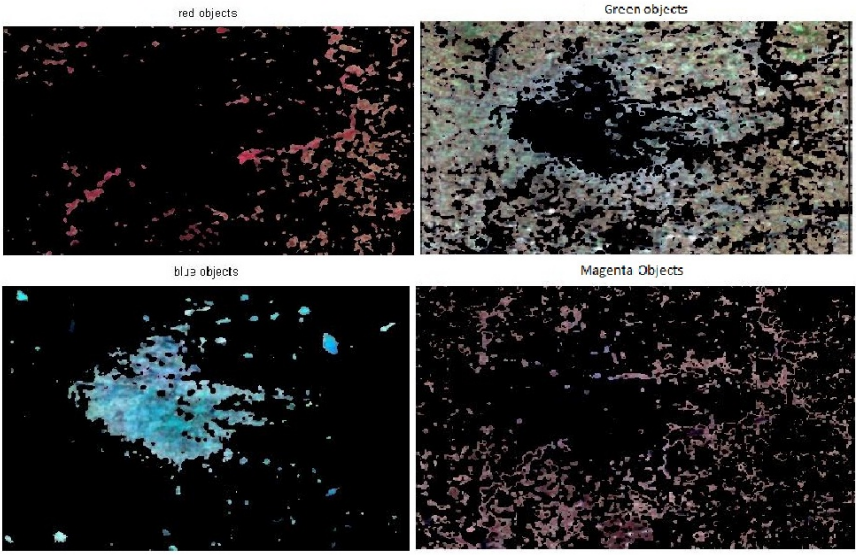


Fig. 3. Outputs of Spatial domain Transformation

8 Validate the Network

Self-organizing feature maps (SOFM) learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on. This can tell you how many data points are associated with each neuron. SOM weight positions plots the input vectors as green dots and shows how the SOM classifies the input space by showing blue-gray dots for each neuron's weight vector and connecting neighboring neurons with red lines.

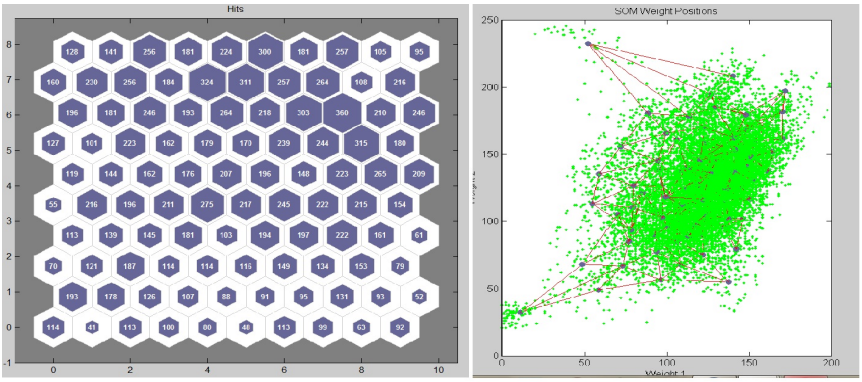


Fig. 4. Sample hits by Nctool and SOM weight positions

9 Creating a Regression Plot

The next step in validating the network is to create a regression plot, which shows the relationship between the outputs of the network and the targets. If the training were perfect, the network outputs and the targets would be exactly equal, but the relationship is rarely perfect in practice. It calculates the trained network response to all of the inputs in the data set. The final command creates three regression plots for training, testing and validation. The three axes represent the training, validation and testing data. The dashed line in fig.5 shows each axis represents the perfect result – outputs = targets. The solid line represents the best fit linear regression line between outputs and targets.

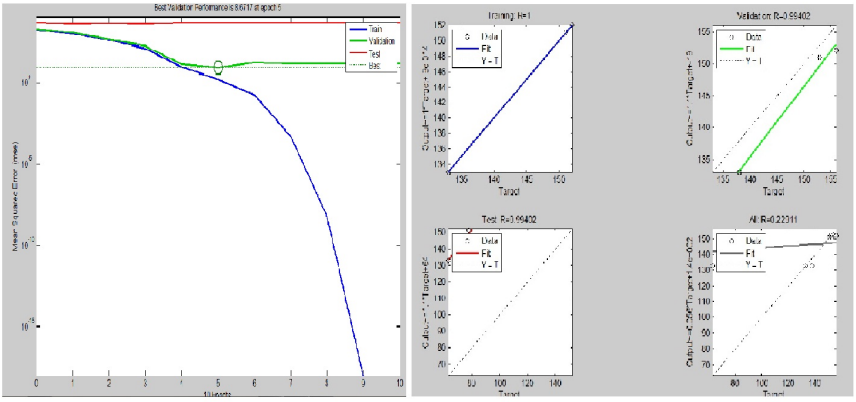


Fig. 5. Validation performance and Plot regression

The R value is an indication in fig.5 shows the relationship between the outputs and targets. If R = 1, this indicates that there is an exact linear relationship between outputs and targets. If R is close to zero, then there is no linear relationship between outputs and targets.

Abbreviations and Acronyms

NC - neural network cluster, SOM - self-organization tool.

10 Conclusion

The method presented here estimates the amount of Color (RGB) content in a satellite image. Neural networks also contribute to other areas of research such as neurology and psychology. In this system initial image processing stage is implemented at first and image segmentation procedure is done. Then suggested features are calculated for each region and a three layer perceptron neural network is trained for detection of Texture in satellite images. After these processes the system is evaluated using nctool and nntool. The plots value of these calculated parameters shows the percentage of Texture content in desirable level of a satellite image.

References

[1] Chen, S., Mulgrew, B., Grant, P.M.: A clustering technique for digital communications channel equalization using radial basis function networks. IEEE Trans. on Neural Networks 4, 570–578 (1993)

[2] Duncombe, J.U.: Infrared navigation—Part I: An assessment of feasibility. IEEE Trans. Electron Devices ED-11, 34–39 (1959)

[3] Lin, C.Y., Wu, M., Bloom, J.A., Cox, I.J., Miller, M.: Rotation, scale and translation resilient public watermarking for images. IEEE Trans. Image Process. 10(5), 767–782 (2001)

[4] Ardö, J., Pilesjö, P., Skidmore, A.: Neural networks, multi temporal Landsat The-matic Mapper data and topographic data to classify forest damages in the Czech Republic. Canadian Journal of Remote Sensing 23(3), 217–229 (1997)

[5] Rajani, T., Mangala, Bhirud, S.G.: An Effective ANN-Based Classification System for Rural Road Extraction in Satellite Imagery. European Journal of Scientific Research 47(4), 574–585 (2010) ISSN 1450-216X

[6] Boggess, J.E.: Identification of Roads in Satellite Imagery Using Artificialneural Networks: Acontextual Approach. Computer Science Department, vol. (601), pp. 2325–2756. Mississippi State University, P. O. Drawer CS, Mississippi State, MS 39762, U.S.A

[7] Rekik, A., Zribi, M., Hamida, A.B., Benjelloun, M.: An Optimal Unsupervised Satellite image Segmentation Approach Based on Pearson System and k-Means Clustering Algorithm Initialization. International Journal of Signal Processing 5(1), 38–45 (2009)

Identifying Refactoring Opportunity in an Application: A Metric Based Approach

Syamala Kumari Dora and Debananda Kanhar

Computer Science Dept., NIST
{dora.syamala,devanand.kanhar}@gmail.com

Abstract. This paper defines a set of program restructuring operations (refactorings) that support the design, evolution and reuse of java application. Long parameter list and Shotgun surgery most complex refactorings are defined. Degree of coupling and cohesion is calculated to check whether these bad smell having low cohesive and high coupling or not. Removing these bad smells is one way of avoiding problems that arise due to the presence of bad smells. This makes the source code more maintainable and more comprehensible.

Keywords: Bad smells, Refactoring, NRV, NAV, DIT, NOC, CBO, RFC.

1 Introduction

The size and complexity of software systems increases with time. The maintenance and debugging of such system is more costly, more difficult and burdensome. Object oriented software systems are known to evolve during their whole lifetime. This evolution leads to several anomalies (Bad smells) decreasing desired system properties like readability, flexibility, scalability, modifiability and maintainability. Refactoring allow re-engineering of these desired properties by changing the internal structure of software while preserving its external behavior. [4]

1.1 The Problem

Most of the authors detect duplicate code [11, 5] and using code cloning to avoid the complication of maintenance and evaluation of software. The design of reusable software is very hard. Reusable software usually is the result of many design iterations, some of these iterations occur after the software (i.e. backward engineering) [6].

Some changes to object-oriented (JAVA) software can be made simply by adding new subclasses or by adding new operations on existing classes, while leaving most of the original software unchanged [8]. However, JAVA software is harder to change than it might at first appear to be. Changing in java software often requires changing the abstractions embodied in existing object classes and the relationships among those classes. This involves structural changes such as moving variables and methods between classes and partitioning a complicated class into several classes. When a structural change is made to a class or set of classes, corresponding changes may also be needed elsewhere in a program, due to naming, typing and scoping (inheritance) dependencies. When we are talking about dependencies and consistently updating the

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 107–112.
springerlink.com

program, it can be time consuming, difficult and error prone. The reusability benefits of object-oriented programming (JAVA) can be difficult to realize without some form of automated support for making these structural changes [1].

2 Related Work

Object –oriented programming is often touted as promoting software reuse. Sometimes however the benefits of object-oriented approach are overstated, and claims are made that features can be added to an object-oriented system without disturbing the existing implementation [1].

2.1 Software Reuse

The high costs of developing software motivate the reuse and evolution of existing software. Software reuse in its broadest sense involves reapplying knowledge about one software system to reduce the efforts of developing and maintaining another system [10]. Closely related to software reuse is software maintenance, where knowledge about a software system is used to develop a version that refines or extends it. Approaches that support reuse address one or more of the following four important aspects [2]: (1) Finding a reusable component, (2) Understanding the component. (3) Modifying a component or a set of components, (4) Composing the components together.

While some software reuse techniques have focused at the code level [7], others have focused on design-level reuse [9]. There are limitations on the reuse of code [7]: it works best when the domain is narrow and well understood and the underlying technology is very static. Sometimes the design of software is reusable even when the code is not. However, a major problem with design-level reuse [9] is that there is no well-defined representation system for design. Reuse does not happen by accident; one needs to plan to reuse software and look for software to reuse. Reuse requires the right attitude, tools and techniques [3]. Tools and techniques to support software reuse include compositional and generational approaches [10]. The composition-based model of reuse is based on the notion of plugging components together, with little or no modification of those components, in order to create target software systems. The components might be code skeletons [2], subroutines [9] or methods [1].

Restructuring a program can make it easier to understand the design of a program and can assist in finding reusable components. Some restructurings modify a component to make it more reusable; such components can be easier to compose together for an application.

2.2 Software Restructuring

Software sometimes needs to be restructured before it is reused. Software restructuring as “the modification of software to make the software: (1) Easier to understand and to change or (2) Less susceptible or error when future changes are made [1].

A major goal of software restructuring is to preserve or increase the value of a piece of software. Restructuring a system may make it possible to add more features to the existing system [1], or make the software more reusable in other systems.

Software restructuring approaches have become increasingly attractive as the cost of programmer time relative to computer time has increased. Software restructuring is most often used during software maintenance, where the lack of software structure often is most evident and expensive. However, it can also be applied in the earlier design and development phases.

3 Proposed Approach

Our approach is to detect two bad smells long parameter list and shotgun surgery. Long parameter list says, a method is having large number of parameters as its input and shotgun surgery says if a super class is modified then that change will effects to the derived classes. To measure these bad smells we used some metrics (NRV, NAV, DIT, NOC, and CBO). Software metrics have been proved to reflect software quality evaluation methods [6]. These software metrics is useful to provide measurable information about the structure of the software system. The result of these evaluation methods can be used to indicate which parts of a system to be reengineered.

3.1 Detecting Bad Smell

3.1.1 Long Parameter List

No consistent or precise definition of large parameter list [7] is currently available. Large parameter lists are often operationally defined by individual detection methods. One of the detection methods is categorized as follows:

In large parameter list we can find three coupling types:

- Data and control flow coupling,
- Global coupling,
- Environment coupling

Using these measures, a module coupling indicator, Mc, is defined in the following way:

$$Mc=k/M \tag{1}$$

Where k=1, a proportionality constant and

$$M=Di + (a*Ci) + (b*Co) + Gd + (c*GC) + W + R \tag{2}$$

Where a=b=c=2, Di=number of input data parameters, Ci=number of input control parameters, Co=number of output control parameters, Gd=number of global variables used as data, Gc=number of global variables used as control, W=number of modules called (fan-out), R=number of modules calling the module under consideration (fan-in).

The higher the value of Mc, the lower is the overall module coupling.

3.1.2 Shotgun Surgery

This can be identified by the degree of interdependency between modules. And this can be detected by Coupling Model Graph [6] based technique. There are no standard measures of interdependency. For given modules x and y, we can create an ordinary classification for coupling by defining six relations on the set of pairs of modules:

- No coupling relation R0 : x and y have no communication.
- Data coupling relation R1: x and y communication by parameters.
- Stamp coupling relation R2: x and y accept the same record type as a parameter
- Control coupling relation R3: x passes a parameter to y with the intention of controlling its behavior.
- Common coupling relation R4: x and y refer to the same global data.
- Content coupling relation R5: x branches it into, changes data in, or alter statement in y.

Coupling model graph based technique [6] to model CMG we use a directed graph with more detail than a call graph but less than the full module structure chart. The nodes correspond to the modules, and there may be more than one arc between two nodes.

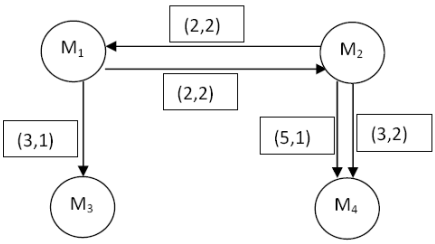


Fig. 1. Coupling Model Graph

Each arc from node x to a node y represents coupling interdependency between modules x and y; specifically, each arc is labeled by a pair (i ,j), where i represents the coupling relation R_i, and j is the number of times the given type of coupling occurs between x and y. e.g. : There are four modules M1,M2,M3, and M4. Where modules M1 and M2 share two common record types, module M1 pass to module M3 a parameter that acts as a flag in M3, and module M2 branches into module M4 and also passes two parameters that act as flags in M4.

3.2 Measure the Bad Smells with Some Metrics

3.2.1 Long parameter List

Long parameter S includes code fragment f1, f2,….., fm. A block/code fragment fi references Si number of variable defined externally, and it assigned t_0 t_i number of variables defined externally.NRV(S) represents the average of the externally defined variables referenced in the code fragments. NAV(S) represents the average of externally defined variables assigned in the code fragments [7].

3.2.2 Shotgun Surgery

The objects are coupled if and only if at least one of them acts upon the other. x is said to act upon y if the history of y is affected by x, where history is defined as the

chronologically ordered sates that a substantial individual traverses in time. The different measurement matrices are Weighted Method per Class(WMC),Depth of inheritance tree (DIT), Number of Children(NOC), Coupling between object classes(CBO) [8].

4 Automated Tool

We implemented an automated tool with java. We developed this tool, having following functionalities:

- Search the bad smells,
- Measure the bad smells taking some metrics,

As a preprocessing, Java programs are parsed and stored in syntax table for each and every file. The application domain is modeled as a hierarchy of classes. This hierarchy can be represented as a tree, called inheritance tree. The nodes in the tree represent as classes.

5 Case Study

Apache-Tomcat (version 6.0.32) was chosen as the target for two reasons. First, tomcat is written in Java language. As previously mentioned, the current implementation can only handle Java language. Second, the tomcat package includes many test cases that can be used to confirm that tomcat's external behavior has not been changed due to refactoring.

6 Result Analysis

We found out the nunmber of parameter as input argument of the method and calculated the degree of coupling using above metrics. Given table shows the result of ArrayElResolver.java class, which having different methods and each method having different parameters as its argument and calculated the degree of coupling for each method.

Table 1. Result of Degree of Coupling

Method Name	No. of parameter	Degree of Coupling
getValue	3	0.1
setValue	5	0.14285714285714285
isReadOnly	3	0.1
checkBounds	2	0.07692307692307693
coerce	1	0.05623157894736842

According to the above table we also plot a graph. According to software design hypothetication high coupling is a bad design. In the graph we can see that the method setValue having 5 parameters as its argument and it has high degree of coupling.

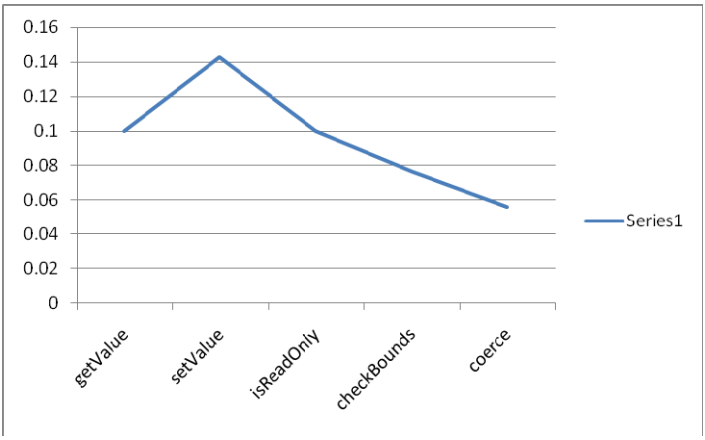


Fig. 2. Graph of Degree of Coupling

References

[1] Binkley, D., Ceccato, M., Harman, M., Ricca, F., Tonella, P.: Automated Refactoring of Object Oriented code into Aspects. In: 21st IEEE International Conference on ICSM 2005, 1063-6773/05. IEEE (2005)

[2] Garrido, A., Johnson, R.: Refactoring C with Conditional Compilation. In: 18th IEEE International Conference on ASE,1527-1366/03 (2003)

[3] Maruyama, K., Tokoda, K.: Security-Aware Refactoring Alerting its Impact on Code Vulnerabilities, 1530-1362/08. IEEE (2008), doi:10.1109/APSEC.2008.57

[4] Fowler, M., Beck, K., Brant, J., Opdyke, W., Roberts, D.: Refactoring: Improving the Design of Existing Code, 3rd edn. International Thomson Computer Press, ISBN: 978-1-4503-1098-7

[5] Rieger, M., Ducasse, S.: Visual Detection of Duplicated Code, NFS-2000-46947.96 and BBW- 96.0015, 21975. IEEE, doi:10.1109/ASPEC, 2008. 27

[6] Fenton, N.E., Pfleeger, S.L.: Software Metrics, A Rigorous & Practical Approach, 2nd edn. International Thomson Computer Press, ISBN: 978-0-470-59717-0

[7] Pressman, R.S.: Software Engineering, A practitioner's approach, 5th edn. McGRAW-Hill International Edition, ISBN:0-07-120251

[8] Counsell, S., Mendes, E.: Size and Frequency of Class Change from a Refactoring Perspective, 0-7695-3002-8/07. IEEE (2007), doi:10.1109/SE.2007.13

[9] Bryton, S., e Abreu, F.B., de e Tecnologia, F.C.: Modularity-Oriented Refactoring, 978-1-4244-2157-2/08. IEEE (2008)

[10] Hayashi, S., Tsuda, Y., Saeki, M.: Detecting Occurrences of Refactoring with Heuristic Search, 1530-1362/08. IEEE, doi:10.1109/APSEC.2008.9

[11] Tilevich, E., Smaragdakis, Y.: Binary Refactoring: Improving code behind the scenes. In: ICSE 2005, ACM, St. Louis, May 15-21 (2005) 1-581 13-963-2/05/0005

Technologies for Cost Efficient Enterprise Resource Planning: A Theoretical Perspective

Shivani Goel[1], Ravi Kiran[2], and Deepak Garg[1]

[1] Computer Science and Engineering Department
[2] School of Behavioral Sciences and Business Studies
Thapar University,
Patiala-147001
{shivani,rkiran,dgarg}@thapar.edu

Abstract. Enterprise Resource Planning(ERP) systems are used in all types of organizations. Many vendors are providing ERP software. The aim of ERP system is to reduce the costs and increase the benefits in terms of increased revenues and sales. However, all ERP implementations are found to be costly as compared to benefits gained from it. Also the costs in using the ERP creep up from time to time . This paper highlights the relation between main contributors of costs in ERP systems which include architecture, implementation, integration, maintenance, training and vendor consultation and provides ways to reduce these. Cloud computing, SOA, EAI and ASP model for ERP implementation provide a number of ways of reducing costs.

Keywords: Cost, Cloud technology , EAI, ERP architecture, ERP integration, SOA.

1 Introduction

In today's globalized economy and competitive environment, a total business solution implementation is required which is capable of supporting various business processes within the organization as well as maintains value chains with its customers as well as suppliers. Enterprise resource planning (ERP) is a software solution which integrates various functional areas of an organization using best industry and management practices. Modern Enterprise Resource Planning software is used not only in business for increasing the profits, speed up the delivery and maintaining a healthy customer relationship but also for improving operational efficiency and effectiveness of information services .All the business organizations may aim at the similar structure as is currently adopted by ERP software. ERP software has been used by capital intensive industries, such as manufacturing, construction, aerospace, defense, finance, education, insurance, retail, and telecommunications sectors. ERP has been selected worldwide for its integration capability, reputation, standard software, three-tier client/server architecture, business engineering, and migration tool from the mainframe[3].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 113–120.
springerlink.com

The ERP solutions will likely continue to define the IT standards that could enable end users to integrate most of their information systems into one cohesive technology infrastructure. The main benefits of ERP applications are improved organizational efficiency, implementation of best practices, better alignment of organizational processes, and improved data accessibility .The business objectives of ERP solutions are customer services and lower cost for the organization. All benefits though can be achieved with high cost. The aim of this paper is to identify various costs in ERP to be considered and which processes or technologies can help in cost reduction. First the issues in ERP cost estimation process are discussed, then various costs in ERP software are identified . Various ways of reducing different costs are identified. The service oriented architecture , enterprise application integration , application service provider model and cloud computing are the best technologies identified for reducing the ERP costs in a number of ways.

2 ERP Costs and Estimation

ERP planning is the first phase in the overall ERP implementation life cycle. Cost estimation is an important activity of the planning phase. Cost overrun is the one of the most critical risk in ERP implementation. Though properly planned, ERP projects end up incurring more costs than estimated. This is due to the fact that there are many hidden costs associated with ERP implementation which are ignored during formal cost estimation. ERP costs include:

1. Establishment costs (cost of platforms, cost of hardware, software, cost of interfaces and data migration requirements, deployment of cost), Hardware costs (leased, networking services costs) and Software cost (Application license costs, database license cost, OS license cost, Monitoring and management tool cost, costs of the software, external services) .
2. Recurrent costs (schedule requirements and constraints, training and associated training material and consultation services) i.e. operating costs.
3. Avoided costs (costs for platforms no longer in use or required to be decommissioned)
4. Post establishment costs (maintenance, licensing updates, version upgradation)

Though cost estimation is difficult task but it is very important in order to meet with the budget requirements. All organization are in pressure to be accurate at this. So in order to calculate the cost of an ERP system, proper plan, a process or a model should be designed to calculate the various types of costs in an organization. The process or cost model should also aim at reducing the time and effort for doing cost estimates at regular intervals. The cost estimates made should always improve as the process becomes mature. The process should also allow to compare various cost estimates. A framework developed using system dynamics simulation modelling of a case study organization can help the organizations to better predict the long-term cost of ERP systems, identify key cost drivers, and determine what dynamic relationships customizations have on total cost of ownership [6].

In order to estimate the costs there are generally three basic methods[15]: analog estimation method, top-down estimation method and parameter model estimation method.

(i) Analog estimation method : in this method, the actual costs of previous projects are used as the basis for the current cost estimate. Using this method we must carefully analyze the current projects and past projects.

(ii) Top-down estimation method (bottom-up estimating): This method works on work breakdown structure in a project .It includes estimates of individual work items and the summary of individual work items into the overall project. In order to estimate correctly, individual work items and estimated the size of the staff experience are required .

(iii) Parameter model estimation method: It is a mathematical model to estimate the project cost. It uses the project characteristics as parameters to estimate the project cost. This model if provided with accurate historical information and easy to quantify project parameters, can model the size of the project is capable of estimating the cost reliably.

In the ERP project cost estimation, emphasis should be placed on the software development cost estimates because other costs can be easily obtained in the market reference price. Along with this, the other costs for which the estimates should be included are infrastructure cost, cost of consultation by ERP vendor , cost of training the employees within the enterprise, the cost of integration, implementation and maintenance, estimated cost of licensing for first time and renewing the same, various subscription fees for shared resources and for version upgradation.

3 Ways to Reduce Costs in ERP

Various ways are suggested here for reducing the costs associated with ERP. Main costs considered are consultation cost, architecture cost, implementation cost, integration cost and maintenance cost.

3.1 Consultation Cost by ERP Expert

The professional charges payable to the outsider also depends on the extent of the services availed by the company. This has also been verified by Ziaee et al. that the costs of the consultation before the procurement are a big portion of ERP system [13]. Instead of conducting refresher programs and correcting the error during implementation, training should be done before implementation which will prove to be less costly.

3.2 ERP Architecture Cost

Most ERP systems have three distinct features in their architecture. These integrated features could facilitate compatibility between task and technology in the ERP system.

1. *Data dictionary :* which specifies thousands of domains that are associated with supporting fields and arranged in numerous tables. This data dictionary could be used across all functional areas within an organization. Once data are entered into the ERP system, it could be shared across an entire value chain in the firm.
2. *A middleware :* which could make distributed systems possible by allowing users to set up application modules and databases at different locations. Data could be moved from a central system to a remote system, permitting applications to exchange information between them. The middleware not only routes data, but also knows what data are needed in a given situation.
3. *A Data warehouse or a repository :* This is the foundation of the business framework, because the repository captures all semantics in the business processes, business objects, and organization model. It contains a comprehensive description of the ERP applications, including all meta information about models, technical programming objects, and business objects. The ERP repository is able to exchange information via application programming interfaces (API).

These three technology features are used to coordinate marketing, manufacturing, distribution, and human resources tasks in the organization. When an integrated ERP is in place, an organization can build whole enterprise applications on top of it. These enterprise applications could provide a timely feedback to enable optimal responses to changing conditions of customer demand and manufacturing capacity. All the ERP systems cannot fit as per the complete requirement of any organization in spite of the fact that they have business practice processes in their repository. So the organization needs to select those applications available from software vendors for its specific requirements, and integrate both the applications and ERP system into the organization's IT backbones.This has also been verified by that a big portion of ERP system acquisition costs are the costs of ERP architecture. Most of these costs are related to the analysis of the organizational processes and a careful determination of the required modules by ERP vendors and consultants.

A large amount of the consulting costs are saved if the modules are studied by the customer organization in the ERP software selection process. A two phase method can be used to select ERP vendor and software where the preliminary actions are forming a project team and doing business process re-engineering (BPR), collecting information about ERP software packages and vendors and filtering unqualified vendors out[13]. While in the second phase (selection phase),a modular approach for ERP vendor and software selection is presented.

3.3 ERP Integration Cost

Integration of other applications or software with ERP not easy. So the organization needs to modify or adapt the current ERP which is a costly affair[3]. The major cost factor is in integrating ERP with other applications or software. Implementation of Capability Maturity Model Integration can also help in reducing integration costs[5]. This model defines three levels of process maturity in a system: initial development process, managed development process and defined development process. Many of the ERP vendors use service oriented architecture (SOA) which is design paradigm which uses loosely coupled services. The services implemented are independent of any particular technology. This reduces the integration cost for ERP because of using open standards. Also using cloud computing, Software as a Service can help in making making ERP integration with other applications or hardware faster, easier, and less risky[9]. Enterprise Application Integration (EAI) can help in reducing integration costs to a great extent. EAI is used for application integration across multiple enterprises[14]. The sharing of business logic and information across multiple enterprises results in reduced integration costs.

3.4 ERP Implementation Cost

ERP project implementation road map: Project preparation; BPR; System developing and tuning, final testing and system go-live. This term will include all the exercises from business process engineering to gap analysis to actual restructuring, training, modifying and transferring data and systems from the old form to new form as costly affairs. The nonmonetary costs include manpower and time spent. In order to save the costs, the company can go ahead with the process of implementation with the help of In-house IT staff than engaging the services of an ERP consultant.

Using SOA, the applications are composed of common business services which can be reused and shared among many business units. This reduces the implementation cost and increases flexibility.

3.5 Training Cost

This is also a crucial determinant of ERP costs. There are two modes of training offered in companies. Companies hire trainers to update their IT staff on use of ERP. They in turn train the user to get acclimatized to ERP's functioning. This method costs less but has lot of drawbacks but still many companies go for it not only because of doing away with the need to train everyone in the company. In spite of the drawbacks this method has claimed relative success in some companies. The other method is training the users and the IT staff as well .In this method the IT staff will be trained on technical parameters while the users will be trained on usage. This method though is costly but is highly successful. Training both the users and IT staff is essential .So ERP training costs can be reduced if in-house IT staff are competent to handle other areas without training. Another solution is to have trainers in the organization itself to save the cost for training all users from outside consultants.

3.6 Infrastructure Cost

The cost of infrastructure includes the cost of hardware and IT resources required in ERP system. One way to reduce this cost is to use the cloud computing using infrastructure as s service. Using infrastructure as a service, the ERP user can store and process the information on computing resources available at another places. This reduces the infrastructure cost and allows the ERP user to deploy and run arbitrary software and hardware through virtualization. Infrastructure cost can also be reduced using platform as a service (PaaS) Platforms that can be used to deploy applications provided by customers or partners of the PaaS provider. It provide the user the capability to deploy onto the cloud infrastructure customer-created or acquired applications created using programming languages and tools supported by the provider.

Application Service Provider(ASP) model for ERP implementation can also reduce fixed costs and overall hardware cost[14]. This is done by deploying, hosting and managing access to applications to multiple parties from centrally managed data server facilities. The ERP customer need only to pay the subscription fee for network components, server-level computing hardware and software.

3.7 Maintenance Cost

Since maintenance is the longest phase of the ERP lifecycle, there is ample opportunity to improve the system in a variety of ways including business process reengineering (BPR) and extending the use of delivered functionality [12].The companies should have better insights in and control over the processes like maintenance and evolvability to improve business and software development processes in order to increase productivity, reduce costs, improve quality, and thus strengthen their position relative to competitors [1]. In order to reduce long-term maintenance costs, Business Process Reengineering (BPR) is encouraged in order to take full advantage of the ERP software but it is difficult because it requires significant enterprise-wide change management which results in high upfront costs ([7], [10]). To address evolvability during the whole lifecycle of the system and to maintain the enterprise system at reasonable costs, companies have a strong need at the level of software architecture [2] . In order to take full advantage of ERP systems, and to control TCO, ERP implementations require drastic structural and cultural changes within the organization including BPR [8].

Maintenance cost can also be reduced by using public cloud. When a public cloud is used, the ERP solution is owned and remotely hosted by the vendor. The users pay a subscription for the services they offer, (licensing model) called Software as a Service (SaaS) instead of licensing the software itself. There is no need for costs and resources required for ongoing maintenance, support and version control with SaaS which are provided by the vendor itself. Hence the maintenance cost is minimized.

A summary of technologies for reducing costs in ERP is shown in figure 1:

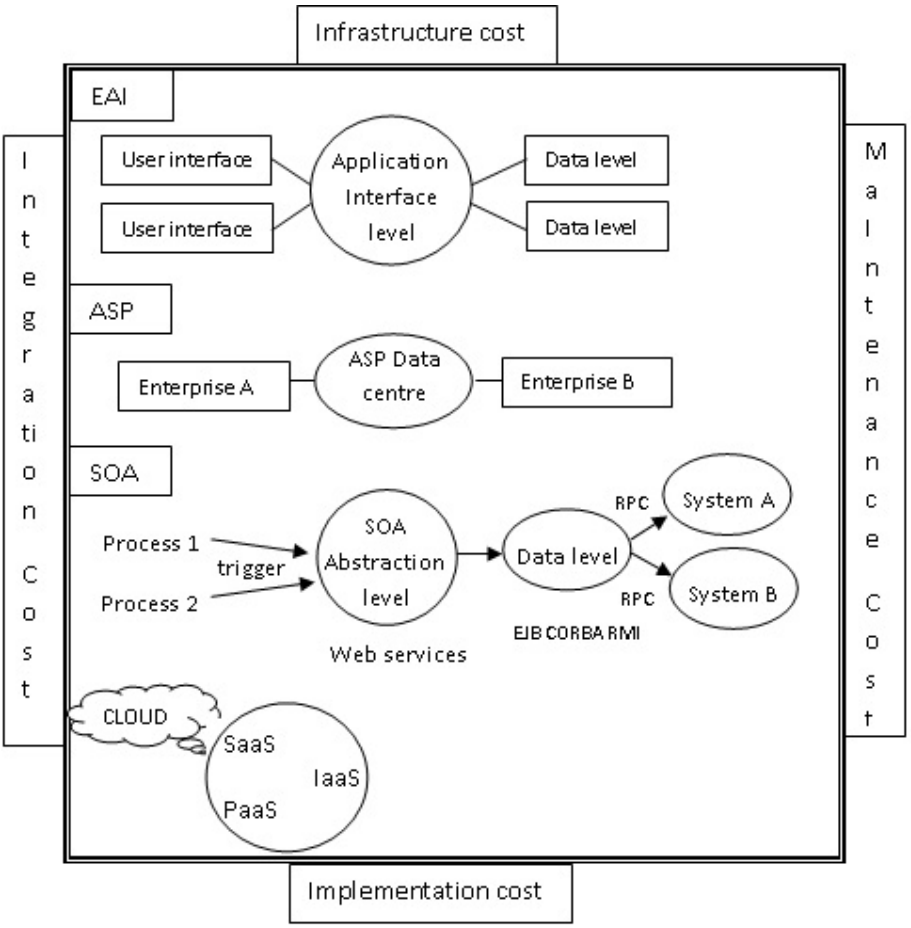


Fig. 1. Technologies for reducing costs in ERP

4 Conclusion

In order to provide lower cost for organization and customer services, continuous business reengineering is supported by ERP [4]. This was not supported by MRP systems and MRP II systems. The adoption of an ERP system brings about new changes to the organization and its information systems. The ERP system with its integrated built-in controls becomes an enabling technology for internal auditors to maintain effective controls over operations and provides assurance of reliable transaction information consistent with the organization's goals and objectives. Traditional controls, such as separation of responsibilities, will not be cost-effective in the ERP system and may not be able to deliver the required level of control.

Almost half of ERP software implementations fail from an investment perspective. A key driver for the success of an ERP implementation begins with choosing the right application with the right pricing model. The ERP system that is selected should be based on total cost of ownership and pricing models as well as for the features and functionality of the technology. Cloud technology would be referred to as an infrastructure tool that can be used to reduce hardware and IT costs while SaaS would be referred to as a deployment and business model that provides customers with a new way to purchase software and reduce cost. Similar technologies are SOA which allows the reuse of various business services using open standards to be integrated easily with less cost among various enterprises. EAI and ASP model for ERP implementation can also be used which reduce the cost of infrastructure by proving access to centrally managed network resources.

References

1. April, A., et al.: Software Maintenance Maturity Model(SMMM): The Software Maintenance Process Model. J. of Soft. Maintenance and Evol.: Res. and Prac. 17, 197–223 (2005)
2. Breivold, H.P., Crnkovic, I., Eriksson, P.J.: Analyzing software evolvability. In: Proc. of the 32nd Annual IEEE International Computer Software and Applications Conference (COMSAC 2008), July 28 -August 1, pp. 327–330. IEEE Comp. Soc., Turku (2008)
3. Chung, S.H., Snyder, C.: A ERP Adoption: A technological Evolution Approach. Int. J. of Agile Manage. Sys. 2(1), 24–32 (2000)
4. Curren, T., Keller, G.: SAP R/3 Business Blueprint. Prentice-Hall, Englewood Cliffs (1998)
5. Day, B., Lutteroth, C.: Climbing the ladder: Capability Maturity Model Integration Level 3. Ent. Info. Sys. 5(1), 125–144 (2011)
6. Fryling, M.: Estimating the Impact of Enterprise Resource Planning Project Management Decisions on Post- Implementation Maintenance Costs: A Case Study Using Simulation Modeling. Ent. Info. Sys. 4(4), 391–421 (2010)
7. Hammer, M.: Reengineering Work: Don't Automate, Obliterate. Harvard Business Review, 104–112 (July/August 1990)
8. Li, L., Markowski, E.P., Markowski, C., Xu, L.: Assessing the Effects of Manufacturing Infrastructure Preparation Prior to Enterprise Information-Systems Implementation. Int. J. of Prod. Res. 46(6), 1645–1665 (2008)
9. Goscinski, A., Brock, M.: Towards Dynamic and Attribute Based Publication, Discovery and Selection for Cloud Computing. J. of Future Gen. Comp. Sys. 26, 947–970 (2010)
10. O'Brien, J.A., Marakas, G.: Management Information Systems, 7th edn. McGraw-Hill/Irwin, New York (2006)
11. Orlicky, J.: Material Requirements Planning. McGraw Hill, New York (1975)
12. Willis, T.H., Willis-Brown, A.H.: Extending the Value of ERP. Ind. Manage. and Data Sys. 102(1), 35–38 (2002)
13. Ziaee, M., Fathian, M., Sadjadi, S.J.: A Modular Approach to ERP System Selection: A Case Study. J. of Info. Manage. and Comp. Sec. 14(5), 485–495 (2006)
14. Alshawi, et al.: Integrating diverse ERP systems: A case study. J. of Enterprise Info. Manage. 17(6), 454–462 (2004)
15. http://www.abouterp.com/erp-system-implementation/erp-project-valuation.html

Test Case Generation Using Activity Diagram and Sequence Diagram

Abinash Tripathy and Anirban Mitra

Department of CSE & IT, M.I.T.S.,
Rayagada - 765017, Odisha, India
{abi.tripathy,anir.mitra}@gmail.com

Abstract. In this paper, we present an approach to generate test cases by usingtogether the UML activity diagram and the sequence diagram. Our approach consists of transforming the sequence diagram into a graph called Sequence Graph (SG) and transforming the activity diagram to the Activity Graph (AG). Henceforth, System Graph (SYG) is formed by integrating the two graphs i.e. SG and AG. The SYG is then traversed to form the test cases. We have used DFS (Depth First Search) method as a Graph Optimization technique .for traversing the SYG.It was observed that the test cases obtained from this method is not only exhaustive but also optimized. The test case thus generated is suitable for system testing and detect the operational, interact and, scenario faults.Our approach is also capable of handling the state explosion problem in case of concurrent systems.

Keywords: Sequence Graph, Activity Graph, System Graph, DFS.

1 Introduction

In a typical software development project, more than 50% of the software development is being spent on testing in terms of time as well as finance [1]. As the complexity and size of software grow, the time required to carry out the process of testing also increases. Manual testing is time-consuming and error-prone. Therefore, automated testing processes are more preferred. The process of testing effort can be divided into three parts: test case generation, test execution and test evolution. The latter two parts are relatively easy to implement provided the proper passing condition is provided. However, the first part, i.e. to determine the test cases generation requires knowledge up to certain level.

UML is known as the language for creating Models. UML provides life-cycle support in software development and is widely used to describe analysis and design specifications of software [2]. It is a big challenge to study the test case generation from UML diagram. An activity diagram shows the flow of activity in a system. An activity is an ongoing non atomic execution in some action. Activity ultimately results in some action. Action encompasses calling another operation, sending signal, creating and destroying objects. Graphically, an activity diagram is a collection of vertices and arcs [3].The Figure 1 (a) shows the activity diagram for the card validation in ATM transaction. On the other hand a sequence diagram is an

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 121–129.
springerlink.com

interaction diagram that emphasizes the time ordering of messages. Graphically, a sequence diagram is a table that shows objects arranged along the X axis and messages, ordered in increasing time, along Y axis. Figure 2 (a) shows the sequence diagram for card validation.

In this paper, we proposed the test case generation from UML diagrams. We use sequence diagram and activity diagram as sources of test case generation. Then by using an optimization technique DFS, we try to optimize the test case that is generated in respect to the number of test cases. Our generated test suite aims to cover various interaction faults, scenario faults and operational faults.

The rest of the paper is organized as follows. In Section 2, we discuss the existing work done on test generation techniques using different UML diagrams. In Section 3, we discuss how we generate the graph from the respective UML diagrams and also propose an algorithm that generates a system Graph(SYG) by integrating both graphs. In Section 4, we present generation of test cases from SYG using the optimization technique DFS. In Section 5 we give a snapshot of the test case generated using the proposed algorithm on the given example. Finally, in Section 6 we conclude the paper with an inside to the future work.

2 Related Work

In this section, we survey the different test case generation technique using different UML diagrams.

Mall et.al [4] proposed an algorithm to generate test cases from a combination of use case and sequence diagrams. First, they convert the use case diagram in to use case diagram graph (UDG) and the sequence diagram in to Sequence Diagram Graph (SDG). Then by integrating the SDG and UDG, they generate system testing graph (STG). In the algorithm for UDG, there have mentioned that there is an edge from the use case diagram to the sequence diagram, but when the edge comes it is not clearly mentioned in the paper.

Wang et al [5] proposed a method to generate the test cases from activity diagram. In this paper, another diagram called IOAD (input/output explicit Activity Diagram) is generated from activity diagram. In the IOAD diagram only the external elements i.e. send and accept signals are represented. The non-external inputs and outputs are suppressed and the data objects like invoice and order are dropped as these objects are implicit tasks. As these fields are dropped this method does not convey the total information to the programmer, they take it as an abstract view.

3 Proposed Approach

In our proposed algorithm we convert the system under test into a graph called System Graph (SYG) which is an integration of activity graph and sequence graph. We first transform sequence diagram (SD) graph into a sequence graph (SG) [4], the activity diagram (AD) into activity graph (AG) [9] and then integrate the SG and AG to form SYG. Next, we generate the test cases by using the System Graph. In the following section, we discuss the different steps of our approach.

3.1 Transformation of AD into AG

In this section first we present, the definition of activity graph (AG). Then we discuss the methodology to generate AG from AD.

Definition 1: Activity Graph: The activity graph (AG) is defined as**AG={ A, T, F, C, V, a_i, a_f },** where

A= a finite set of activity states, T= a finite set of completion transactions, C=a set of guard condition, and C_i is the corresponding transaction t_i, F ⊂ (A X T X C)∪(A X T X C) is the flow relationship between activities and transaction.V= a condition which depend upon previous node which contain a condition. The content is null if previous node has no condition and 1 if the activity is the successful for the previous condition and 0 if it is unsuccessful.a_i∈ A, is the initial activity state. a_f∈ A, is the final activity state. In the activity diagram, at any time its current state (denoted by CS) is represented by a set of activity states.

Now, we discuss the transformation of AD to AG. Each activity in the activity diagram can be mapped as a node. A directed edge from a node A_i to $A_j$is used to represent the sequential dependency of A_i on A_j. Figure1 (a) shows the AD for the *card validation in an ATM transaction* and the corresponding AG is shown in Figure1 (b).

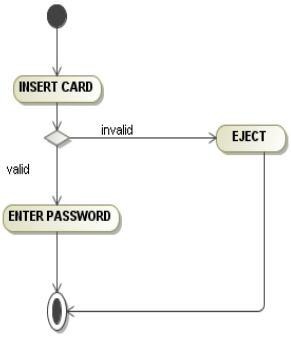


Fig. 1 (a). Activity Diagram for card validation in ATM Transaction

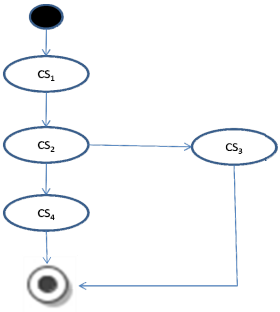


Fig. 1(b). Activity Graph for the Activity Diagram in figure 1(a)

3.2 Transformation of a SD into a SG

In this section we first present, the definition of the sequence graph (SG). Then we present the methodology to generate the SG from SD

Definition 2: sequence graph SG: The sequence graph (SG) is defined as: **SG = {State, Edge, First, Last},** where**State** set of all the nodes representing various states of a scenario.**Edge** set of edges representing transaction between different states.**First**is the initial node representing the starting state.**Last**is the final node representing the final state.

In order to formulate a method, we define a scenario as a quadruple *Scenario: < Id, Start State, message, Successful / Unsuccessful> .Id* is a unique number used to

identify each scenario. *Start State* is the starting point of Scenario i.e. where the scenario start *Message* is a set of all the events that occur in scenario. *Successful / Unsuccessful*, this state is the final state which conveys whether the system's output is successful or not. The successful/ unsuccessful condition totally depend upon the user's choice)

An event in a message can be denoted by a triplet. *Event: < message id; form; to; [/condition]*>where *message id* is a unique identification number for a particular message, *from* is the sender of the message, *to* is the receiver of the message, */condition* is the guard condition subject to which an event takes place. An event with * indicates an iterative process.

Figure 2(a) shows the sequence diagram for the card validation. It has 3 scenarios as shown in Figure 2(b). Figure 2(c) is the SDG for sequence diagram given in Figure 2(a).

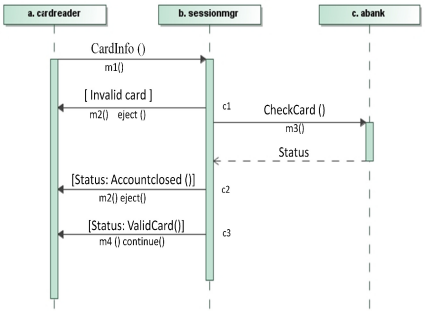


Fig. 2(a). Sequence Diagram for Card Validation

< id1	**<id2**	**<id3**
State X	**State X**	**State X**
S1 : <m1,a,b>	**S1:<m1,a,b>**	**S1:<m1,a,b>**
S2 : <m2,b,a>|c1	**S3:<m3,b,c>**	**S3:<m3,b,c>**
Unsuccessful>	**S4:<m4,b,a>|c2**	**S5:<m5,b,a>|c3**
	Unsuccessful>	**Successful>**

Fig. 2(b). Three scenarios represented in the form of Quadruples

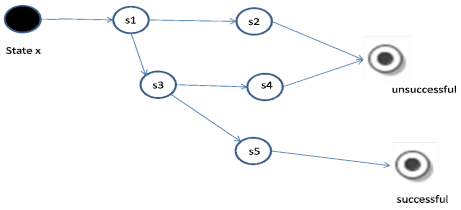


Fig. 2(c). SDG for Sequence diagram in figure 2(a)

3.3 Integrating AG and SG into SYG

After creating the AG and SG, the next step is to integrate these two graphs into a single graph called System Graph (SYG). in the following algorithm1, GEN-SYG we integrate the AG and SG to form SYG. The definition of the SYG is as follows

Definition3: system graph SYG: The sequence graph (SYG) is defined as**SYG = {State, Edge, First, Last},** where, State = $State_{SG} \cup A$ is the set of all nodes of the SG and AG

Edge = $Edge_{SG} \cup F \cup Edge_G$ where $Edge_{SG}$ is the edges of the SG, F is the flow in case of AG and $Edge_G$ is the edge from the AG to SG.

First= a_i the first node of the AG. Last = $Last_{SDG} \cup a_f$ is the final set of nodes in SYG.

Starting with AG, we integrate the SGs into it as per the definition of the SYG. Figure3 shows the SYG for the system after combination of the SG and AG.

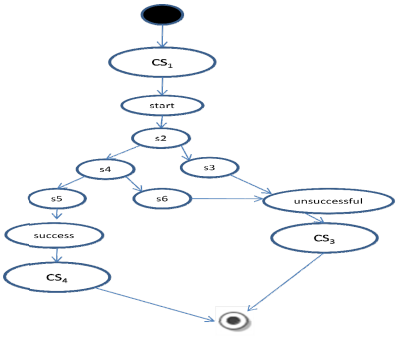


Fig. 3. SYG formed after the integration of AG and SG

Now we present the algorithm to generate SYG from AG & SG.

Algorithm1: GEN-SYG
Input: Activity Graph (AG) and Sequence Graph (SG)
Output: System Graph (SYG)

1. P = EnumerateAllPaths(AG)
2. For each path $p_i \in$ P do
3. $CS_j = CS_i$ // start with the first node
4. $preC_i$ = FindPreCond(CS_i)
5. T ←Φ
6. For each node CS_j of path p_i do
7. If $c_i \in CS_i$ // current stage has any conditional statement
8. $\alpha = CS_{i-1}$→ SG // edge from previous node to sequence graph
9. β = SG(Last) →CS_{i+1}

//edge from the lastnode of SG to next node of AG. Edge from unsuccessful final node of SG to node CS_{i+1} where the value of V = 0 else edge from Successful final node of SG to node CS_{i+1} where the value of V = 1

10. EndIf
11. T ←TU T1
12. If c_i ∈ CS_i
13. $\gamma = CS_i \rightarrow CS_{i+1}$
// there is edge from the present node to the next node of the of the same Activity Graph
14. EndIf
15. EndFor
16. End

4 Test Case Generation

After storing all essential information for test case generation using SYG, we now traverse the SYG to generate test cases. We purpose an algorithm *TCG-SYG*that automatically traverses the SYG so as to generate the test cases.

While traversing the system graph (SYG) we use DFS (Depth First Search) [8] optimization technique. As its name implies, DFS traverses "deeper" in the graph whenever possible. In depth-first search, edges are explored out of the most recently discovered vertex v that still has unexplored edges leaving it. This process continues until we have discovered all the vertices that are reachable from the original source vertex. If any undiscovered vertices remain, then one of them is selected as a new source and the search is repeated from that source. This entire process is repeated until all vertices are discovered**.**Now we present own algorithm TCG-SYG, in pseudocode form.

Algorithm 2: ***TCG - SYG***
Input: System Graph (SYG)
Output: Test Suite (T)

1. Enumerate all paths P = {P_1, P_2, P_3,…..,P_n} fromstart node to a final node in the SYG.
2. For each path P i∈P do
3. $n_j = n_x$ // n_j is the current node ; start withn$_x$the start node
4. $preC_i$ is the precondition of the node corresponding to scenario stored in n_x
5. t_i←Φ // initially the test case for the path P_i is empty
6. while ($n_j \neq n_z$) do n_z being the final node
7. Select Test Case
t= {preC, I (a_1, a_2, a_3,.., a_n), O(d_1,d_2,d_3,.,d_m), postC}
where preC = precondition of the method m
I (a_1, a_2, a_3,.., a_n) = set of input values for the method m(…)from *fromObject*
O(d_1,d_2,d_3,.,d_m)= set of resultant values in the*toObject* when the method m(…) is executed
postC = the postcondition of the method m(…)
8. Add t to the test set t_i, that is, $t_i = t_i \cup t$

9. $n_j = n_k$ // Move to the next node nk on the path Pi
10. $T = T \cup t_i$
11. Endwhile
12. Determine the final output O_i and $postC_i$for the node stored as n_z
13. $t = \{preC_i, I_i, O_i, postC_i\}$
14. Add the test case t to the test case T, that is,
    $T \leftarrow T \cup t$
15. EndFor
16. Return (T)
17. Stop

The algorithm TCG-SYG starts by enumerating all paths in the SYG, from the start node to the different final nodes. Steps 2 to 15 are iterated for each path in the SDG. Step 4 determines the initial precondition of the node from the start node n_x. For each considered path, Steps 7 to 11 determine the various pre conditions, input, output and post conditions for each interaction of the considered scenario. This gives the test cases for finding out interaction faults if any. And finally Step13 gives the test case corresponding to the scenario as a whole.

The test cases generated using theses algorithms and for the case *ATM card Validation* is shown in Figure 4.

Test name: = “ATM CARD VALIDATION”
Precondition : ATM is displaying an welcome message, Ask the User to enter the ATM card

Test case 1
Input: user input ATM card
Condition: Not a valid ATM card
Output: Eject card
Postcondition : display the welcome message

Test case 2
Input: Card = ATM, status =“ valid”
Condition: Account =“ closed”
Output: Eject card
Postcondition : display the welcome message

Test case3
Input: Card= ATM, status = “valid”, Account=“open”
Output: Display “enter pin”
Postcondition: pin is enter and checked for validity

Fig. 4. Test cases generated using GEN-SYG and TCG-SYG for the case ATM card Validation

5 Conclusions and Future Work

We have proposed an approach to use activity diagram and sequence diagram as UML diagram for generating test cases. We convert the diagrams into intermediate representations called System Graph (SYG), which is an integration of intermediate representation of activity diagram and sequence diagram. The test cases obtained in this method is exhaustive i.e. no more valid test cases can be generated apart from the test cases generated in this method. In the activity diagram, a conditional statement is required for having the possibility of multiple paths. The solution of the statement leads us to the optimum result. Sequence diagram represent the various interactions possible between the objects during the operation. For developing the sequence diagram, an experienced developer will consider all the cases. So whenever we integrate the two UML diagrams it will cover all the possibility. Apart from this characteristic, the system is able to solve faults like operational, integration and scenario faults using sequence diagram. Activity diagram also solves the problem of concurrent execution problem which leads to state explosion problem.

In this paper while traversing the graph, we use the method of Depth First Search (DFS) as the method of traversing each graph. As DFS is one of the optimization algorithm while graph traversal, so that the test cases obtained in this methods are not only exhaustive but also optimum.

During the process of test case generation, we have tried to solve many problems but still some of the problems related to combination of activity and sequence diagram remained unsolved. In our future work, we will try to combine one or more UML diagrams with this existing system, so that the system is able to handle all types of error. This step will lead us to develop a generalized method. Further, we have only used the DFS algorithm to optimize the test cases generation. A further analysis on results and performance of our model using other graph optimizing technique will be an interesting work to carry out in future.

References

1. Mall, R.: Fundamental of Software Engineering, 2nd edn. Prentice-Hall of India Private limited, New Delhi (2004)
2. Object Management group, UML Specification 1.5 (2000), http://www.omg.org/uml
3. Booch, G., Rambaugh, J., Jacobson, I.: The United Modeling Language User Guide. Addition-Weseley (2001)
4. Mall., R., Sarma, M.: Automatic Test case Generation from UML Models. In: The Proceeding of IEEE Conference on Software Maintenance (2007)
5. Wang, L., Yuan, J., Yu, X., Hu, J., Li, X., Zheng, G.: Generating Test cases from UML Activity Diagram based on Gray-box Method, National Natural Science Foundation of China. National Natural Science Foundation of China (2005)
6. McGregor, J.D., Sykes, D.A.: A practical guide to testing object-oriented software. Addition Wesley, NJ (2001)
7. Binder, R.V.: Testing Object-Oriented System Models, Patterns and Tools. Addition-Wesley, NY (1999)

8. Coreman, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithm, 2nd edn. The MIT Press. McGraw-Hill Book Company, Massachusetts, London
9. Kim, H., Kang, S., Baik, J.: Test Case Generation from UML Activity Diagram. In: Proceeding of 8th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. IEEE Computer Society (2007)

User Authentication Using Keystroke Recognition

Urvashi Garg and Yogesh Kumar Meena

MNIT, Malaviya Nagar JLN Marg Jaipur
{urvashi.garg.24,yogimnit}@gmail.com

Abstract. This Paper represents technique for user authentication using keystroke dynamics. In this paper we have included Inter key time, Key hold time as well as some other keystroke features to verify the user. As the user types a combination of string, its key hold time and inter key time is noted, which is then compared with trusted user values using Euclidean distance. Then token system is used to check whether user is valid or not. It has improved user authentication such that Type 2 error is 0 % but type 1 error is 30 %.

1 Introduction

With the increasing use of Computer in every field the main issue that arises is the security of the system. Various techniques like user id and passwords are used for security of system but it is easy to shoulder surf or spoof those passwords [7]. After that it was decided to use randomly generated passwords but that was very difficult to remember [6]. Keystroke analysis is the cheapest method for identification of user. We can use timing constraints on user's typing pattern. As the user types a particular pattern its key hold time and inter key time is recorded that can be used for user authentication [2]. This will help us to distinguish different users. Six features are used in order to authenticate the valid user. These feature sets are used in classification method for identifying valid user. In order to evaluate the performance of the feature sets we can check error values. Two types of errors are possible in this case: Type 1 error in which decision is wrong but the person is right and Type 2 error is that in which system is unable to identify false user and recognize it as valid user [2].

2 Related Work

A number of methods have already been developed for keystroke authentication.

In [1] P. Campisi et al. proposed that keystroke dynamics can be used for authentication over mobile devices. In [2]Features of keystroke dynamics are explained by Heather Crawford in order to further improve the methods. In [3] Deian Stefan and Danfeng Yao focused on TUBA framework to detect synthetic forgery attack.

In [4] ukree Sinthupinyo et al. used Back propagation method over non fixed length string for keystroke dynamics. In [5] Obaidat and Balqies Sadoun proposed that several neural network techniques can be applied on inter key and key hold time for user authentication. In [6] Mariusz Rybnik et al. used keystroke dynamics with short fixed text. In [7] Himon Modi et al. solved the problem of spontaneously

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 131–136.
springerlink.com

generated password authentication as well as focused on error. In [8] Sungzoon et al. considered Inter key time and Key hold time with password authentication to get more accuracy.

3 Proposed Set of Features

In this paper we have used some additional features for user authentication. We have applied Euclidean distance between data collected from trusted users and incoming user's typing characteristics and find most suitable match using token based approach. We have used following features.

3.1 Inter Key Time

Time elapsed between release of first key and pressing of second key.

3.2 Key Hold Time

Time elapsed between any key press and it's release.

3.3 Key Type Change Time

It is basically the Inter Key Time. It is of two types, first in which first key is alphabet and second is other than alphabet, second in which first is other than alphabet and second key is alphabet.

3.4 No. of SHIFT

It is the no. of times SHIFT key is used while typing the pattern.

3.5 No. of BACKSPACE

It is the no. of times BACKSPACE key is used while typing the pattern.

3.6 No. of CAPS

It is the no. of times CAPS key is used while typing the pattern.

4 Background Details

The main focus of this paper is to develop a technique for user verification that is simple, robust and cheap. Keystroke recognition is such type of technique. It is a good authentication tool but not as powerful as it should be. This section will cover: (A) Types of Errors, (B) Pattern Typing Characteristics and (C) Classification Method.

4.1 Types of Errors

Main concept behind Keystroke recognition is to detect the validity of user depending upon typing style and timing. The result of classification can be measured using two types of error: Type 1 error i.e. authorized user is not verified and is not allowed to access resource; Type 2 error i.e. unauthorized user is allowed access to resources. Values of Type 1 and Type 2 errors depends upon the sensitivity of resource. If the value of Type 1 error is more than system is over secure and it requires a lot of authentication for trusted user to get access to the resource and if Type 2 error is more than system is not so secure because it can grant access to untrusted users. In this way we can say that Type 1 error is acceptable up to a certain limit. These two types of error cannot be removed completely because one error decreases if other increases. Since both of these are inversely proportional to each other, therefore it is very difficult to build a system that is completely prevented from these two errors.

4.2 Pattern Typing Characteristics

Since we can verify a user by using only user name and password but password can be easily stolen by anyone. Therefore we have used this password protection scheme with Dynamic entry of text i.e. we will determine the typing characteristics of user. Firstly we use Inter key time and Key Hold time. As given in figure. Another terms are key symbol type change time. It is of two types first one is the time in which alphabet is pressed before any other character and second in which alphabet is pressed after other symbol. Another characteristic is whether user has used SHIFT key or CAPS key for typing capital letters and how many times it has used BACKSPACE.

4.3 Classification Method

We can use any distance algorithm or any other clustering algorithm for this purpose but we have used Euclidean distance algorithm in which we have calculated the difference between timing values of trusted users and incoming user.

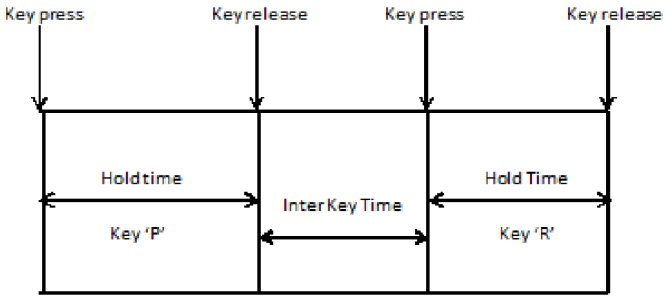


Fig. 1. Pattern Characteristics

5 Data Collection

Now our next step is to collect the data that consist of all pattern features values for all trusted users. For this we have prepared a form that consists of two columns one for entering user name and second consists of string that has to be entered by user. During entering of that string in text box it will calculate number of times SHIFT or CAPS is pressed and number of times BACKSPACE is used as well as it will calculate Inter key time, Key Hold time and Key type change time. We have taken string for experiment ASdffgh123jkfdsdSDuytr. Now the final dataset will consist of keystroke latencies depending upon the key combination given in Table 1-2. After getting all Inter key, Key Hold, Key type change time; their average is calculated and all these values are placed in a separate file for future reference.

Table 1. Inter Key time and Key Hold Time Key Combination

For Inter Key Time	For Key Hold Time
AS	A
SD	S
Df	d
ff	F
fg	f
gh	g
h1	h
12	1
23	2
3j	3
jk	j
kf	k
fd	f
ds	d
sd	s
dS	d
SD	S
Du	D
uy	u
yt	y
tr	t
	r

Table 2. Key Type Change Key Combination

Type 1	Type 2
h1	3u

6 Experiment and Results

We have to perform an experiment to check the validity of user. For this we have prepared a form that consists of multiple columns. First one is for user name and then for password entry. If user has entered user id then he has to enter password and for wrong password he has given three chances and if again no correct password is entered then form disappeared saying that you are not allowed, otherwise first string is enabled which has to be entered by user and this procedure go on up to five strings. For all these five strings we have calculated key hold time, inter key time, key type change time as well as number of times user have pressed SHIFT, BACK and CAPS. After that their average values are calculated which were then compared with each key hold, inter key, key type change time of all trusted users list and their Euclidean distance is calculated and minimum Euclidean distance value is given as maximum token and that user is taken as suspected user which is then compared for number of SHIFT or CAPS used and if they match then number of BACKSPACE is compared and some constraints are applied if all these conditions matched then user name is compared with that user if it matches then only he is granted access otherwise not. Results are given in table 3. It is clear from the results that out feature space method achieved type 2 error as zero percent.

Table 3. Accuracy Achieved At Various Combinations

Attributes	Accuracy in %	Type 1 Error in %	Type 2 Error in %
Key Hold + Inter Key Time	30	40	70
Key Hold Time + Inter Key Time + Key Type Change Time	50	35	50
Key Hold Time + Inter Key Time + Key Type Change Time + no. of SHIFT or CAPS used	60	35	30
Key Hold Time + Inter Key Time + Key Type Change Time + no. of SHIFT or CAPS used + no. of BACKSPACE used	70	30	0

7 Conclusion and Future Work

Previous researches shows the static level of analysis and have only a limited range of accuracy but as we have used dynamic analysis which shows a great accuracy level because in this we have reached to 0 % type 2 error and type 1 error is also minimized and it reaches to high accuracy level. But it can further be improved. If we use more number of attributed to classify pattern then accuracy level can be increased. We can also use randomly generated password to check the timing constraint and also for collection of data we can used a single string that is generated randomly. Therefore further research must be carried out to get full benefits of keystroke dynamics.

References

1. Campisi, P., Maiorana, E., Lo Bosco, M.: User Authentication Using Keystroke Dynamics in Cellular Phones. IET, Signal Processing, pp. 333–341 (2009) ISSN: 1751-9675
2. Crawford, H.: Keystroke Dynamics- Characteristics and opportunity, pp. 205–212 (2010) ISBN: 978-1-4244-7551-3
3. Stefan, D., Yao, D.: Keystroke-dynamics authentication against synthetic forgeries, pp. 1–8 (2010)
4. Inthupinyo, S., Roadrungwasinkul, W., Chantan, C.: User recognition via keystroke latencies using SOM and Back propagation neural network, pp. 3160–3165 (2009) ISBN: 978-4-907764-34-0
5. Obaidat, M.S., Sadoun, B.: Verification of Computer Users Using Keystroke Dynamics. IEEE Transactions on Systems Man and Cybernetics 27(2), 261–269 (1997)
6. Rybnik, M., Panasiuk, P., Saeed, K.: User Authentication with Keystroke Dynamics Using Fixed Text. In: International Conference on Biometrics and Kansei Engineering, pp. 70–75 (2009) ISBN: 9780769536927
7. Modi, S., Elliott, S.J.: Keystroke Dynamics Verification using Spontaneously Generated Password Carnahan Conferences Security Technology, pp. 116–121 (2006) ISBN: 1424401747
8. Cho, S., Han, C., Han, D.H., Kim, H.-I.: Web based Keystroke Dynamics Identity Verification using Neural Network. Journal of Organizational Computing and Electronic Commerce 10(4), 295–307 (2000)

On Decidability and Matching Issues for Regex Languages

Praveen Alevoor, Pratik Sarda, and Kalpesh Kapoor

Department of Mathematics,
Indian Institute of Technology Guwahati, India
{p.sarda,a.praveen,kalpesh}@iitg.ernet.in

Abstract. Several programming languages, such as Perl and Java, provides an extended variation of regular expressions to match patterns in a text. We study one such extension, Regex expression, that includes back-reference operator in addition to standard regular expression operations. Although matching a regular expression can be done in polynomial time, it is known to be NP-complete for Regex expressions. We study decidability properties for Regex languages and show that it is not possible to improve matching complexity based on semantic analysis of Regex expressions. We also give an algorithm for matching that is efficient for certain types of Regex. We compare our results with another algorithm proposed earlier.

1 Introduction

The theory and applications of finite automata and corresponding regular languages has been an extensively studied topic. Regular expression is a convenient notation to describe regular languages and they closely relate to Non-Deterministic Finite Automata (NDFA). However in practice many applications, such as Perl and Java, use an extended notation for regular expression by introducing several operators that are not studied in theory. These new operators often allow to express languages that are no longer regular.

This paper is concerned with study of such an extension called as Regex expressions (or simply Regex). We assume that the reader is familiar with finite automata and regular expressions. For details, we refer to [7].

Regex are an extension of regular expressions that are used to match patterns in applications.

Definition 1.1 ([2]). Let $\sum$ be a finite alphabet. A Regex over $\sum$ is a well-formed parenthesized formula, consisting of operands in $\sum^* \cup \{\backslash i | i \geq 1\}$, the binary operators . and +, and the unary operator * (Kleene star). By convention, () and any other form of "empty" expression is a Regex denoting ϕ (consequently, $()^*$ will denote ϵ). Besides the common rules governing regular expressions, a Regex obeys the following syntactic rule: every control character $\backslash i$ is found to the right of the i^{th} pair of parentheses, where parentheses are indexed according to the occurrence sequence of their left parenthesis. ◊

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 137–145.
springerlink.com

An example of Regex is $({}_1a + b)^*\backslash 1$. We number the brackets to make it easy to match them with corresponding \ operator (also referred to as back-reference operator).

Given a Regex, r, the language, $L(r)$, represented by r is the set of all words matching it in the sense of regular expression matching, with the following additional semantic rules:

I. During the matching of a word with r, a control $\backslash i$ should match a sub-word that has matched the parenthesis i in r. There is one exception to this rule:
II. If the i^{th} pair of parentheses is under a Kleene star and $\backslash i$ is not under the same Kleene star, then $\backslash i$ matches the content of the pair of parentheses under the Kleene star, as given by its last iteration.

For example, $L(({}_1a + b)^*\backslash 1) = \{ww \mid w \in (a + b)^*\}$. Note that this language is not context-free.

Given a Regex, r, and a string, s, the matching problem requires to answer the question whether $s \in L(r)$? The membership problem for regular expression is solvable in polynomial time [9]. On the contrary, the Regex matching problem is known to be NP-Complete [1]. Thus, we look for possible improvements that guarantees to perform well on certain class of Regex. In particular, we address the following questions in rest of the paper.

Semantic Classification. Given a Regex, r, is it decidable that $L(r)$ is context- free? (Section 2).

Syntactic Classification. Is it possible to identify syntactic structures in Regex such that their presence will guarantee matching to be done in polynomial time? (Section 3).

Finally, conclusions are presented in Section 4.

2 Decidability for Regex Languages

Given a Regex, r, is it possible to design an algorithm that tells whether $L(r)$ is context-free? The motivation for asking this question is that if it is feasible to design such an algorithm, then an existing algorithm for parsing Context Free Grammars (CFG) can be used for Regex matching. Since parsing problem for a CFG is solvable in polynomial time, it will be guaranteed that at least for those Regex that defines context-free languages, the matching can be done in polynomial time.

Our proofs about decidability of different properties of Regex languages are based on the following theorem.

Theorem 2.1 ([6]). Let $\mathcal{F}$ be effectively closed under union and under concatenation by regular sets and let "$L_1 = \sum^*$" be undecidable for $L_1 \in \mathcal{F}$. If P is any property that is defined on $\mathcal{F}$ and (a) is false for at least one L_2 in $\mathcal{F}$, (b) is true for all regular sets, (c) is preserved by inverse gsm[1], union with ϵ, and intersection with regular sets, then P is undecidable for $\mathcal{F}$.

Let $\mathfrak{R}$ be the set of Regex languages. We know that $\mathfrak{R}$ is closed under union and concatenation with regular sets. We also know that universality property of Regex is undecidable from [3]. Therefore, $L_1 = \sum^*$ is undecidable for L_1 in $\mathfrak{R}$. Thus, for any property P satisfying the above three conditions, P will be undecidable for $\mathfrak{R}$. The following lemma is proved in [3]. We give an alternative proof of the same.

Lemma 2.1 ([3])**.** It is undecidable whether the language of given Regex is regular.

Proof. Let regularity be a property defined on $\mathfrak{R}$. Let $L = \{a^n ba^n | n \geq 0\}$, then $L \in \mathfrak{R}$ and L is not regular. Regularity is true for all regular sets, preserved by inverse gsm, union with ϵ, and intersection with regular sets. Thus, by Theorem 2.1, it is undecidable whether the language of given Regex is regular. □

Lemma 2.2. It is undecidable whether the language of given Regex is linear context-free.

Proof. Let linear context-free be a property defined on $\mathfrak{R}$. Let

$$L = \{a^n ba^n c^m dc^m \mid n \geq 0, m \geq 0\}$$

then $L \in \mathfrak{R}$ and L is not linear context-free. Linear context-free is true for all regular sets and is preserved by inverse gsm [5]. It is also preserved by union with ϵ, and intersection with regular sets. Thus, by Theorem 2.1, it is undecidable whether a Regex language is linear context-free. □

Lemma 2.3. It is undecidable whether the language of given Regex is deterministic context-free.

Proof. Let deterministic context-free be a property defined on $\mathfrak{R}$. Let $L = \{a^n ba^n ba^n \mid n \geq 0\}$, then $L \in \mathfrak{R}$ and L is not deterministic context-free. All regular sets are deterministic context free languages, and deterministic context free property is preserved by inverse gsm, union with ϵ, and intersection with regular sets [4]. Thus, by Theorem 2.1, it is undecidable whether the language of given Regex is deterministic context-free. □

Lemma 2.4. It is undecidable whether the language of given Regex is context-free.

Proof. Let context-free be a property defined on $\mathfrak{R}$. Let

$$L = \{a^n ba^n ba^n | n \geq 0\}$$

then $L \in \mathfrak{R}$ and L is not context-free. All regular sets are context free languages and context free property is preserved by inverse gsm [5], union with ϵ, and intersection with regular sets [7]. Thus, by Theorem 2.1, it is undecidable whether the language of given Regex is context-free. □

Using Greibach theorem, similar proofs can be given about undecidability of complement of Regex to be regular or context-free. Thus, we conclude that it is not

directly feasible to analyze a Regex as context-free and then use an existing algorithm for parsing CFG to speed up Regex matching. In other words, the semantic classification of Regex expressions is not possible.

3 Regex Matching Algorithm

In this section, we give an algorithm for matching Regex patterns. In [8] based on a syntactic analysis of Regex, an algorithm for matching is presented. We compare our algorithm with that given in [8].

In order to perform matching, a Regex is converted into a suitable format and a machine is built using the same, which is used for matching a given string. The rules of parenthesis indexing used for the Regex are as follows [2]:

a) Entire expression is enclosed by a pair of parentheses.
b) Inner pairs of parentheses have an index smaller than those of the parentheses that surrounds them.
c) If two pairs of parentheses are enclosed in an outer parentheses and are not nested, then the left pair has a higher index than the right one.

For a Regex, r, following the above mentioned rules of parentheses indexing, every pair of parenthesis u_k is associated with a Regex, r_k, over $\sum_k = \sum \cup \{u_1, v_1, ..., u_{k-1}, v_{k-1}\}$, where v_k denotes the back-reference operator for the k^{th} parenthesis $\backslash k$. The machine for the Regex $r(= r_n)$ is constructed as follows:

a) The machine for v_k is constructed with a start state O_{vk} and a final state F_{vk}. The inner construction for the machine is determined dynamically.
b) The machine A_k for u_k is constructed using the machines of u_1 to u_{k-1} and v_1 to v_{k-1} as done in [1] using NDFA's.

The input string, s, is matched with the Regex, r, using a Regex Machine which consists of the machine (A_n) constructed above, two stacks and a counter. Configuration of Regex Machine has the form $(w,c,\Gamma^{(1)},\Gamma^{(2)})$, where,

- w is the portion of input string which is yet to be evaluated, $w \in \sum^*$.
- c is a counter which denotes the number of characters which have been evaluated.
- $\Gamma^{(1)}$ is a stack containing the current set of states.
- $\Gamma^{(2)}$ is a stack containing the set of states reachable from the states in $\Gamma^{(1)}$ using the next character to be evaluated.

The initial Configuration of the Regex Machine is $(s,0,O_n^x,\phi)$, where s is the input string and O_n is the start state of $r_n(= r)$ and x is a array of length $2(n - 1)$ filled with -1.

Type of Transitions

System changes it's configuration in one of the six ways,

$$(\alpha w, c^{(t)}, \Gamma^{1,(t)}, \Gamma^{2,(t)}) \rightarrow (w, c^{(t+1)}, \Gamma^{1,(t+1)}, \Gamma^{2,(t+1)})$$

These are shown in Table 1.

Table 1. Types of Transition for Regex Machine. *Ar*(*i*) indicates i^{th} element of *Ar*

E transition	**letter transition**	**stack switch**
$\alpha = \epsilon$	$\alpha \in \sum$	$\alpha = \epsilon$
$s^x = pop(\Gamma^{1,(t)})$	$s^x = pop(\Gamma^{1,(t)})$	$\Gamma^{1,(t)} = \phi$
$s \in Q^k$	$s \in Q^k$	$\Gamma^{1,(t+1)} = distinct(\Gamma^{2,(t)})$
$q \in \delta_k(s, \epsilon)$	$q \in \delta_k(s, \alpha)$	$\Gamma^{2,(t+1)} = \phi$
$y = x$	$y = x$	$c^{(t+1)} = c^{(t)}$
$y(c_{k,1}) = c$ if $s \in O_k$	$y(c_{k,1}) = c$ if $s \in O_k$	-
$y(c_{k,2}) = c$ if $s \in F_k$	$y(c_{k,2}) = c$ if $s \in F_k$	-
$push(q^y, \Gamma^{2,(t+1)})$	$push(q^y, \Gamma^{2,(t+1)})$	-
$c^{(t+1)} = c^{(t)}$	$c^{(t+1)} = c^{(t)} + 1$	-
rtb	**btr**	**I transition**
$\alpha = \epsilon$	$\alpha = \epsilon$	$\alpha = input(i+1)$
$s^x = pop(\Gamma^{1,(t)})$	$s^x = pop(\Gamma^{1,(t)})$	$s^x = pop(\Gamma^{1,(t)})$
$s = O_{vk}$	$s = I_{i,i,k}$	$s = I_{i,j,k}$
$q = I_{x(c_{k,1}),x(c_{k,2}),k}$	$q = F_{vk}$	$q = I_{i+1,j,k}$
$y = x$	$y = x$	$y = x$
$push(q^y, \Gamma^{2,(t+1)})$	$push(q^y, \Gamma^{2,(t+1)})$	$push(q^y, \Gamma^{2,(t+1)})$
$c^{(t+1)} = c^{(t)}$	$c^{(t+1)} = c^{(t)}$	$c^{(t+1)} = c^{(t)} + 1$

Dynamics of Regex Machine

Each state of Regex Machine is of the form Q^x, where *x* is an array of the form $(c_{1,1},c_{1,2},...,c_{(n-1),1},c_{(n-1),2})$ and of length $2(n-1)$. Here $c_{k,1} + 1$ to $c_{k,2}$ represents the portion of input string to be matched by v_k and this array is updated during the evaluation. The *epsilon*-function is used to compute the reachable states from any set of states using only ϵ-transitions [1]. To perform this, a simple reachability algorithm is used. The *epsilon*-function uses E transition, rtb and btr given in Table 1. The *goto*(*Q*,*c*) function is used to compute the set of states reachable from any state using letter *c* and it uses the letter- and I-transitions given in Table 1. The *SS* function performs stack switch. Final configuration of the Regex Machine is $(\epsilon,|s|,\Gamma^{(1)},\Gamma^{(2)})$, where $|s|$ is the length of the input string. The string is accepted only if $\Gamma^{(1)} \cap F_n \neq \phi$. These steps are summarized in Table 2.

Table 2. Execution of Regex Machine

```
begin
    Q := SS(epsilon(O_n))
    for j = 1 to |s| do
      begin
          Q := SS(epsilon(SS(goto(Q, s_j))))
      end
    if Q ∩ F_n ≠ ∅ then return yes
    else return no
end
```

Optimization of array size

The optimization is based on reducing the size of the array x, used with every state of the Regex Machine. The terminology required for the optimization is as follows.

Interval: Interval for the i^{th} parentheses is defined as the portion of a Regex starting from the i^{th} left parenthesis and ending at the last back-reference of that parenthesis. It is not defined for those parentheses for which back-references do not exist.

Maximum Overlapping Intervals (*MOI*): It is defined for a Regex as the maximum number of overlapping intervals at any point in the Regex. The reduced length of the array, which is based on the *MOI* is computed using the algorithm given below.

I. Read the Regex and store the indexes of parentheses which have at least one back-reference in the Regex.
II. Create an array of i's and \$i$'s placed according to the position of the i^{th} left parenthesis and last i^{th} back-reference operator in the Regex. For example for $r_1r_2\backslash 1\backslash 2r_3$ the array will be 1,2,\1,\2.

Read the array created in the previous step and allot the smallest free natural number to the index i when it is encountered and free that index i when \i is encountered.

Let $k(i)$ be the natural number allocated to the index i in the previous step. Then, in the array x associated with the states of Regex Machine, $c_{i,1}$ refers to the $2k(i)$ index of the array and $c_{i,2}$ refers to the $2k(i) + 1$ index of the array.

Let p be the maximum number of natural numbers used, then the array size is reduced to $2p$. Here $MOI(r) = p$.

An Example

Consider a Regex $r = (a + b)^* a\backslash 1$ and an input string $w = abab$ to be matched. Applying the rules of parentheses indexing and rewriting the Regex r we get, $r = ({}_2({}_1a + b)^* a\backslash 1)$, $r_1 = a + b$ and $r_2 = r_1^* a v_1$. We do not require the optimization steps for this example as only one back-reference is present in r. The machines A_1 and A_2 for r_1 and r_2 constructed using the above rules are shown in Figure 1 and 2, respectively.

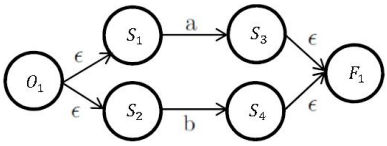


Fig. 1. Machine A_1 for regular expression, r_1

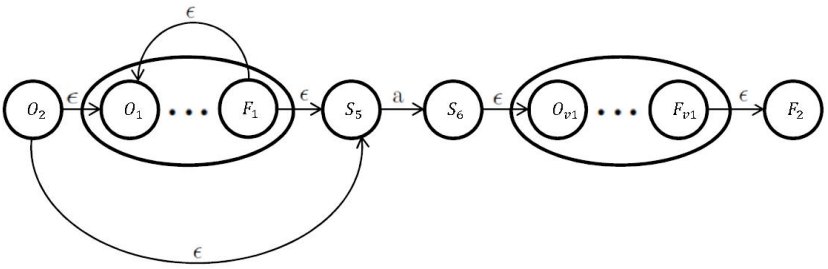


Fig. 2. Machine A_2 for regular expression, r_2

The current set of states in the stack after every transition is shown Figure 3. The stack contains the final state after the string is consumed, hence the string is accepted.

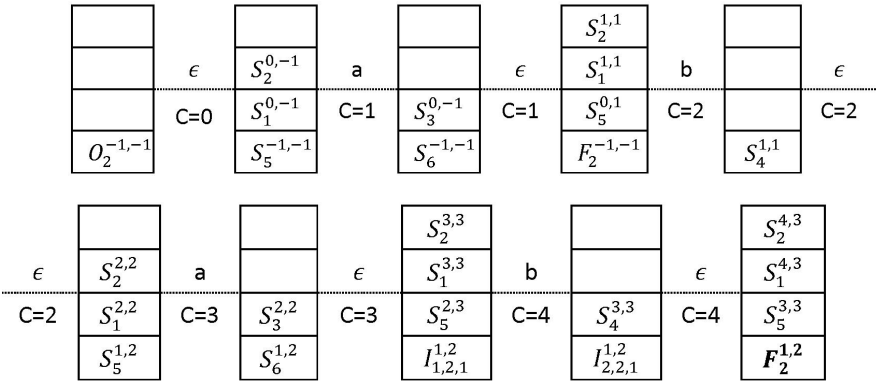


Fig. 3. Stack contents after every transition for the example

Correctness

The correctness of matching algorithm is straight forward as NDFA's of the regular expressions match the string correctly in every round of evaluation. The back-references are matched correctly, as the index of the string matched by every parenthesis is stored and updated as needed and these indexes are used to match the back-references. Thus, without optimization, the matching is correct.

For the correctness of the optimization, if a natural number $k(i)$ is associated with only one index i then $c_{i,1} + 1$ to $c_{i,2}$ represents the portion of input string to be matched by \i and thus, this is same as what was being done before.

If a natural number $k(i)$ is associated with more than one index, for instance $k(i) = k(j)$ and $i > j$ then in the Regex, parenthesis j starts after the last back-reference operator of parenthesis i, thus the indexes $2k(i)$ and $2k(i) + 1$ can be used by both i and j indexes of the Regex.

Complexity

Let r be the Regex, $|r|$ the length of r, M the total number of possible states in NDFA of r, n the number of parentheses in r and s the length of the input string to be matched. The time complexity can be analyzed as follows.

Maximum number of states of the form $I_{i,j,k}$ is $s*s*n$. Length of array on each state is $2MOI(r)$ thus, maximum possible states is $O((M + ns^2)s^{2MOI(r)})$. Each stack in the Regex Machine contains distinct states thus having a maximum size of $O((M + ns^2)s^{2MOI(r)})$. A total of $O(s)$ computations are performed to match a string of length s. Thus the time complexity is $O((M + ns^2)s^{2MOI(r)+1)})$.

Every state in the machine has at most two out-going states, M is at most equal to $2|r|$. Thus, the time complexity is $O((|r| + ns^2)s^{2MOI(r)+1})$. We can write $O(|r| + ns^2)$ as $O(ns^2)$ and hence the time complexity will be $O(ns^{2MOI(r)+3})$.

The space complexity can be analyzed as follows. Since every state in the Regex Machine has an array of length $2MOI(r)$, the total space required by the stack is $O(MOI(r)ns^{2MOI(r)+2})$.

An Example

Let Regex $r = ({}_5(_4a)\backslash 4({}_3b)\backslash 3({}_2a)\backslash 2({}_1b)\backslash 1)$. Optimization is performed on the given Regex as described below.

a) Indexes 1, 2, 3 and 4 have back-references in the Regex.
b) Create the array 4,\4,3,\3,2,\2,1,\1.
c) Processing the above array, $k(4) = 1$, 1 is free, $k(3) = 1$, 1 is free, $k(2) = 1$, 1 is free, $k(1) = 1$, 1 is free.
d) $c_{1,1}$, $c_{2,1}$, $c_{3,1}$, $c_{4,1}$ all point to first position in array and $c_{1,2}$, $c_{2,2}$, $c_{3,2}$, $c_{4,2}$ all point to second position in the array.
e) $MOI(r) = 1$, thus size of the array is 2.

Hence the time complexity of matching for this example is $O(s^5)$.

3.1 Comparison with Earlier Work

The above algorithm creates sub-classes of Regex languages for which the membership problem can be solved efficiently using parameter *MOI*. Similar class has been created by Reidenbach in [8] using a parameter known as variable distance. Both the algorithms outperform each other under certain cases.

For example a Regex $r = r_1r_2\backslash 2r_3\backslash 3r_4\backslash 4r_5\backslash 5r_6\backslash 6\backslash 1$. Here $MOI(r) = 2$ and $vd(r) = 5$. Therefore complexity of our algorithm is $O(ns^7)$. While the complexity using Janus automata is $O(|r|^3 s^{(vd(r)+4)}) = O(|r|^3 s^9)$.

On the contrary for the Regex $r = r_1r_2r_3r_4\backslash 1\backslash 2\backslash 3\backslash 4$. Here $MOI(r) = 4$ and $vd(r) = 4$. Therefore complexity of our algorithm is $O(ns^{11})$. While the complexity using Janus automata is $O(|r|^3 s^{(vd(r)+4)}) = O(|r|^3 s^7)$. A hybrid algorithm based on these two algorithms can be constructed by calculating the *MOI* and *vd* of the given Regex and then using the appropriate algorithm based on them.

4 Conclusions

We have shown that Regex matching algorithms cannot be improved based on the semantic classification of Regex as regular, linear-context free or context free languages. This is done by proving that these properties are undecidable for Regex expressions.

We introduce the notion of Regex Machines and use it to design an algorithm to match Regex patterns, which runs in polynomial time for a certain class of Regex. We have compared our algorithm with an existing algorithm [8] and provided examples where each algorithm performs well. We suggest to use a combination of the two algorithms for pattern matching of Regex expressions.

References

1. Aho, A.V.: Algorithms for finding patterns in strings, pp. 275–300. MIT Press, Cambridge (1990)
2. Câmpeanu, C., Santean, N.: On the intersection of regex languages with regular languages. Theoretical Computer Science 410, 2336–2344 (2009)
3. Freydenberger, D.D.: Extended Regular Expressions: Succinctness and Decidability. In: 28th International Symposium on Theoretical Aspects of Computer Science, Dagstuhl, Germany. Leibniz International Proceedings inInformatics (LIPIcs), vol. 9, pp. 507–518. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik (2011)
4. Ginsburg, S., Greibach, S.: Deterministic context free languages. Information and Control 9(6), 620–648 (1966)
5. Ginsburg, S., Rose, G.F.: Operations which preserve definability in languages. Journal of the ACM 10(2), 175–195 (1963)
6. Greibach, S.: A note on undecidable properties of formal languages. Theory of Computing Systems 2, 1–6 (1968)
7. Hopcroft, J.E., Ullman, J.D.: Introduction To Automata. In: Theory, Languages, And Computation, 3rd edn., Pearson Education (2008)
8. Reidenbach, D., Schmid, M.L.: A polynomial time match test for large classes of extended regular expressions. In: 15th International Conference on Implementation and Application of Automata, pp. 241–250. Springer (2011)
9. Thompson, K.: Programming techniques: Regular expression search algorithm. Communications of the ACM 11, 419–422 (1968)

Randomized Algorithms: On the Improvement of Searching Techniques Using Probabilistic Linear Linked Skip Lists

Yogeesha C.B.[1], Ramachandra V. Pujeri[2], and Veena R.S.[3]

[1] Department of Computer Science and Engineering,
Karpagam University, Coimbatore, Tamilnadu, India – 641 021
ycb@cisco.com
[2] Department of Computer Science and Engineering,
KGiSL Institute of Technology, KG Campus, Coimbatore, Tamilnadu, India – 641 035
sriramu_vp@kggroup.com
[3] Department of Computer Science & Engineering,
K.S. School of Engineering & Management, Bangalore-560 062, Karnataka, India
veena_me@yahoo.com

Abstract. This paper explores a concept of randomized algorithms [6] combined with linked skip list. A skip list is a probabilistic data structure where elements are kept sorted by key. It allows quick search, insertions and deletions of elements with simple algorithms. It is basically a linked list with additional pointers such that intermediate nodes can be skipped. It uses a random number generator to make some decisions. Skip Lists are used as an alternative to balanced trees. A skip list is a practical data structure that gives good results while keeping the implementation simple.

1 Introduction

Data Structure [1][7] is a mathematical or logical representation of particular way of storing and organizing data in a computer so that it can be used efficiently. One of the type of non-primitive linear data structure is Linked List. Linked list is a relatively easy data structure to implement. It is simple to keep a linked list of n elements sorted. To perform a search n comparisons are required in the worst case. Now, suppose a second pointer pointing two nodes ahead is added to every other node, the number of comparisons required goes down to ceil (n/2)+1 (in the worst case). Adding one more pointer to every fourth node and making them point to the fourth node ahead reduces the number of comparisons to ceil(n/2)+2. If that strategy is continued so every node with i pointers points to 2^{i-1} nodes ahead, $O(log\ n)$ performance is obtained and the number of pointers has only doubled (n + n/2 + n/4 + n/8 + n/16 + ... = 2n).

2 Skip List Problem Definitions: A Probabilistic Data Structure

A typical representation of linked list and skip lists node [2][3][4][7] of the data structure described in Fig. 1. The nodes representation of a skip list are in the following

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 147–153.
springerlink.com

proportions: 50% of them are level 1 node, 25% are level 2 nodes, 12.5% are level 3 nodes, etc. The number of levels required is log2 n where n is the expected number of elements. The data structure described above is great for searching but insertions and deletions would be really difficult to implement and inefficient as almost all the pointers must be modified. Maintaining perfect balancing is time consuming. This data structure can be made more flexible for simple insert and delete operations. This can be done quite easily while keeping a good (though not perfect) balancing as follows: The level of every new node to be inserted is chosen randomly with a probability distribution that keeps approximately the proportions of nodes of level i as described above. Probabilistic analysis shows that the average cost of insert, delete and search operations is O(log n).

Skip list [5] operations are O(n) in the worst case. It happens when all (or almost all) the nodes are given level 1 at insertion. In that case, the skip list becomes an ordinary linked list. The motto of Probabilistic Data Structure [8] such as Skip List is Don't worry, be happy! as this case is highly unlikely to happen as n increases. In this case we simply accept the result of random Level 1 and expect that probability eventually work in our favor. The advantage of this approach is that the algorithms are simple, while requiring only O(log n) time for all operations in the average case.

In practice [8], the Skip List will probably have better performance than a Binary Search tree (BST). The BST can have bad performance caused by the order in which data are inserted. For example, if n nodes are inserted into a BST in ascending order of their key value, then the BST will look like a linked list with the deepest node at depth n−1. The Skip List's performance does not depend on the order in which values are inserted into the list. As the number of nodes in the Skip List increases, the probability of encountering the worst case decreases geometrically. Thus, the Skip List illustrates a tension between the theoretical worst case (in this case, O(n) for a Skip List operation), and a rapidly increasing probability of average-case performance of O(log n), that characterizes probabilistic data structures.

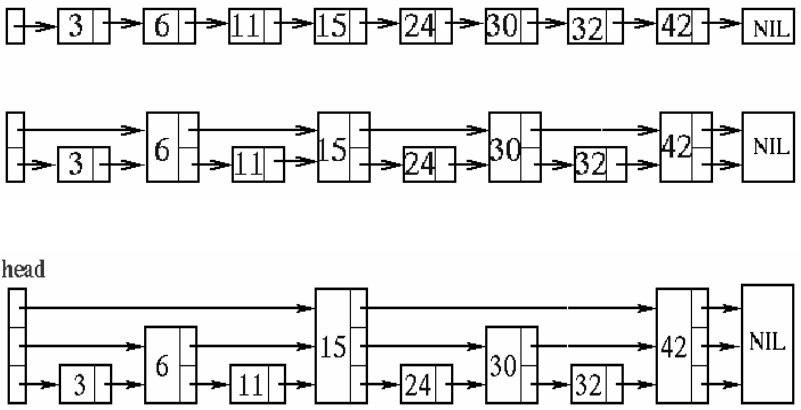


Fig. 1. Linked list and Skip Lists

3 Designs and Implementation: Algorithms

In the following algorithms

- Forward pointers of a level i node are stored in an array called forward indexed from 1 to i
- There is a constant called MaxLevel. All levels must be smaller than or equal to this constant
- The level of a node is not stored
- The level of a list = maximum {levels of the nodes in the list}.

3.1 Initialization

A new list is initialized as follows:

In Fig. 2 a node called NIL is created and its key is set to a value greater than the greatest key that could possibly be used in the list (i.e. if the list will contain only keys between 1 and 999, then 1000 may be taken as the key in NIL). Every level ends with NIL. The level of a new list is 1.All forward pointers of the header point to NIL.

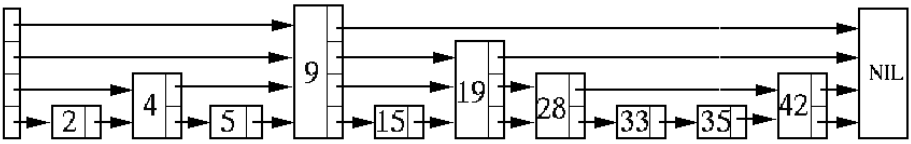


Fig. 2. Skip List with 4 levels

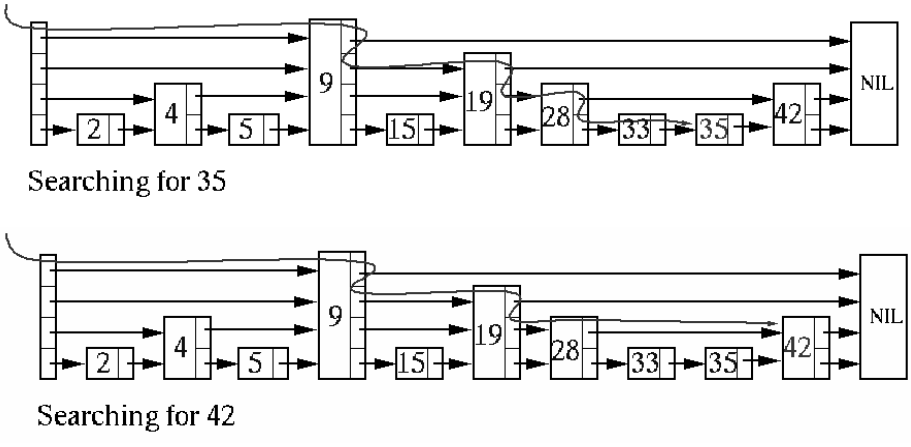


Fig. 3. Searching for the key element in skip list

3.2 Searching

Start at the highest level of the list.

Move forward following the pointers at the same level until the next key is greater than the searched key If the current level is not the lowest, go down one level and repeat the search at that level from the current node Stop when the level is 1 and the next key is greater than the searched key If the current key is the searched key return the value of that node. Otherwise, return a failure. The steps are demonstrated in Fig. 3.

```
SEARCH(list, searchKey)
        1.  x <- list.header
        2.  for i <- list.level downto 1
        3.      do  while x.forward[i].key < searchKey
        4.              do  x <- x.forward[i]
        5.  x <- x.forward[1]
        6.  if x.key = searchKey
        7.      then return x.value
  else return failure
```

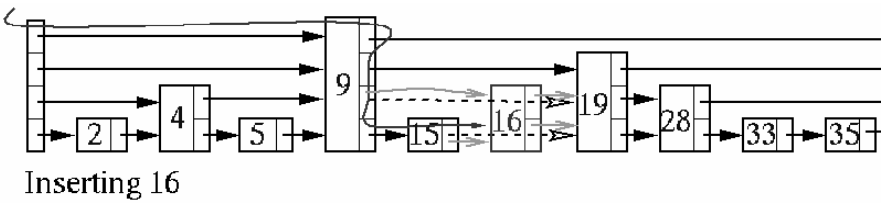


Fig. 4. Inserting an element to skip list

3.3 Insertion / Deletion

The insertion or deletion (Fig. 4 and Fig. 5) of a node consists mainly in a search followed by pointers update. An array update is used to store the last node reached at every level. It is used when changing the pointers after a node has been inserted or deleted.

The level of the new inserted node is determined randomly by the function RANDOM-LEVEL.

```
INSERT(list, searchKey, newValue)
        1.  x <- list.header
        2.  for i <- list.level downto 1
        3.      do  while x.forward[i].key < searchKey
        4.              do    x <- x.forward[i]
        5.          update[i] <- x
        6.  x <- x.forward[1]
        7.  if x.key = searchKey
```

```
8.      then    x.value <- newValue
9.      else    newLevel <- RANDOM-LEVEL()
10.             if newLevel > list.level
11.                then    for i <- list.level + 1 to newLevel
12.                           do    update[i] <- list.header
13.                        list.level <- newLevel
14.             x <- MAKE-NODE(newLevel, searchKey, newValue)
15.             for i <- 1 to newLevel
16.                do     x.forward[i] <- update[i].forward[i]
17.                       update[i].forward[i] <- x
```

```
DELETE(list, searchKey)
1.   x <- list.header
2.   for i <- list.level downto 1
3.      do    while x.forward[i].key < searchKey
4.               do     x <- x.forward[i]
5.            update[i] <- x
6.   x <- x.forward[1]
7.   if x.key = searchKey
8.      then    for i <- 1 to list.level
9.                 do     if update[i].forward[i] <> x
10.                           then break
11.                        update[i].forward[i] <- x.forward[i]
12.             FREE(x)
13.             while list.level > 1 and list.header.forward[list.level] = NIL
14.                do list.level <- list.level - 1
```

```
RANDOM-LEVEL()
1.   newLevel <- 1
2.   while RANDOM() < p
3.      do     newLevel <- newLevel + 1
4.   return MIN(newLevel, MaxLevel)
```

3.4 Random Level

This last algorithm deserves some explanations.

The function RANDOM( ) returns a number between 0 and 1.0. -p is a constant between 0 and 1.0 (suppose p = 0.5). RandomLevel works like flipping a coin. Say head is a win and tail is a lost. The electronic coin is flipped until a tail is obtained. Every time it is head the level is increased by 1 and the coin is flipped again.

Note: if p = 1/4, there is an average of 1.33 pointers per node. This saves space without reducing significally the search time.

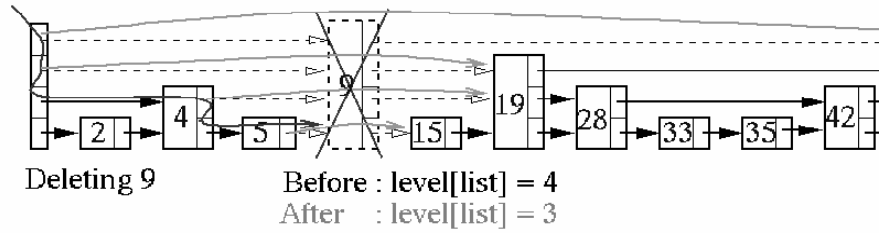


Fig. 5. Deleting an element from skip list

4 Results and Benefits

The BST is efficient but may become easily unbalanced after several insertions and deletions. Balanced trees (2-3 tree, red-black trees, ...) are guaranteed to remain balanced and thus does the basic operations in O(log n) in the worst case. Theoretically, they are efficient but their implementation is complicated. On the other hand skip lists are easier to implement. The algorithms for insertions and deletions are simpler and faster. They do not guarantee O(log n) performance but they do have an O(log n) performance in the average case (for insert, delete, search) and the probability of a high deviation from the average is quite high. Therefore, a very bad performance (O(n)) is extremely unlikely and its probability decreases exponentially as n increases. For most applications they are as efficient as balanced trees structures (see timings table below). They are also space efficient as no balance information needs to be stored in the nodes and they can work well with an average of only 1 1/3 pointers per node. Unlike the BST, the performance of a skip list does not depend in the order of insertion. Their implementation simplicity makes them a good alternative to BST and other tree structures.

Table 1. Timings of Implementations of Different Algorithms

Implementation	Search Time	Insertion Time	Deletion Time
Skip lists	0.051 msec (1.0)	0.065 msec (1.0)	0.059 msec (1.0)
non-recursive AVL trees	0.046 msec (0.91)	0.10 msec (1.55)	0.085 msec (1.46)
recursive 2-3 trees	0.054 msec (1.05)	0.21 msec(3.2)	0.21 msec (3.65)
Self-adjusting trees :			
Top-down splaying	0.15 msec (3.0)	0.16 msec (2.5)	0.18 msec (3.1)
Bottom-up splaying	0.49 msec (9.6)	0.51 msec (7.8)	0.53 msec (9.0)

5 Conclusions and Further Work

A skip list is a probabilistic data structure where elements are kept sorted by key. It allows quick search, insertions and deletions of elements with simple algorithms. It is basically a linked list with additional pointers such that intermediate nodes can be skipped. It uses a random number generator to make some decisions. The paper explores different techniques adapted to skip lists like insertion, deletion and searching for a key element. The functions are implemented in 'C' programming language under Linux environment and results are compared with other existing methods. The algorithms for insertions and deletions are simpler and faster. They do not guarantee O(log n) performance but they do have an O(log n) performance in the average case (for insert, delete, search) and the probability of a high deviation from the average is quite high. Therefore, a very bad performance (O(n)) is extremely unlikely and its probability decreases exponentially as n increases. For most applications they are as efficient as balanced trees structures. They are also space efficient as no balance information needs to be stored in the nodes and they can work well with an average of only 1 1/3 pointer per node.

Acknowledgments. The authors wish to thank member of Collaboration and Communications Group (CCG), Cisco Systems, Bangalore and Department of Computer Science and Engineering, Karpagam University, Coimbatore for their help in completing this work.

References

1. Langsam, Y., Augenstein, M.J., Tenanbaum, A.M.: Data Structures using C and C++. PHI/Pearson Edition (2003)
2. Shaffer, C.A.: A Practical Introduction to Data Structures and Algorithm Analysis–Java, pp. 365–371. Prentice-Hall, Inc. (1998)
3. Pugh, W.: Skip Lists: A Probabillistic Alternative to Balanced Trees. Communications of the ACM 33, 668–676 (1990)
4. Sen, S.: Some observations on skip lists. Information Processing Letters 39(4), 173–176 (1991)
5. Schneier, B.: Skip lists. Dr. Dobb's Journal 19, 50–52 (1994)
6. Motwani, R., Raghavan, P.: Randomized Algorithms. Cambridge University Press (1995)
7. The NIST website. National Institute of Standard and Technology, Dictionary of Algorithms and Data Structures, `http://www.nist.gov/dads/`
8. ShafferJanuary, C.A.: A Practical Introduction to Data Structures and Algorithm Analysis, 3rd edn., C++ Version, January 19 (2010), `http://www.scribd.com/doc`

Review of Proposed Architectures for Automated Text Summarization

Tejas Yedke[1,2], Vishal Jain[1,2], and R.S. Prasad[1]

[1] Dept. of Computer Engineering,
Vishwakarma Institute of Information Technology, Pune
{tejasyedke,rsprasad_viit}@yahoo.com
jainvishal@y7mail.com
[2] B.R.A.C.T's VIIT Sr. No. 2/3/4 Kondhwa (Bk.) Pune – 48,

Abstract. Automatic Summarization is the creation of a shortened version of the text by a Computer Program. It is a brief and accurate representation of input text such that the output covers the most important concepts of the source in a condensed manner. The summarization process could be extractive or abstractive. Extract summaries contain sentences that are copied exactly from the source document. In abstractive approaches, the aim is to derive the main concept of the source text, without necessarily copying its exact sentences. It is generally agreed that automating the summarization procedure should be based on text understanding that mimics the cognitive processes of humans. However, this is a sub problem of Natural Language Processing (NLP) and is a very difficult problem to solve at this stage. Through this paper, we intend to review various architectures which have been proposed for automated text summarization.

1 Introduction

Over the past decade the research and progress in the field of text summarization has led to development of methods such as Computational Intelligence (CI), Artificial Neural Networks (ANN), Fuzzy Systems (FS), Evolutionary Computation, Hybrid systems and other methods & techniques. However, there are a number of problems [5] while applying these techniques to Text Summarization problem:

1. Difficulty in preselecting the system's architecture.
2. Catastrophic forgetting.
3. Excessive training time required.
4. Lack of knowledge representation facilities.

To overcome the above problems, improved and hybrid methods and techniques are required both in terms of learning algorithms and systems learning [6].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 155–161.
springerlink.com

2 Methods for Automated Text Summarization

2.1 Text Summarizer Based on Graph Theory

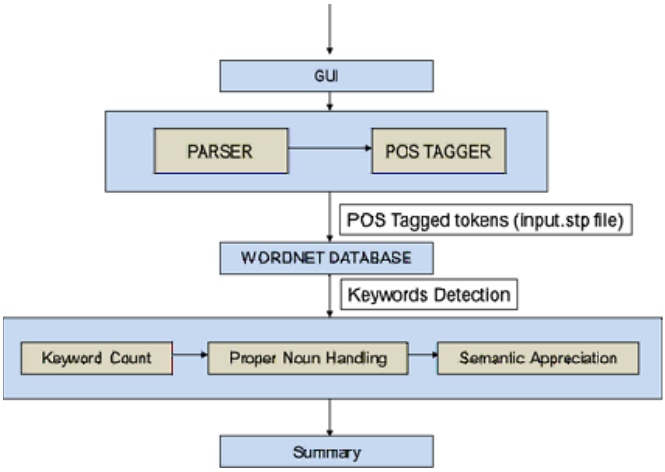


Fig. 1. Architecture of Text Summarizer based on Graph Theory

Generally, this methodology requires statistical induction of synset clusters and entails costly training of specific key domains. The present study word net is rich enough to obtain useful results in text categorization and summarization without training the tagged corpora. Stanford Parser is used to generate the parse tree of a sentence and extract typed dependencies among the words of a sentence. The typed dependencies provide a description of grammatical relationships between the words of sentence.

After the semantic grading of the nouns and verbs, also called nuclei, has been done, keywords among the nuclei are identified these keywords are nuclei having a semantic grading or polysemy count of <=5. After all the keywords have been determined, keyword counts of each and every sentence, the semantic unit of summary, are determined. Then, the semantic appreciation of each sentence i.e. modifier effect on nuclei is determined. Using these two criteria and considering proper nouns, sentences for the summary is picked. Our study suggests that there are some limitations of this system. They are as follows:

1. The software is subject to the memory limitations of the parser. If the parser runs out of memory while processing a long passage, the software generates an error.
2. The parser also sometimes trips up when considering apostrophes and double quotes. It is unable to process them appropriately.
3. Due to the humungous size of the WordNet database, it takes a considerable amount of time (40 seconds to 1 minute) for the summary to be generated. The entire database must be traversed for each token in order to calculate their precedence and semantic grading.

2.2 Text Summarizer Based on Fuzzy Logic [7]

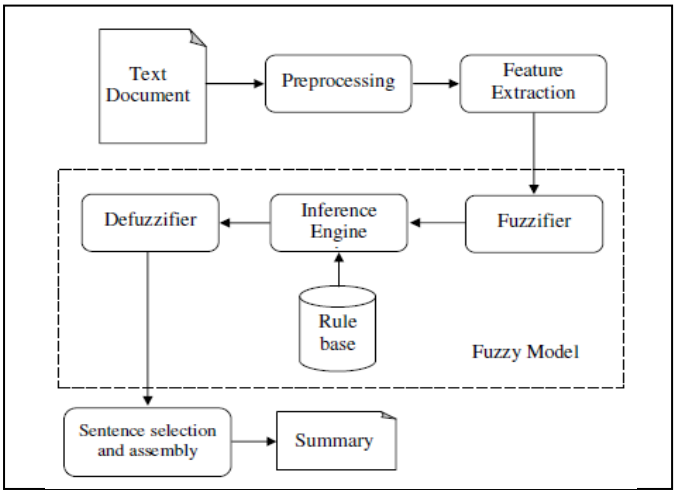


Fig. 2. Text Summarizer based on Fuzzy Logic

This architecture is based on decision module (fuzzy logic) and its essence is that this system has decision making capabilities. It has four major components, as follows:

1. Pre-processor: The source text document is fed to the pre-processor which includes pre-processing activities such as sentence segmentation, stop word removal, word stemming. Segmentation means to split the sentences based on a delimiter full stop. Stop word removal means to remove words(noise) such as 'is', 'are', and' etc. Word stemming is achieved using Porter Stemmer. Stemming converts word into its root form. For e.g. 'Worked' and 'Working' get converted to 'Work'.
2. Feature Extraction: Feature vector is computed for each sentence. Every sentence in the document along with its ID has a feature vector with nine fields for the significant text features. All the features will have the value range between 0 and 1. Significant text features considered in the design of the proposed approach are word similarity among sentences, word similarity among paragraphs, iterative query score, format based score, numerical data, cue-phrases, term weight, thematic features, and title features.
3. Fuzzy Logic Module: Triangular membership function [3] and fuzzy logic are utilized to score a sentence based on the extracted text features. This module consists of four parts: Fuzzifier, Rule base, Inference Engine and Defuzzifier.
   - A fuzzifier converts the input feature values into linguistic values (Very Low, Low, Medium, High, and Very High) using the membership function. The linguistic value denotes a fuzzy set (E.g. Low) to which a given sentence feature belongs

- Rule base: The most important procedure in any fuzzy system is defining the fuzzy IF-THEN rules. Sample of fuzzy rule is given below:

  IF (Word similarity among sentences is H) **and** (Word similarity among paragraphs is VH) **and** (Iterative query score is H) **and** (Format based score is M) **and** (Numerical data is H) **and** (Cue-phrases is VH) **and** (Term weight is M) **and** (Thematic feature is VH) **and** (Title features is H) **THEN** (Sentence is important)

- Inference Engine: The Inference engine compares the fuzzy input obtained from the fuzzifier with the Knowledge /rule base and decides the importance of a sentence. The output of inference engine is one of the linguistic values from the set {Unimportant, Average, and Important}.
- Defuzzifier: The input for the defuzzification process is a fuzzy set and the output is a single number. As much as fuzziness helps the rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. The linguistic values obtained from the inference engine are converted into crisp values by the defuzzifier. The crisp value denotes how close the sentence is to the given linguistic value. Centroid defuzzification method has been used to defuzzify values in this system.

4. Sentence Selection & Assembly: This module has two steps. First, determining the no. of sentences to be included in the summary, based on the amount of compression given by the user. Second, extracting the appropriate sentences for the summary.

This system has the following advantages:

1. Decision module is modeled using a fuzzy inference system. The summary of the document is created based upon the degree of importance of sentences in the document.
2. The semantic relation, among the extracted sentences in the summary, is maintained by the features like word similarity among sentences and word similarity among paragraphs
3. Summary is generated quickly and the results (though subjective) are better.

There are some disadvantages in this system. They are:

1. The major problem with purely statistical methods is that they do not account for context. Specifically, finding the aboutness of a document relies largely on identifying and capturing the existence of not just duplicate terms, but related terms as well.
2. This concept, known as cohesion links semantically related terms which is an important component in a coherent text is missing.
3. There is no training or learning method to enhance the capability of system. The system also lacks in adaptive learning.

2.3 Text Summarizer Based on Evolutionary Connectionism [6]

It consists of the following components:

1. Pre-processing
2. Feature extraction
3. Fuzzy model
4. Evolutionary Programming (EP) model
5. Connectionist model
6. Sentence selection and assembly

Pre-processing, Feature Extraction, Fuzzy Model and Sentence Selection & Assembly are the same as explained in former part of this study. The added variations are EP Model and Connectionist Model, which essentially overcome the disadvantages of the Text Summarizer explained in Sec 2.2.

Evolutionary Programming (EP) module generates large number of feature vectors (chromosomes) iteratively utilizing cross over and mutation operators subsequently, fuzzy logic is employed on the chromosomes and it returns the fuzzy score for each chromosome. Then, the chromosomes with their fuzzy score are fed to the neural network for training.

A lot of research using neural networks is made under the more common name "connectionist". Here, Multi-layer Perceptron Neural Network (MLPNN) has been used. A multilayer perceptron is a feed forward artificial neural network model that has at least one layer in-between the input and the output layer. A neural network MLP couples, through functions and weights, certain variables (called inputs) with certain other variables (called outputs) [4]. The neural network used in this system has configured with a nine inputs, one hidden and one output layer.

3 Evaluation and Results

DUC 2002 [10] dataset contains documents on different categories and this dataset have been used as experimentation material to test these systems. There are many ways by which we can evaluate the retrieval quality of an Automatic Summarization System. The quality of summary (Intrinsic Evaluation) is determined using precision, recall and f-measure. Precision is used for calculating the ratio of correctness of the sentences in the summary. Recall calculates the ratio of number of relevant sentences included in the summary. The weighted harmonic mean of precision and recall is called as f-measure.

Table 1. Comparative Results of Subjective Evaluation

Summarizer	Precision	Recall	F-measure
Copernic	0.8	0.775	0.786
Intellexer	0.825	0.7083	0.7559
MS-Word	0.5916	0.625	0.5913
Graph Theory	0.7666	0.6555	0.7
Fuzzy Logic	0.83	0.79	0.8095
Evolutionary Connectionism	1	0.77	0.87

The subjective evaluation is carried out on parameters such as Content, Readability and Overall Responsiveness (OR). The study also reveals that the subjective quality of the generated summary is acceptable and it is as follows:

Table 2. Comparative Results of Intrinsic Evaluation

Summarizer	**Content**	**Readability**	**OR**
Copernic	9	8.5	8.5
Intellexer	9	8	8
MS-Word	6	6.5	6.5
Graph Theory	8	8	8
Fuzzy Logic	9	9	9
Evolutionary Connectionism	9	9	9

4 Conclusion

This study explicitly reveals the amount of work done in the domain of text summarization. The authors [8] had initially proposed an Evolving Connectionist environment for Text summarization. This was followed by development of a summarizer based on Semantic Network, maintaining in mind the previous ideology of creating a network and an environment.

Now, Semantic Net did have some limitations but the results were promising and this led to development of a summarizer based on fuzzy model. The results with this system were better as it had decision making capabilities. However, the concept of creating a network was not much focussed upon here. Also there was an environment but the system didn't possess incremental learning capabilities rather it was just a system capable of learning and without any feedback mechanism or ability to learn from past experiences.

The next stage of this development (as per our study) is a summarizer based on Evolutionary Connectionism, which also has an added advantage of Fuzzy Decision Making Capabilities. This study clearly reveals that the idea of creating a summarizer whose architecture involves learning from the environment and a network structure has been implemented successfully by the authors [6]. They once quoted that to overcome the problems in the field of text summarization, *improved and hybrid methods and techniques are required both in terms of learning algorithms and systems learning* and have now provided a solution to the problem.

As an ending note we wish to point out that, initially the authors proposed evolving connection but the idea got translated to evolutionary connectionism. Now, as per our study, these are two different concepts. While Evolutionary Connectionism is an adaptive, incremental learning and knowledge representation system that evolves its structure and functionality, where in the core of the system is a connectionist architecture (neural network) that consists of neurons (information processing units) and connections between them. The former is a CI that is based on neural networks, but using other techniques of CI that operate continuously in time and adapt their structure and functionality through a continuous interaction with the environment. Evolutionary Connectionism is much closer to the way human brain works.

References

1. Amari, S., Kasabov, N.: Brain-like Computing and IntelligentInformation System. Springer, Singapore (1998)
2. Hebb, D.: The Organization of Behavior. Wiley, New York. Haykin, S.: Neural Networks: A Comprehensive Foundation, 2nd edn. Prentice Hall (1998)
3. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
4. Lahoz, D., Migue, M.S.: MLP neural network to predict the wind speed and direction at Zaragoza. Monografías del Seminario Matemático García de Galdeano 33, 293–300 (2006), http://www.unizar.es/galdeano/actas_pau/PDFIX/LahSan05.pdf
5. Prasad, R.S., Kulkarni, U.V., Prasad, J.R.: Machine learning in evolving connectionist text summarizer. In: Proceedings of IEEE International Conference on Anti-Counterfeiting, Security and Identification, August 20-22, pp. 539–543. IEEE Xplore Press, Hong Kong (2009a), doi:10.1109/ICASID.2009.5277001
6. Prasad, R.S., Kulkarni, U.V.: Implementation and Evaluation of Evolutionary Connectionist Approaches to Automated Text Summarization. Journal of Computer Science 6(11), 1366–1376 (2010) ISSN 1549-3636
7. Prasad, R.S., Kulkarni, U.V.: An Automated Approach to Text Summarization using Fuzzy Logic. International Journal on Computer Engineering & Information Technology, IJCEIT 23(01), 07–21 (2010)
8. Prasad, R.S., Kulkarni, U.V., Prasad, J.R.: Connectionist Approach to Generic Text Summarization. World Academy of Science, Engineering and Technology (2009), http://www.waset.org/journals/waset/v55/v55-63.pdf
9. Zhijun, L.I., Minghong, L.: EOS: Evolutionary overlay service in peer-to-peer systems. Am. J. Applied Sci. 2, 1401–1406 (2005), doi:10.3844/.2005.1401.1406
10. DUC 2002 dataset (2002), http://www-nlpir.nist.gov/projects/duc/

Steer-By-Wire Implementation Using Kinect

Rohan Sadale[1], Roshan Kolhe[1], Sachin Wathore[1], Jagannath Aghav[1], Saket Warade[1], and Sandeep Udayagiri[2]

[1] College of Engineering Pune (COEP)
{sadalerb08,kolherd08,wathoresu08,
jva,waradesp10}.comp@coep.ac.in
[2] John Deere Technology Center India, Pune
UdayagiriSandeepC@johndeere.com

Abstract. Steer by wire is one of the most advanced technologies used in the automobile industry. This paper describes implementation of Kinect based Steer-by-wire system. It elaborates a novel concept of vehicle steering by use of hand gestures. The key idea is to replace the steering wheel and angle sensor by Kinect and use its gesture recognition capability for performing steering actions. This paper focuses on advanced control design method to carry out steering functions using gesture recognition.

1 Introduction

Steering control is a core part in vehicle's design as it controls the actual movement of a vehicle. Recent advancements in automotive industry make use of electronics and computers for safety and comfort of drivers. The use of electronic components (sensors or encoders) in place of mechanical and hydraulic controls to control a wide range of operations such as acceleration, braking, steering etc. is known as 'by-wire' technology [16]. The implementation of electronics elevates the performance, provides safety and reliability with reduced manufacturing and operating costs [7]. Conventional steering system comprises of steering wheel, steering shaft, power assist unit and gear assembly. When the driver steers, input through steering wheel is transmitted by steering shaft through gear reduction mechanism, enabling steering motion of front wheels [14]. Steer by wire system substitute electronic components in place steering shaft and introduces feedback motor attached to steering wheel.

By implementing algorithms and accomplishing tasks as mentioned, interaction between driver and steering control can be made more efficient and convenient. What we propose is use of Kinect, which is a motion sensing device capable of gesture and speech recognition [11], to control the steering through gestures. Here we replace the steering wheel, angle sensor and feedback actuator of a vehicle by Kinect. The purpose of this paper is to introduce a modified steer-by-wire system which would be capable of sensing angle from driver's hand gestures and convert it into vehicle movement accordingly without actual need of steering wheel.

The use of Kinect in steer-by-wire system provides many benefits. The main benefit of such system is for physically disabled and elderly people. As the person driving

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 163–170.
springerlink.com

the vehicle just needs to specify the gesture or command instead of actually rotating the steering wheel, these people can drive vehicles without much effort. Also as the Kinect is a programmable and reconfigurable device, it can be used in various vehicles with slight modifications in the code. It can also be configured as per driver's convenience to signal gestures. Thus as a whole, the use of Kinect in steer-by-wire system simplifies driving experience.

2 Related Work

Need for a simplified interior design and better space utilization leads to the rise of steer-by-wire technology. Many physical modifications are required to change characteristics of handling in conventional steering systems. But, vehicles equipped with steer-by-wire can accomplish same characteristics through active steering interventions. In conventional steer-by-wire system, mechanical linkage between the steering wheel and the front wheel is removed to provide vehicle stability and to assist driver for autonomous steering control. Fu Xiuwei, Fu Li and Kong Feng [12] proposed a steer-by-wire system using MATLAB environment, active steering control and the controlling scenario of integral separation PID. Their research showed that steer-by-wire controller based on Integral Partition PID Control [12] gives better dynamic characteristics than conventional controller.

Handling and stability of steer-by-wire system is very important at high speed and affects the difficulty level of driving. Duan Jianmin, Wang Ran and Yu Yongchuan [3] researched two control strategies viz. the immobile steering sensibility type control strategy and the yaw rate and sideslip angle control strategy [3] and increased handling and stability of steer-by-wire at high speed.

Paul Yih and J. Christian Gerdes [13] proposed a method for altering vehicle handling characteristics by augmenting driver's steering command with vehicle state feedback. The vehicle's response can be reduced or enhanced based on driver's preference and road condition. A vehicle state was accurately estimated using global positioning system and inertial navigation system sensor measurement [13]. They experimented that such a system can achieve modified handling behavior that is exactly equivalent to physically changing the cornering stiffness of the front tyres.

Compared to other steering systems, steer-by-wire system provides variable steering ratio, easy assembling and enhancement of active safety. To improve return ability of steer-by-wire system and reproduce realistic driving feeling, Ba-Hai Nguyen and Jee-Hwan Ryu [6] proposed a method by measuring road wheel motor's current directly. The steering torque on the rack is measured by current sensor. The current sensors are available at low costs and thus offer a simple and cost-effective solution to reproduce a real driving feel. They developed the force feedback control algorithm which not only gave a realistic driving feel, but also improved the return ability.

On-board controller translates higher-level vehicle commands into vehicle motion or activation of the vehicle's equipment. Prototype of Distributed Control System (DCS) developed by RedZone Robotics [1] consists of a network of small, intelligent

nodes mounted throughout the vehicle. The nodes communicate via a controller area network bus, with a master unit to oversee the network operation. DCS provides cost-effective, flexible control and monitoring of a variety of robotic vehicle functions [1].

The introduction of Kinect by Microsoft led to a new enhancement in natural user interface at a considerably lower cost compared to that of sensors and cameras. Initially it was just used with the XBOX gaming applications, but after the release of OpenKinect [8] and Microsoft Kinect SDKs [5], it became a platform to develop more useful applications or integrate it with a huge domain of applications other than gaming. The release of above SDKs made it easier for the academic researchers and enthusiasts to create rich experiences by using Kinect. The main intention was to explore the development of natural user interfaces.

Kinect Identity technique [4] made use of multiple technologies and careful user interaction to achieve the goal of recognizing and tracking player identity. It tracks the identity in two ways viz. biometric sign-in and session sign-in. The identity system consists of 3 techniques namely face recognition, clothing color tracking and height estimation [4].

Jun-Da Huang [2] used gesture tracking capability of Kinect in physical rehabilitation system viz. Kinerehab [2]. In this system, gestures are used to find out whether the rehabilitation has reached a particular standard and whether the movements of students are correct or not. An interactive interface using Kinect also enhances student's motivation, interest and perseverance with rehabilitation.

Hand gesture detection is an important aspect of HCI. The authors of [9] used Kinect for hand detection and gesture recognition. But typical resolution of 640*480 for Kinect sensor provides problem in recognition of hand. It was eliminated using a novel shape distance metric called Finger-Earth Mover's Distance to measure the dissimilarities between different hand shapes [10].

3 Proposed Method

The system consists of two main parts, steering section and wheel section. Steering section is actually a modification of usual steer-by-wire system. The steering wheel, angle sensor and feedback motor are replaced by Kinect. Kinect acts as a device which performs function of both the steering wheel and the angle sensor.

Microcontroller (μC) is a core part of system design and it works as an intermediate between steering section and wheel section. It contains algorithms to convert output of Kinect into the desired input for the actuator which is usually represented as voltage levels.

The other part of system viz. wheel section consists of actuator, pinion angle sensor and gear assembly. Kinect provides the steering actuator with input angle. The actuator with steering gear (rack and pinion arrangement) is responsible for corresponding turn of the tyres. The dashed rectangular portion in the system architecture (Fig. 1) depicts implementation of Kinect in steer-by-wire system.

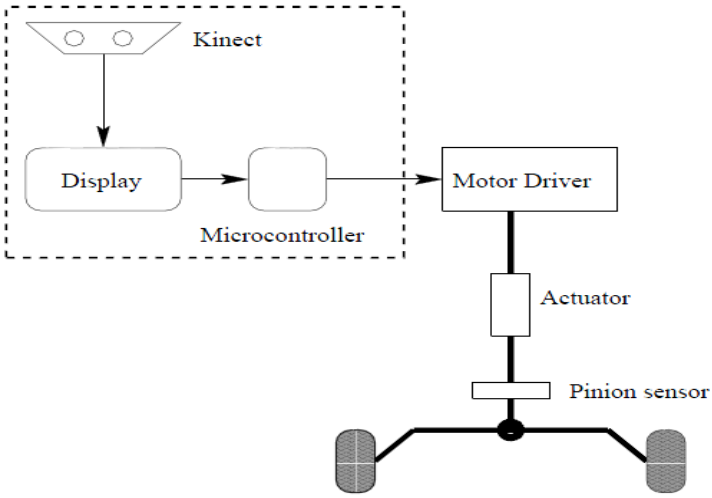


Fig. 1. System Architecture

3.1 Components

3.1.1 Kinect

Kinect is a device which is capable of gesture and speech recognition. It contains a RGB camera, 3D depth sensors and multi-array mic. It is a highly advanced and low cost device for effective natural user interface [5]. Kinect makes use of skeletal tracking and 3D depth data capabilities for detecting gestures [15]. With Kinect, hand gestures of the driver are tracked. As the driver's hands move in a 3-Dimensional plane, the corresponding co-ordinates are tracked and are converted into an angle which is given to μC. Corresponding hand gestures should be within the range of Kinect, so that positions of hand with respect to Kinect can be shown on the screen of a computer. A graphical display is used as an interface to user for displaying the steering actions through an image of a steering wheel and corresponding hand movements. A specific set of voice commands such as I AM READY to start the steering, STOP to stop steering etc. are also used so as to make the system flexible.

3.1.2 Micro-controller

Most of the control mechanism of a steer-by-wire system depends on the microcontroller (μC). It takes the angle from Kinect and converts it into the pulse (voltage level) with which the actuator should rotate. μC consists of the algorithms so as to convert the data from Kinect and give it to the motor driver. The output of the μC doesn't provide enough power (0-5 V) to rotate the actuator. Here the motor driver comes into picture.

3.1.3 Motor Driver (Controller)

Motor driver is used to supply motor with necessary input. It helps in obtaining required torque with which actuator rotate. Motor driver works as a DC to DC step up power converter. It boosts the voltage level as per the requirements.

3.1.4 Actuator (Servomotor)

DC servomotor is used to provide steering actuation. It is attached to the remainder of the connecting shaft via flexible coupling. DC servomotor is used because it reduces noise and maximizes the motor life. Servomotor consists of motor, gear head and feedback circuit. If the torque applied by the motor is sufficient to overcome the friction and road forces, wheels begin to move.

3.2 Gesture and Speech Recognition

A predefined set of gestures is implemented in system. Particular actions occur according to the gesture of user. The main gesture implemented is virtual steering action by hands. This gesture would specify whether to turn in left or right direction. Also, it would determine by what angle wheels should turn. The gestures are triggered by audio commands from the user. They are used to indicate whether user is ready or not.

3.3 Determination of Turning Angle of Wheels

From a gesture, we determine the angle by which wheels are to be turned based on a steering ratio and lock to lock turns. The steering ratio determines the angle by which wheel should be turned based on rotation of steering wheel. The steering ratio 12:1 means turn the wheel by 1 degree when steering wheel is rotated by 12 degrees. The steering ratio may range from 12:1 to 20:1 depending on the design of the steering system in a particular vehicle. The lock to lock turns specifies the number of rotations of steering wheel when it is rotated from a lock on one side to the lock on the other. Thus knowing the steering ratio and lock to lock turns, we determine the angle by which wheels to be turned based on the angle specified by gesture. Here, we have considered gesture frame of 90^0 for specifying the angle by which wheels to be turned. Whether to turn right or left, will depend on corresponding position of right and left hand.

$$\delta = \frac{\theta \times \frac{\frac{LTL}{2} \times 360}{90}}{\mu} \tag{1}$$

Where,

θ - Gesture angle
δ - Angle by which wheels to be turned
μ - Steering ratio
LTL - lock to lock turns

3.4 Workflow

The driver should be given some instructions regarding use of gestures and audio commands. Kinect waits for triggering of audio command to start the gesture recognition (e.g. I am ready). Next, it waits until both hands of the driver are recognized. If both hands are recognized and distance between the two hands is greater than some threshold value, whole steering is controlled by hands. Gestures are detected as hands move making virtual action of steering. The gesture system is explained in algorithm 1. In algorithm 1, lines 6-14 comes into picture only when command isn't stop or halt. As driver move his hands for performing virtual steering action, Z and X coordinates of his hands are tracked and angle between them is measured (Fig 2).

This angle is used to determine the angle by which wheel needs to be turned. When the command recognized is stop, the gesture system temporarily stops measuring the angle made by the hands. When the command recognized is halt, the gesture system stops tracking (steering operation stops). The tracking of user hands along with the illusion of hand wheel is shown on the display screen, so that the driver can actually see by what angle the wheel is turned. The corresponding angle is passed to μC. It then converts the angle into a pulse (0-5 V) and sends it to motor driver (motor controller). Motor driver supplies appropriate current/voltage to actuator, as voltage output from μC is very low for rotation of actuator. As actuator rotates uniformly, pinion rotates correspondingly making rack to move horizontally. Thus the tie-rods move and the wheels turn by an angle.

Feedback system is used to provide feedback of actual rotation of wheels to driver. Actual position of wheels is measured by pinion angle sensor. Pinion angle sensor gives feedback to μC. Thus the micro controller gives its output to computer application which runs Kinect and shows the effect of rotation of steering on display screen (an illusion of steering wheel on display is created which rotates). Thus even though driver applies much gesture on rough road, the wheels rotate by less degree and corresponding steering rotation is shown on screen.

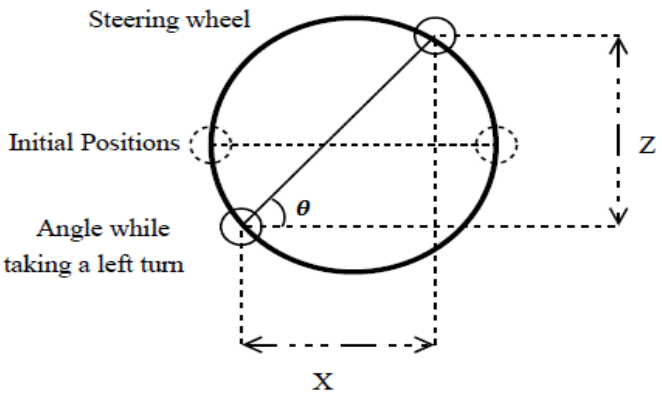


Fig. 2. Gesture Angle Determination

Algorithm 1. *Measuring Gesture angle*

Require: HANDS SHOULD HAVE SAME DISTANCE FROM KINECT
Ensure: DRIVER IS TRACKED

1. Start the system
2. Start gesture tracking by a audio command *I AM READY*.
3. GOTO 5
4. Continue recognizing commands until *RESUME* or *HALT* is recognized
5. while (command != *STOP* and command != *HALT* and *distance between hands* > *threshold*) do
6. if (righthand.Z < lefthand.Z) then
7. Z = lefthand.Z – righthand.Z
8. X = lefthand.X – righthand.X
9. tan θ = Z/X
10. else if (righthand.Z > lefthand.Z) then
11. Z = righthand.Z – lefthand.Z
12. X = righthand.X – lefthand.X
13. tan θ = Z/X
14. end if
15. Pass angle θ to μC
16. end while
17. if (command == *HALT*) then GOTO 20
18. else GOTO 4
19. end if
20. STOP

3.5 Fault Tolerance

The steering control system (μC) can diagnose faults by detecting input and output signals and driving current of motor. If this system fails, there must be some alternative to control the steering. The control unit (μC) then stops responding by activating the fail safe relay mode. This is indicated on display screen, activating the manual steering (conventional mechanical steering). The mechanical system of steering control which is used can be folded inside the dashboard by using telescopic cylinder. On activating fail safe mode, the telescopic cylinder can be opened to have the mechanical steering popped up. The vehicle will now be driven by conventional steering system.

4 Conclusions

The proposed gesture controlled steer by wire system provides flexibility to driver and efficiently controls vehicle movement. Important aspect of this research is how to interact between user gestures and mechanical components of steer-by- wire system. Introduction of angle tracking system based on driver's gesture provides a new way to

replace steering angle sensors and other mechanical components. Feedback system is used for steering angle correction due to obstacles and unexpected disturbances while driving. In the simplest form, the proposed system provides a new steering system which works and feels like conventional steer-by-wire system without actual steering wheel. By introducing Kinect based steer-by-wire system, our research shows a new way for driver assistance, flexibility and added new features like voice recognition to initiate steering procedure. While dealing with steer by wire, Kinect accuracy should be dealt properly, as it is very sensitive to human actions.

References

1. Callen, J.N.: Distributed control for unmanned vehicles. IEEE Concurrency 6(2), 16–20 (1998)
2. Chang, Y.-J., Chen, S.-F., Huang, J.-D.: A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. Research in Developmental Disabilities 32(6), 2566–2570 (2011)
3. Duan, J., Wang, R., Yu, Y.: Research on control strategies of steer-by-wire system. In: International Conference on Intelligent Computation Technology and Automation (ICICTA), vol. 2, pp. 1122–1125 (May 2010)
4. Leyvand, T., Meekhof, C., Wei, Y.-C., Sun, J., Guo, B.: Kinect identity: Technology and experience. Computer 44(4), 94–96 (2011)
5. Microsoft SDK, http://www.microsoft.com/en-us/kinectforwindows/
6. Nguyen, B.-H., Ryu, J.-H.: Direct current measurement based steer-by-wire systems for realistic driving feeling. In: IEEE International Symposium on Industrial Electronics, ISIE 2009, pp. 1023–1028 (July 2009)
7. Nice, K.: How car steering works, http://auto.howstuffworks.com/steering5.htm
8. OpenKinect. libfreenect, http://openkinect.org/
9. Ren, Z., Meng, J., Yuan, J., Zhang, Z.: Robust hand gesture recognition with kinect sensor. In: Proceedings of the 19th ACM International Conference on Multimedia, MM 2011, pp. 759–760. ACM, New York (2011)
10. Ren, Z., Yuan, J., Zhang, Z.: Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In: Proceedings of the 19th ACM International Conference on Multimedia, MM 2011, pp. 1093–1096. ACM, New York (2011)
11. Wikipedia. Kinect, http://www.wikipedia.org/kinect
12. Xiuwei, F., Li, F., Feng, K.: Research of automotive steer-by-wire control based on integral partition pid control. In: 3rd International Conference on Genetic and Evolutionary Computing, WGEC 2009, pp. 561–564 (October 2009)
13. Yih, P., Gerdes, J.C.: Modification of vehicle handling characteristics via steer-by-wire. IEEE Transactions on Control Systems Technology 13(6), 965–976 (2005)
14. Yih, P.: Steer-By Wire: Implications for Vehicle Handling and Safety. PhD thesis, Department of Mechanical Engineering Stanford University (2005)
15. MSDN blogs: Unleash the power of Kinect, http://blogs.msdn.com/b/eternalcoding/archive/2011/06/13/unleashthe-power-of-kinect-for-windows-sdk.aspx
16. Saha, S.: Steer by wire, http://saha.suvankar.tripod.com/id5.html

An Efficient Incentive Compatible Mechanism to Motivate Wikipedia Contributors

Mane Pramod[1], Sajal Mukhopadhyay[1], and D. Gosh[2]

[1] Department of Information Technology,
National Institute of Technology,
Durgapur, India 713209
{pramodcmane,sajmure}@gmail.com
[2] Department of CSE,
National Institute of Technology,
Durgapur, India 713209
profdg@yahoo.com

Abstract. Wikipedia is the world's largest collaboratively edited source of encyclopedic information repository consisting almost 1.5 million articles and more than 90,000 contributors. Although, since its inception on 2001, the numbers of contributors were huge, A study made in 2009 found that members (contributors) may initially contribute to site for pleasure or being motivated by an internal drive to share his knowledge. But latter they are not motivated to edit the related articles so that quality of the articles could be improved [1] [5].In our paper we address above problem in economics perspective. Here we propose a novel scheme to motivate the contributors of Wikipedia with the mechanism design theory that is the most emerging tool at present to address the situation when data is privately held with the agents.

1 Introduction

Wikipedia is the world's largest collaboratively edited source of encyclopedic information repository. The simple editing system for user contribution is distinctive feature of Wikipedia. The simple editing process inspires a person with almost no technical background of Wikipedia to contribute by writing his/her known information. Over 1.5 million article and more than 90000 contributors are key factors of Wikipedia to consider as one of the largest information reference website.

As Wikipedia is the best example of user created encyclopedia and collaborative user generated platform, its success and failure is decided by user participation. Indeed, content enrichment of Wikipedia is defined by overwhelming participation of motivated users. A 2010 study has found that although members may initially contribute to the site for pleasure, they are motivated primarily by an internal drive to feel efficacious and self-confident. Surprisingly, aside from potentially fueling a first-time contribution, enjoyment was found to have no significant impact on knowledge sharing behavior in Wikipedia [1].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 171–181.
springerlink.com

Table 1. Detailed categories of respondents

Type	2009(%)	2010(%)
Readers	62.55	65.92
Occasional Contributors	24.49	23.25
Regular Contributors	9.18	7.42

This is obvious that readers' response is higher due to information or knowledge reference. However the contributors response is an unsettling issue. Table1. Figures are signifying that occasional contributors and regular contribution rate is declined to 1.24%, and 1.76% respectively, see also [4].

Above information is addressing the primary question of how to motivate users (both reader and present contributors) for Wikipedia contribution. Though research [12] make conclusion that intrinsic motivation can play an important role in online communities but table-1 and [4] statistic shows controversial picture. Whereas a present effort in research on motivation of contributors is focused on extrinsic motivational factor and approach is to interprete Wikipedia contributor motivation in economic module. In [9] author presents an interesting social reward scheme in Wiki System. In this work author gives rank of contributors on the basis of amount of references, rating of articles, most viewed articles and to represent resulted rank author use star and spark-lines data visualization technique. To cope up with motivation problem and to motivate contributors, Wikipedia community started to honor its contributors by giving Barn-stars award .There- after some other awards viz. motivation award, personal user award and barns- star ribbons[2] also given to contributors for motivation. Wiki cup [3] one more interesting scheme run by Wikipedia to create competitive environment among contributors. But apart from this effort Wikipedia community has not reached at conclusive figure.

In Wikipedia information edited or articulated in collaborative way and be considered as public goods or property. Let us suppose that number of readers is 'r' and number of contributors is 'w'. Then in Wikipedia $r > w$, i.e. readers are larger than contributors [Table.1]. If we analyze this information from economic aspect then we will be eligible to see net utility of reader and contributor.

1.1 Problem Definition

Let us say R is a set of reader $\{r_1, r_2, r_3, \ldots, r_n\}$, W is a set of writer $\{w_1, w_2, \ldots, w_n\}$, t is a time spend by writer to collect and edit it, β is an amount spend by writer for internet and other resources, is an information gain common for both, paycost-effort for $i \in w$ writer is evaluated in following eq. $p = (t + \beta)$.Utility function denotes paycost-effort of contributor and reader as follows.) Indicate no efforts.

$$u(i) = \begin{cases} no\ effort\ (0) & \forall r_i \in r \\ paycost\ effort\ (0 - p) & \forall w_i \end{cases}$$

Aim:

$$\text{Maximize} \sum_{i=1}^{k} u_i$$

Subject to:

$$\text{Minimize} \sum_{i=1}^{k} p_i$$

Where $r_i \in \{R - W\}$ and $W, i = 1,2 \ldots \ldots n$, $k \in W$ and $k < n$

We are supposing, 87% reader [table 1] are not interested to contribute due t $'p'$ i.e. pay cost factor, which decrease the utility of contributor, hence their true e orts for information contribution is not revealed out. Here our approach is to make $-p \cong 0$ by providing incentive. In other words we try to make contributor's e ort utility nearly equal to reader hence every one motivated to contribute.

In our work we use an **Incentive Compatible Mechanism-Design** [19] [20], an emerging and important economical tool to module the situation when data is privately held by agents. Mechanism design is subset of game theory and set of protocols to achieve social-welfare. Mechanism design considers those problems in which multiple agents are selfish and organized by some set of rules or protocols to meet specified goal. Incentive compatible mechanism design tries to motivate the agents to reveal their true value or character (here we consider willing to take truth e ort for information contribution which agents have as private or true value).Thus eliciting true value of agents can be achieved by providing some payment or incentives to these selfish agents. Thus these set rules or a protocol of mechanism design is able to guide the agents towards a social-welfare.

1.2 Terms

Before we proceed further here we are defining some terms.

- Paycost-Effort: This term indicate that contributor pay some amount, his/her time and e ort to contribute wikipedia. It is indicated by '*p*', Our mechanism try to reduce '*p*' to increase paycost-effort utility of '*k*' contributor by providing incentive to them for contribution in competitive environment.
- Reputation Points: It is very hard to major the paycost effort of the contributor in some specific unit. Hence we are converting paycost effort of contributor in reputation points by designing polynomial algorithm with small modification in the existing wikipedia edit history mechanism.It is indicated by *'R'*. From now onwards in this paper, we use reputation points of contributor '*i'* as her paycost-effort.
- Utility: In motivation problem the utility of bidder is the form of monetary. Here utility of contributor is in the form of paycost effort he/she save by getting incentive with respect to reader and his competitor. It is indicated by '*u*'.

While proposing an IM (Incentive mechanism), we consider following issues;

- Incentive structure on the basis of *social need* which helps to reduce paycost effort 'p' and helps to increase utility with nearly equal to reader.
- Competitive Environment in Wikipedia user to gain incentive (to increase utility).
- To achieve, *social matching* for community enrichment.

In IM an incentive in the form of *Byline credit*[1] to contributors (we explain incentive structure in our mechanism section).Proposed scheme is considering only registered users not to anonymous. In [14] the results show that there is a positive correlation between user registration and the quality of contributed content. The majority of the revisions are submitted by registered users.

1.3 Why Incentive Compatible Mechanism

If we consider paycost-effort factor & table-2 statistics then this make reader as rational or selfish agent in information contribution. The issue of motivation for participation is more complex when a production of information is done by collaborative way. Wikipedia is a best example of public goods where information is produced in collaborative way. Most of the research concludes that intrinsic altruism behavior does not sustain in social capital, if the losses incurred by the altruistic agent are relatively high, then the degree of altruism goes down.

- *Free Riding*-Free riding shows selfish or rational nature of agents in cooperative information production. Agents take more benefit with less share or contribution in information production. Fischer shows that high cooperation is not sustained as equilibrium and degree of free riding decides cooperation. If there is more free riding then there is less cooperation and vice versa.
- *Social Loafing*-Social loafing is another phenomenon that indicates the rationalism of agent. In social loafing, persons make less e orts to achieve a goal when they work in a group than when they work alone.
- *Survey Analysis*-Rishab [Wikipedia survey] point out non contributing reasons in Wikipedia. Among several reasons, which de ne selfish nature of user. We consider that following some of them.

Table 1. Detailed categories of respondents

Reasons for not-contributing	(%)
Others are already doing it, there is no need for me	19.08
I don't have time	31.31
I don't feel comfortable editing other people's work	23.68
I don't think I have enough information to contribute	53.06
I am happy just to read it, i don't need to write it	45.49

[1] A line in a newspaper naming the writer of an article. In some cases, bylines may be used to give credit for who wrote an article. They are a small element in books, magazines, newspaper, or newsletter design but certainly important to the author.

In brief our aim is:

1)To reveal out the true e ort value of every contributor through competitive environment by means increasing their utility 2)To increase utility of contributor, we try to reduce paycost effort *p* of contributor by providing incentive to them.

The rest of the part of this paper is organized as follows: Section-2: describes related work. In section-3: we discuss about our tool mechanism design in brief and our motivation Problem Definition. In section-4: Problem Definition .In section-5: Our scheme to motivate to contributors and also show simulation result and limitation of our work .In section-6, we come on conclusion of our work.

2 Related Work

The work which is mostly closed to ours is [6]. In this vivek et al. use a game theoretic framework for motivating contributors in social media. Author proposes technique from a system designer's perspective. In this work user is considered as rational agent and finds out the optimal incentive level to in enhance the selfish agents for maximize the system utility. In other work like information providing in terms of resource exchange perspective [a elsever paper] shows that altruistic triats, social rewards and reciprocity influence willingness to provide information. We consider research work of [7][8] is nutritious food for our work. Sheizaf Rafaeli et el [8] is given considerable weight age incentive based motivation. Authors compel us to concentrate to-wards following point for wikipedia participation. 1) Professional versus non professional partici-pation. 2) Content contribution. 3) Continuous versus one time participation. In [7] Andrea Forte et el. highlight the importance of incentive structure for open-content publishing like wikipedia. In this paper author accentuate on byline to credit authors unlike scientific community for their hard work. In this work, our outlook to consider motivation of contributors is quantitative and qualitative to get solution for how people will contribute to Wikipedia.

3 Background and Motivation Problem

3.1 Mechanism Design

Mechanism design is the subfield of microeconomics and game theory that considers how to implement good system-wide solutions to problems that involve multiple self-interested agents, each with private information about their preferences [21]. In following way mechanism is-

Let *N* be the number of agents or participants. Each agent $i \in N$ can have private information or type or valuation t_i, e.g. in an auction the type of agent would be his valuation price for the item offered. A_i is the set of possible strategies or actions for agent *i*. Depending on his type, the agent will pick an action or strategy, $a_i \in A_i$, e.g. in an auction, a strategy of *i* would be a bid of a certain amount.

The mechanism provide as output function o = o $(a_1, a_2, a_3 \ldots\ldots\ldots, a_n)$ Agents $\forall_i \in N$ to optimize the utility u_i. The objective or certain outcome can be achieved by using payment p_i given to the agent(s).

4 Our Scheme

We are proceeding our scheme in following subsections:

-Incentive Structure and Solution of a problem
- Reputation Algorithm
- Incentive Allocation
- Claim: Our mechanism is truthful.
- Time Complexity of Mechanism
- Byline-Credit and Social Matching

4.1 Incentive Structure and Solution of Problem

While proposing incentive compatible mechanism design, we consider social need of people. In social network, self-marketing is an effective way to demonstrate someone's capability and skillfulness in his/her work area. Self-marketing or self-presentation is coming forward as a basic need of netizen. People namely employee, students, scholars, some govt. employee is interested to express or represent themselves. All these people need some medium to fulfill this need. Personal web pages provide a facility for express or advertise to an individual. In proposed mechanism, we consider this fact and design incentive structure which will fulfill this social need.

Here we consider k number of memory slots of x unit each to create personal web page for Wikipedia contributors, i.e. provide a service as web-host and web-domain (wiki-webhosting) for their contribution. Let us say W_{DS} and W_{HS} is a charge of web domain and web hosting respectively. Then $I = W_{DS} + W_{HS}$, where I is a reward or incentive for contributor. We know in $p = t + \beta$, its hard to evaluate a time of contributor in money. But if we consider of contributor then $\beta \ll I$. if we consider t of contributor then $p \cong I$.

4.2 Formulation of Reputation-Points

Here effort of a contributor i is a function $\theta_i : p_i \rightarrow r_i$, where p_i and r_i is a paycost-effort and reputation points respectively of contributor i . Let us say u_p is Unique page edited, Avg_p is Average Edits per Page, $D_{ji}(x)$ denote contributor i deletes x amount edition of himself, $D_{ii}(x)$ shows contributor i deletes x amount edition of contributor i, $D_{ji} + (x)$represent contributor j deletes x amount edition of contributor other than i .

Reputation points will be provided to each contributor on the basis of constructive edition. These reputation points will be consider as type (or private value) of contributor in resource allocation problem. Wikipedia has its own mechanism to count constructive edition done by contributor. Whereas Wiki counts edit history of its contributor with respect to A) Unique page edit. B) Average edit per page. C) LiveEdit. D) Deleted Edit. E) Total Edit. But we are interfering small change in existing system of Wiki for counting reputation points of contributor with help of

Algorithm-1. In wiki Edit variable can be consider as either new edition, updation or deletion in page. Every edit in page made by contributor is considered as positive edition. But consequently, edit count alone does not directly correlates with the effort put into improving Wikipedia. $TR_i = R_{St} + \sum_{i=1}^{A} R_i$ Where R_i is reputation points of contributor i, A is auction period and $y = 1,2, \ldots, n$ and R_{st} is saved reputation points of contributor i from previous computation.

4.3 Incentive Allocation

First come first serve mechanism cannot help to reveal out the truth effort of contributor. Hence it is not possible to calculate incentive evaluation of contributors'. As a result we adopt auction mechanism to create competitive environment for incentive allocation. Where auction mechanism provides a base to understand an incentive evaluation of contributor. We represent incentive allocation as follows:

n: is a set of contributor(bidder) who make bid.

$r: \{r_1, r_2, \ldots, r_n\}$ is a bid vector.

$r_i \in r$ is reputation points, indicate maximum effort taken by contributor $i = 1,2, \ldots, n$.

k: is number of rewards(incentive) for bid. Where $k < n$.

Mechanism sort out all $r_i \in r$ in following order

$r_1 \geq r_2 \ldots \geq r_k > r_{k+1} > \square r_n$, First r_k contributor is considerd as winner for reward. Where in case of $t_i = t_j$ winner is considered as random.

These k winner pay $(k+1)^{th}$ contributor's reputation point as a payment.

We have to note that each bidder i 2 n bids so as to maximize her profit. Where as $R_{st} = r_i - r_{k+1}$ is balance reputation points of contributor which is used for next reward allocation computation.

Algorithm 1 Calculate Reputation Points of *i*

```
1: REPCOUNT((UP)i, (Avgp)i))
2: {
3: (Edit)i ← (Up)i * (Avgp)i
3: if Delete = = TRUE then
4:          if di == true && (deletetime)i ≤ N
5:              Ri ← Prev [(Edit)i + di] (Avgp))
6:              (deletetime)i ← (deletetime)i + 1
7:          else if dj == TRUE && (deletetime)j ≤ N then
8:                  Ri ← Ri − dji (Avgp)
9:          (deletetime)j ← (deletetime)j + 1
10:                 Rj ← Rj
11:         else
12                  Ri ← Prev (Edit)i + dii (LiveEdit)i
13:                 Rj ← Rj − dji (Avgp)
```

14: end if
15: else
16: $R_i \leftarrow Prev\ (Edit)_i + (LiveEdit)_i$
17: end if

Truthfulness

Theorem 1. Our k-round Wiki-reward Mechanism is Truthful

Proof. We divide the proof for 1st round and then for any arbitrary tth round. In first round carry or saved reputation points of ith contributor is R_{si} = 0.And also the contributor joining for the first time in any tth round, his or her R_{si} = 0.

First Round:

Say r_i is the reputation points of contributor earned by giving maximum effort, that we call revealing his true effort and $r_{(k+1)}$ is reputation points of (k + 1)th contributor. We show in our mechanism s/he can't gain any pro t by deviating from r_i. And Let us Say contributor has given less e ort and say his reputation points earned is r_i' *i.e.* $r_i' < r_i$

- Case1: $r_{k+1} < r_i' < r_i$
Utility of contributor i is $u_i^* = r_i' - r_{k+1}$ and $u_i = r_i + r_{k+1}$ In this case contributor 'i' is still winning but $u_i^* < u_i$ is reduced. So, giving less effort not maximizing his net utility even if he is winning.

- Case2: $r_i' < r_{k+1} < r_i$
Contributor loose reward hence his net utility is zero.

The above two cases confirm that deviation from r_i is not maximizing his utility. So, revealing true e ort is the dominant strategy.

Any t**th Round**

Now consider the process continuous for k rounds after every γ duration. And each contributor tries to save max reputation points that are carry to the next round. Then utility of contributor $u_i = R_{si}^t$, where R_{si}^t reputation points saved by i at round *t*,and $t = 1 \ldots . k$.

Now we look for any t^{th} round process. Here we prove that any t^{th} round,apart from saved reputation $R_{s\,i}^{t}$,contributor i has to give best effort, otherwise he cannot increase his gain or utility. So, in any t^{th} round the saved amount reputation of previous $t-1$ round is R_{si}^{t-1}, and let us say in between $t-1$ round to t contributor i earn r_i reputation points by pay max effort.and r_{k+1} is reputation points of (k + 1)th contributor.

If i **deviate and earn** r_i' **(make less effort)** then

- **Case 1**: $r_{k+1} < r_i' < r_i$

Here Utility $u_i = (r_i + R_{si}^{t-1})$ and $u_i^* = (r_i' + R_{si}^{t-1}) - [(r]_{k+1} + R_{sk+1}^{t-1})$.In this case contributor i is win but as $r_i' < r_i$ hence his utility is reduced *i.e.*$u^* < u_i$.

- **Case 2**: $r_i' < r_{k+1} < r_i$

If we consider this case, contributor i gave less effort than $k+1$, here we consider following two sub cases

1. $R_i s(t-1) > (r_i(k+1) + R_i(sk+1)^{\uparrow}(t-1))$:
   Utility $u_i = (r_i + R_{si}^{t-1})$ and $u_i^* = (r_i' + R_{si}^{t-1}) - [r]_{k+1} + R_{sk+1}^{t-1})$
   agent i win , but utility is $u^* < u_i$.
2. $R_i s(t) < (r_i(k+1) + R_i(sk+1)^{\uparrow}(t-1))$:

This case implies that $r_i' + R_{si}^{t-1} < r_k + R_{sk+1}^{t-1}$. Hence contributor i lose, t round (present) contest hence utility is zero.

From above two cases, we can say for any contributor i give max or showing true effort r_i is a dominant strategy.

4.4 Time Complexity of Mechanism

The overall time complexity of our mechanism could be given as follows:

1. Time to sort out contributors according their reputation points in non increasing order is: $\theta(nlgn)$
2. To calculate utility of contributor i in each round t is; $\theta(k)$ where $k < n$
3. Total time taken to calculate step1 and step2 for t number of round is: $t(\theta(nlgn) + \theta(k))$ where $t \in$ some constant
4. **Over all time complexity is**: $\theta(nlgn)$

4.5 Byline-Credit and Social Matching

Only motivation to contributors' for Wikipedia is not our aim but also how to perform social inter-action among information seeker and contributor (article expert) is key question. Social matching system brings people together in both physical and on-line space [18].With present technique it is not possible for information seeker to find article expert. We consider this problem and provide byline credit to contributor as incentive with webhosting. [7] author suggests for byline credit as a incentive in Wikipedia like scientific community, here we consider this suggestion as incentive as well as tool for social interaction between information seeker. Social interaction technique work in following way.

a. Auction winner (contributor) will create home-pages what is our incentive for personal advertisement.
b. Byline-credit is provided to winning contributor as shown in fig [1] on her contributed articles.
c. Information seeker click on byline credit and home page of contributor will open this will lead further social interaction between information seeker and contributor.

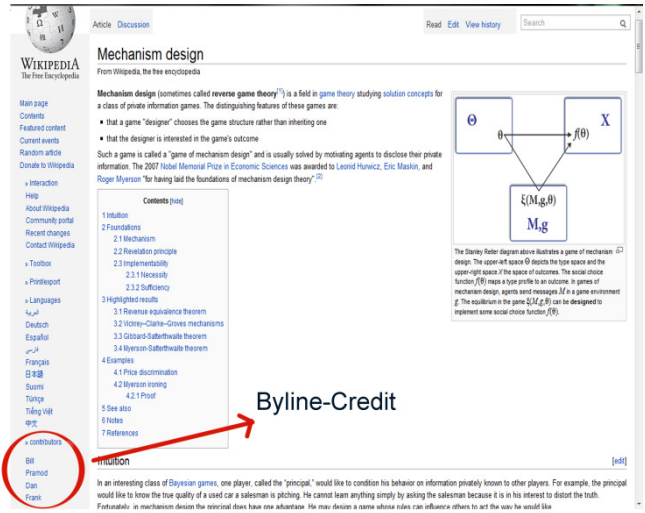


Fig. 1. Byline-Credit

5 Limitation and Future-Work

Suppose contributor i contributes only on mechanism design article and her total reputation points is R_i. Contributor j contributes on several articles as well as on mechanism design. Suppose j' s total reputation points is R_j and consider rj reputation point she collect by contributing on mechanism design. Let us say $R_i < R_j$ and $R_j < r_j$, then according to our scheme R_i is winner then byline credit will provide to each article contributed by her. But $r_j < R_i$ is deserve for by-line credit in mechanism design article. In future work we will try to resolve above problem by designing an incentive compatible mechanism for individual article.

References

1. http://en.wikipedia.org/wiki/Wikipedia:Wikipedians
2. http://en.wikipedia.org/wiki/Wikipedia:Barnstars
3. http://en.wikipedia.org/wiki/Wikipedia:WikiCup
4. http://stats.wikimedia.org/EN/TablesWikipediaEN.html
5. Ghosh, R.: Wikipedia Survey Overview of Results. UNU-MERIT-2009 (2010)
6. Singh, V.K., et al.: Motivating contributors in social media networks. In: WSM 2009, Beijing, China, October 23 (2009)
7. Forte, A., Bruckman, A.: Why do People Write for Wikipedia? Incentives to Contribute to Open-Content Publishing. In: GROUP 2005 Workshop: Sustaining Community: The Role and Design of Incentive Mechanisms in Online Systems, Sanibel Island (2005)

8. Rafaeli, S., Ariel, Y.: Online motivational factors: Incentives for participation and contribution in Wikipedia (2008)
9. Hoisl, B., Aigner, W., Miksch, S.: Social Rewarding in Wiki Systems – Motivating the Community. In: Schuler, D. (ed.) HCII 2007 and OCSC 2007. LNCS, vol. 4564, pp. 362–371. Springer, Heidelberg (2007)
10. Kuznetsov, S.: Motivations of contributors to Wikipedia. ACM SIGCAS Computers and Society 36(2) (June 2006)
11. Nov, O.: What Motivates Wikipedians. Communication of the ACM 50(2) (November 2007)
12. Yang, H.-L., Lai, C.-Y.: Motivations of Wikipedia content contributors. Computers in Human Behavior 26(6), 1377–1383 (2010)
13. Zhang, X.M., Zhu, F.: Intrinsic Motivation of Open Content Contributors: the Case of Wikipedia. In: Workshop on Information System and Economics, Evanston,IL (2006)
14. Javanmardi, S., et al.: User Contribution and Trust in Wikipedia. In: IEEE 5th International Conference on: Collaborative Computing, Networking, Applications and Worksharing (2009)
15. Javanmardi, S., et al.: Modeling user reputation in wikis. Statistical Analysis and Data Mining 3(2) (April 2010)
16. Manoj, P., Whinston, A.: Research Issues in Social Computing. JAIS 8(1), 336–350 (2007)
17. Pierce, M.E., et al.: Social Networking for Scientists Using Tagging and Shared Bookmarks: a Web 2.0 Application. In: International Symposium on Collaborative Technologies and Systems, CTS 2008 (2008)
18. Terveen, L., Mcdonald, D.: Social Matching: A Framework and Research Agenda. ACM Transactions on Computer-Human Interaction 12(3), 401–434 (2005)
19. Nisan, N., Ronen, A.: Algorithmic Mechanism Design. Games Econ. Behav. 35, 166–196 (2001)
20. Nishan, N., Roughgarden, T., et al.: Algorithmic Game Theory. Cambridge University Press (2007)
21. Narahari, Y., et al.: Game Theoretic Problems in Network Economics and Mechanism Design Solutions 2(14), 99–101 (2009)
22. Mukhopadhyay, S., Mane, P.: An e cient auction based ticket booking scheme for NBA all-star event championship. In: IEEE International Symposium on:Computer Communication Control and Automation (3CA) (2010)

Simulating Spiking Neuron for Information Theoretic Analysis in Stochastic Neuronal System

Sanjeev Kumar

Jawaharlal Nehru University, New Delhi, India
sanjeevgreen@gmail.com
Krishna Institute of Engineering and Technology,
Ghaziabad, India

Abstract. A neural model is used to analyze decoding of information from response and reproducing the response from a given stimuli. Extended leaky integrate and fire (LIF) model of neuron proposed by Deco and Scurmann is analyzed to study the effects of diffusion and jump process. Relationship in generated spikes and spike firing rate required to encode stimulus is validated. We have taken input stimuli spike train to be generated by Poisson process and studied the entropy of Poisson process during a small time window. We examined the information theoretic framework to simulate the coding strategy of single neuron for separating two different input spikes trains with use of information theory. Simulations have done to detect the number of output spikes required to differentiates between input signals without decoding the neural code.

1 Introduction

There is challenging question in computational neuroscience relating to the mechanism of information encoding in brain and how does information propagate in the neural system. To this end researcher are looking for the neuron model which can explain complex functionality of neuron [2, 3].

There are several models of spiking neuron which have been proposed in the literature. One of the basic models of neuron H-H was proposed by Hodgkin and Huxley (1952).This model described biophysical mechanism in brain in terms of membrane potential. One of the limitations of H-H model is that it involves a large number of parameters. Attempts have been made to consider simpler which have the advantage of involving a few parameters but at the same time capture neural dynamics one the well known model for generation of spikes is based on integrate and fire model [1,4,7].

A neural model can be used to either decode information from response or it can reproduce the response for a given stimuli. Neuron firings events are characterized by inter - spike intervals. Information in the brain is encoded through pattern of spike firing neuron in response to input signal. Our interest is to know how many spikes are generated or what spike firing rate is required to encode stimulus. Deco and Schurmann [5, 6] studied transmission of information, preciseness of input detection and mechanism of coding by spiking neuron.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 183–190.
springerlink.com

Neural firing event characterizes rate coding which encodes intensity of stimulus. Thus mean firing rate encodes the information. It has been observed by Softky and Koch[11] that spike train of cortical cells exhibit higher degree of variability measured in terms of coefficient of variation in the visual area VI .

They note that coefficient of variation of cortical cell is in the range of 0.5 – 1.0. Variability of ISI is one of the important aspects which enable to ensure that there is information in the input signal. This gives rise to time coding which requires precise time instances when spikes are emitted.

These observations led to the use of the spiking models of neuron therefore to describe mechanism of neuron information processing and coding. Accordingly we are required to employ models which generate spike e.g. integrate and fire model where spiking units (neuron) generate code that propagate information.

We wish to examine the Deco and Schurmann [5,6] information theoretic framework to simulate the coding strategy of single neuron for separating two different input spikes trains with use of information theory [5, 6].We consider the model due to Deco and Schurmann who extended LIF model to study the effects of diffusion and jump processes. This model has enabled us to simulate to detect the number of output spikes required to differentiates between input signals without decoding the neural code. For the purpose of illustration we have taken input stimuli spike train to be generated by Poisson process.

2 Generalized LIF Model

Stochastic differential equation of a single neuron driven by diffusion process is given

$$dV(t) = \left\{-\frac{V}{\tau} + \mu\right\} dt + \sigma dW(t) \tag{1}$$

where W (t) is a standard Wiener process. This equation allows generating ISI distribution in equation (1),

τ = constant decay of membrane potential when no input signal is applied
μ = drift parameter and
σ = measure of the strength of randomness.

Deco and Schurmann [5, 9] introduced another term proportional to increments in Poisson process. The modified LIF model reads as

$$dV(t) = \left\{-\frac{V}{\tau} + \mu\right\} dt + \sigma dW(t) + w dS(t) \tag{2}$$

where S (t) is homogeneous Poisson process. It may be noted that for Poisson process s (t),

$$\frac{d}{dt} S(t) = ds(t) = \sum_i \delta(t - t_i) \tag{3}$$

where t_i are the Poisson distributed random instants with mean rate λ such that λ^{-1} represents the mean value of the time interval between two spiking events.

For the sake of completion we give below the details of Poisson process.

2.1 Poisson Process

S (t) is a Poisson process which is referred to as counting process with rate λ. An important aspect of the process is that it is independent increment process. In a time interval of length t, the number of spikes is given as

$$P\{S(t+s) - S(t) = n\} = \frac{e^{-\lambda t}(\lambda t)^n}{n!}, \; n = 0,1,\ldots \tag{4}$$

It is a known result that if the arrival or generation of spikes is according to Poisson process the inter arrival time is exponential.

3 Spike Generations and Analysis

We closely follow the work by Deco and Schurmann [5, 6]. A spike is generated when membrane potential V (t) > Θ (Threshold value) and is reset to a given initial potential V(0) after generation of spike. The spike generation time is denoted by $t_0' \ldots t_k' \ldots$,

$$o(t) = \sum_k \delta(t - t_k') \tag{5}$$

The input signal is assumed to be described by independent stimuli in time and output spike trains are also independent because after the spike, the model is reset. We had taken the time precision ε so that λ ε << 1.

3.1 Entropy of Poisson Process

Our interest is to obtain an expression for the entropy of Poisson process during a small time window with precision ε. We further assume that λ ε << 1. The entropy is defined as

$$\hat{H}(t) = -\sum_{n=0}^{\infty} P_n(t) \ln P_n(t) \tag{6}$$

$$= -\sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \{-\lambda t + n \ln \lambda t - \ln n!\}$$

$$= \lambda t - \lambda t \ln \lambda t + \{(\ln 2!) \frac{e^{-\lambda t}(\lambda t)^n}{n!} \ldots$$

$$= \lambda t(1 - \ln \lambda t) + O(\varepsilon^2) + \ldots$$

$$H = \lambda - \lambda \ln \lambda \varepsilon$$

$$= \lambda (1 - \ln \lambda \varepsilon)$$

Output inter spike intervals are independent so the mutual information between the input and output spike train per unit time is given by

$$I_{Io} = I\left(s(t);o(t)\right) = I(\{t_0 \dots t_k \dots\};\{t_0^* \dots t_k^* \dots\}) \quad (7)$$

$$= R.I(\{t_0 \dots t_k \dots\};T^r) \quad (8)$$

where $R = \langle T^* \rangle^{-1}$ is the rate of the output spikes.

So spike times are restricted to those in the interval $[t^r, t^r + T^*]$. Where t^r is the timing of the mutual information per output spike last output spike and the timing of the input spike is measure with respect to t^r

Thus

$$I(\{t_0 \dots t_j \dots\};T^*) = H(T^*) - H(T^*|_{t_0 \dots t_k \dots})$$

is the mutual information per output spike where T^* is the ISI of the output train.

The entropies [5,6] are

$$H(T^r) = -\int_{-\infty}^{\infty} p(t^r) \ln p(t^r) dt^r \quad \text{and}$$

$$H(T^*|_{t_0 \dots t_k \dots}) = -\left(\int_{-\infty}^{\infty} p(t^r|_{t_0 \dots t_k \dots}) \ln p(t^r|_{t_0 \dots t_k \dots}) dt^r\right)_{\{|_{t_0 \dots t_k \dots}\}}$$

As given in Deco and Schurmann the loss of information is $L = (H_{in} - I_{io})/H_{in}$.

4 Simulations

Diffusion processes represented with equation 1 is integrated numerically by discretizing as following

$$V(t + \Delta t) = V(t) + \left(-\frac{V(t)}{\tau} + \mu\right)\Delta t + \sigma\sqrt{\Delta t}\, v + w\, \Delta S(t) \quad (9)$$

where v is the standard Gaussian noise and

$$\Delta S(t) = \int_t^{(t+\Delta t)} [\sum_i \delta(t - t_i)] dt$$

represents the number of input spikes t_i between t and $t + \Delta t$.

The simulation results are shown in figure. (For a single neuron)

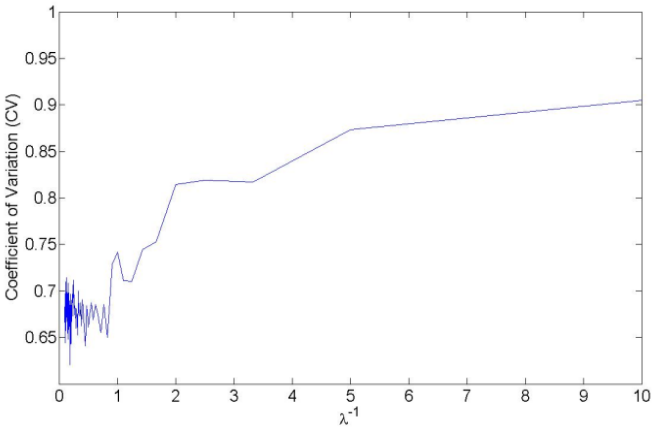


Fig. 1. Input Mean Time vs CV

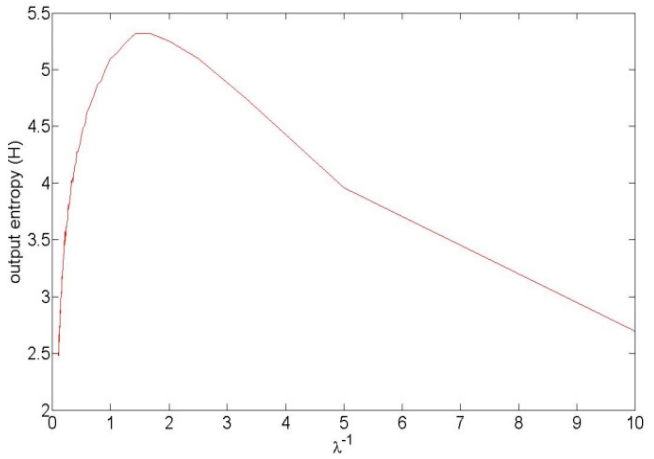


Fig. 2. Input Mean Time vs Output Entropy H

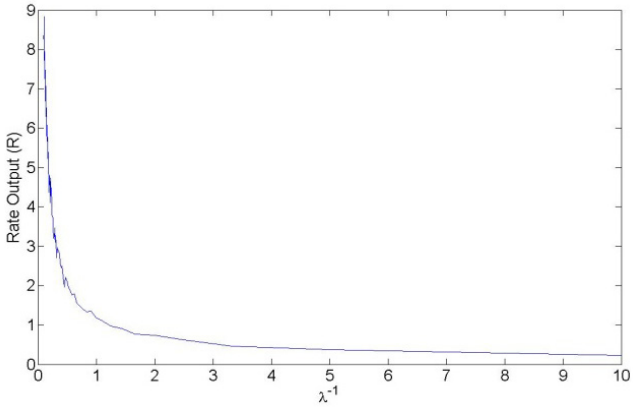


Fig. 3. Input Mean Time vs Rate Output (R)

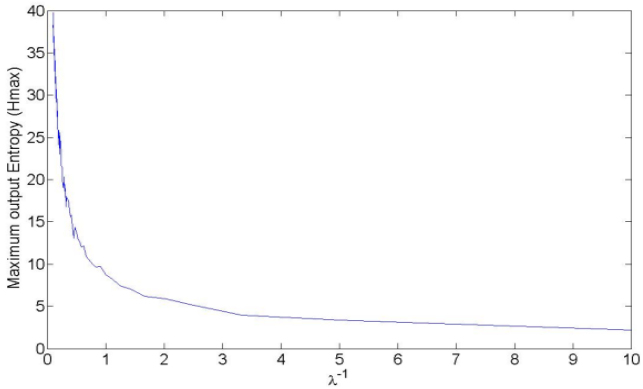


Fig. 4. Input Mean Time vs Maximum Entropy (H_{max})

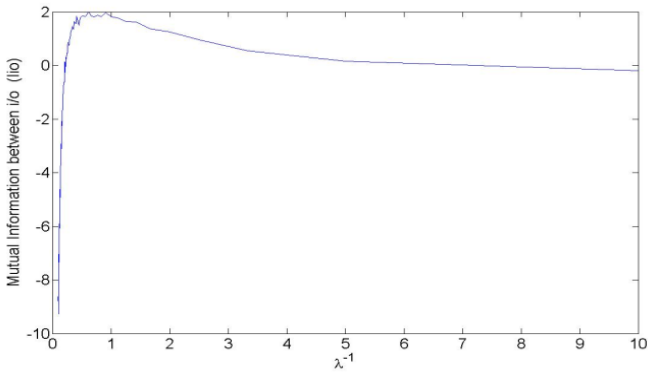


Fig. 5. Input Mean Time vs Mutual Information i/o

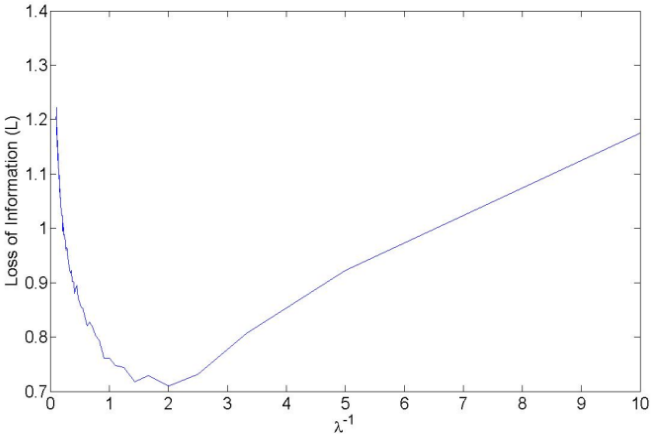


Fig. 6. Input Mean Time vs Loss of information (L)

5 Conclusions and Future Work

While examining information – theoretic analysis of single neuron we observed that the maximum efficiency in the transmission of information is not reached in the Poisson regime but just before it and when the output CV is high than also high transmission of information. Now for future work it is important to examine the effect of correlated inputs. This will throw the light on the effect of degree of correlation in the spiking pattern and ISI distribution.

References

1. Saarinen, A., Linne, M.-L., Yli-Harja, O.: Modeling single neuron behavior using stochastic differential equations. Journal of Neurocomputing 69(10-12) (June 2006)
2. Dayan, P., Abbott, L.F.: Theoretical Neuroscience "Computational and Mathematical Modeling of neural system". MIT Press (2001)
3. Gabbiani, C., Koch, C.: Principles of Spike Train Analysis in Methods in Neural Modeling: From Ions to Network. In: Koch, C., Segev, I. (eds.). MIT Press (1998); Burkitt, A.: A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. Biological Cybernetics 95(1), 1–19 (2006)
4. Gerstner, W., Kistler, W.M.: Spiking neuron models. Cambridge Univ. Press, Cambridge (2002)
5. Deco, G., Schürmann, B.: Information Transmission and Temporal Code in Central Spiking Neurons. Physical Review Letters 79(23) (December 8, 1997)
6. Deco, G., Schürmann, B.: Information Dynamics: Foundation and Applications. Springer (2001)
7. Kistler, W., Gerstner, W., et al.: Reduction of Hodking_Huxley equations to a single variable threshold model. Neural Computation 9, 1015–1045 (1997)

8. Koch, C.: Biophysics of Computation: Information Processing in Single Neurons. Oxford University Press, New York (1998)
9. Di Maio, V., Lansky, P., Rodriguez, R.: Different Types of Noise in Leaky Integrate and fire model of Neuronal Dynamics with Discrete Periodical Input. Gen. Physiol. Biophys. 23, 21–38 (2004)
10. Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., Bialek, W.: Entropy and Information in Neural Spike Trains. Physical Review Letters 80(1) (January 1998)
11. Softky, W., Koch, C.: J. Neuroscience 13, 334 (1993)

Nash Equilibrium and Marcov Chains to Enhance Game Theoretic Approach for Vanet Security

Prabhakar M.[1], J.N. Singh[2], and G. Mahadevan[3]

[1] Anna University of Technology, Coimbatore, Tamilnadu, India
[2] IT, Sambhram Institute of Technology, Bangalore, India
[3] CSE, AMC Engg. College, Bangalore, India
laxmi.prabakar@gmail.com

Abstract. Increased reliance on the Internet has made information and communication systems more vulnerable to security attacks in VANET. Instead of designing a defense against an attack, Game Theory attempts to design a defense against a sophisticated attacker who plans in anticipation of a complex defense in VANET. In addition, Game Theory can model issues of trust, incentives, and externalities that arise in security systems. This paper presents a Game theoretic potential for VANET security. The new technique termed NE (Nash Equilibrium) with Markov chains has been computed based on the game model, for VANET security. In this paper, Markov chains (MC) are used to choose the appropriate model for a given security problem in VANET and we have examined this new approach to extend the basic ideas of using game theory in VANET to predict transition rates for enhancing the security. The experimental evaluation will show that the Game theoretic approach of security measures for VANET using NE and MC perform better.

Keywords: VANET, Nash Equilibrium, Markov Chain.

1 Introduction

A VANET technology which uses every moving car as nodes to form a network and it defines an efficient way of vehicular networking. Since the mobility rate [6] of vehicle has been changed, the topology of the network will never be changed largely. This provides an integrated service for giving an effective and simple communication for vehicle mobility. To make a travel more convenient for the user, the VANET has provided a good service and the users' request has also been processed efficiently. As information and communication technologies have now become an integral and indispensable part of our daily life, there has been a significant shift in the global threat landscape. Today, the greatest security threats are coming to enhance the game theoretic approach for VANET security.

Game theory approach address these troubles, as it features are capable to hold VANET security which is listed below, i) Several players with different goals struggle and interact with each other and they are ii) Used in several authority i.e., finances, decision theory, and control. Game theory provide mathematical framework for analysis, modeling, and decision processes for VANET security. Game theory authorizes

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 191–199.
springerlink.com

extra modeling of attacker [7] behavior and communication between defense and attackers Compared with a pure optimization approach. Mathematical abstraction (framework) is useful for generalization of problems, merging existing ad hoc schemes in single window.

Recently, game theoretic models are used to address network security issues. The concerns are carried over by Game Theory for information security problem [4]. In Game Theory, one player's result depends not only on his decisions, but also on those of his adversary. Likewise, the achievement of a security scheme depends not only on the real protection strategies that have been realized, but also on the deliberate actions taken by the attackers to initiate their attacks. It also depends on the actions of the users that are sharing the systems, and on the actions of their peers situated in other networks. All these agents act realistically according to their various incentives. Game Theory also helps the agents predict each other's behavior and suggests a course of action to be taken in any given situation.

2 Literature Review

In general, a VANET can be modeled as a weighted graph G = (N, E). N contains the network nodes (eg; moving cars), while E is the set of edges. To improve the security of VANET, many techniques have been presented earlier. One of the technique named TPM hardware [Wagan, A.A. ; Mughal, B.M.; Hasbullah, H.,2010] [7], for the security framework of VANET by developing the trusted infrastructure for neighboring nodes.

To improve the communication security with game theoretic approach, [Assane Gueye, 2011] provided a Nash Equilibrium (NE) technique. The NE has been used here for variety of game applications. The variety of games like blocking games, intelligent virus attacking games and so on. [Samara, G. ; Al-Salihy, W.A.H.; Sures.R, 2010] [8] analyzed a security issues of VANET environment.

3 Enhancing Game Theoretic Approach for Vanet Security

The basic assumption of a game theoretic model is that decisions makers are *rational* and take into account the *rationality* [7] of other decision makers. Players are termed as decision makers in a game and comprise the crucial entities of a game. A player can represent a person, a machine or a group of persons. Within a game, players perform *actions* that they describe from their particular *action sets*. The plan of actions that a given player takes during a game play is called the *strategy* of that player. When all players play their strategies, it leads to an *outcome*, which is the vector of the actions taken by the players in that game play. An outcome gives a positive or negative reward (prn) to each player. Being rational, each player is trying to choose the approach that capitalizes the received prn. The prn of a given player is derived from the *preference* that the player has of some outcome compared to others. A player's preference is given by a utility function [3] that dispenses to each outcome a real number.

3.1 Markov Chains for Game Model

The construction of Markov chains are used to measure the security and provide the game model based on the transitions in VANET. It is an arbitrary process differentiated as the next state depends only on the present state and not on the series of events that headed it. The Markov model is used to identify the probability of state transition from one state to another [8]. Depends upon the users' actions states in VANET, it construct the Markov model of the system. To construct the Markov model,

Step1: Let the source be T_i
Step2: Let Destination be T_j
Step3: if transition from the state T_i *to the state* T_j
Step4: Compute actions profiles of $\rho_1, \rho_2 \ldots \rho_k$
With the rates of $\mu\rho_1 \ldots \mu\rho_k$ *consequently,*
Step5: Evaluate the total rate of the state transition from T_i *to the state* T_j*:*
$\lambda_{ij} = \Sigma^{k}_{n=1} \mu\rho_n$.....(1)
Step6: End If

The Markov chains are used to identify the security state of the nodes in the network. The above pseudo code described the state transition of nodes from source to destination. Before transmitting the nodes in the given way, it is necessary to evaluate the action profiles for each node in the network. Then choose the transition path and evaluate the rate of the transition path using Equation 1. Therefore, for computing the transition rates in Markov chain, it is enough to compute the rates of $\mu\rho_i$.

If in the state *s* of the system has equal rate of transition, then for action profile of $\rho_i = \rho_{r1}, \rho_{r2} \ldots \rho_k$ with occurrence probabilities ($\pi^{*,s}{}_1(\rho_{r1}) \ldots \pi^{*,s}{}_k(\rho_k)$), the rate will be as follows:

Step7: if transition has equal state,
Step8: Compute the action profiles $\rho_i = \rho_{r1}, \rho_{r2} \ldots \rho_{rk}$
Step9: for each action profiles,
Step 10: Evaluate the total rate of the state transition
$\mu\rho_i = \mu * \pi^{*,s}{}_j(\rho_j)$.....(2)
Step 11: end for
Step12: end if

Therefore, to compute the rates of action profiles, it is enough to compute the rate of actors and the occurrence probabilities of the actions. The later is computed by solving the game model and the former is computed as follows.

Step 13: If users' actions on the state path fail,
With probabilities as $\gamma_1(1), \gamma_2(2), \ldots \gamma_{n1}(n1)$,
Step 14: The state of those actions rate is computes as
$\mu^{*,i}{}_1 = \mu_1 * \Sigma^{n1}_{k=1}(1 - \gamma_1(k))$.....(3)
Step 15: derive the game model

The formulas referred in [2] for computing the Markov chains, to derive the game model. Based on users' state transition, the Markov model is being designed.

Depends upon the users' action specified, the state of the transition rate is modified. The state transition rate is computed depended upon the action profiles done by the user. The game model is derived based on Markov chains where the current state of a variable or system is free of all past states, except the present state.

The proposed Markov chains and Nash Equilibrium for VANET security is shown in fig 1.

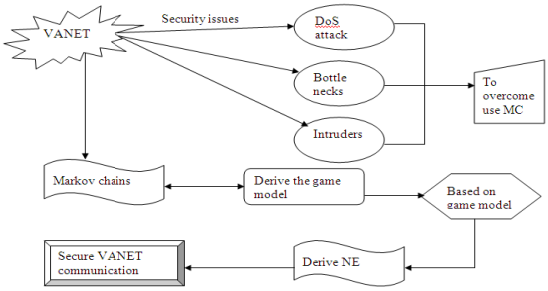


Fig. 1. Architecture Diagram of Game theoretic approach using NE and MC

3.2. Nash Equilibrium

Nash Equilibrium (NE) is applied in game theory when two or more number of players involved in game under VANET. It is computed based on the game model like intelligent virus game, blocking game and so on which are derived from MC (Markov Chains). There is a chance of issue raised for computing the NE for game model. So to overcome the issue, this paper described the process of deriving the NE based on the players involved in game theoretic model in VANET security. It can also be termed as mixed strategies, where players involved in game select the probabilities from possible actions made in it.

The pseudo code for NE to secure the VANET using the game model derived using MC.

Step1: Derive the game model using MC
Step2: Let the game be G, and players be n
Step3: Compute the strategy set Ti
Step4 : For each player,
Step5: Compute the payoff $f_i(x)$
Step6: end for
Step 7: Based on $f_i(x)$
Step 8; Evaluate NE (eqn 4)
Step 9: To identify the state of the game strategy
Step 10: End

Let (T,g) be a game with n players, where T_i is the strategy set for player i, $T=T_1 X g_2 ... X g_n$ is the set of approach and $g=(g_1(x), ..., g_n(x))$ is the payoff function for $x \in T$. Let x_i be a approach of player i and x_{-i} be a approach of all players except for player i. When each player $i \in \{1, 2, ..., n\}$ selects approach x_i resulting in approach profile $x = (x_1, ..., x_n)$ then player i gets payoff $f_i(x)$. Note that the payoff depends on

the approach chosen, i.e., on the approach selected by player *i* as well as the approaches by all the other players. A approach $x^* \in S$ is a Nash equilibrium (NE) if no unilateral variation in approach by any single player is beneficial for that player, that is [13]

$$\forall i, x_i \in S_i, x_i \neq x_i^* : f_i(x_i^*, x_{-i}^*) \geq f_i(x_i, x_{-i}^*) \quad \text{..... (4)}$$

The game model is finally derived using MC and NE successfully to improve the efficiency of VANET security.

4 Performance Evaluation

The proposed VANET security using NE and MC is evaluated in an efficient manner using NS2 simulator. Initially the experiment is evaluated with 100 nodes in a flat area of 100*100 m^2. The nodes' incoming time measured in terms of seconds is noted as t_1, t_2….t_n. Using MC, it derived the game model. Based on game model, we have to apply NE for a secure communication. The simulation results show that it takes 800 secs to travel from source to destination by choosing the path efficiently without any interruption. In the simulations, two particular scenarios are studied: first one rural and the other urban, which vary from each other in road and traffic density.

The penalty for the attacker when both players decide the identical road segment, are initally set to r = 0.2 approximately interpreted as 20 percent loss. From penalty, the equilibrium value (NE) (Equation 5) of the game amounts to v = w = 0.4145. If the penalty is decreased to 0.01, the value raises to 0.6560, *i.e.,* a gain for the attacker as predictable.

5 Results and Discussion

Let us assume a 1000x1000 grid of a map where nodes can travel randomly for vehicular network environment (VANETS). Normally, in the urban city, the vehicle density is high. These nodes attempt and communicate with each other depends on another random set of connections of TCP or UDP. We use a simple maliciousness model to inject malicious behavior into the system. At this situation, NE and MC for VANET security perform well. The parameters are carried over here for VANET security using NE and MC. All simulations are run 20 times over and averaged.

Table 1. Attack and Defense rate of VANET security in rural scenario

Vehicle Density	Probability (rural scenario)	
	Attack factor	Defense factor
10	0.06	0.09
20	0.13	0.15
30	0.16	0.2
40	0.21	0.25
50	0.25	0.29

Table 1 describes the probability of attack and defense factor based on vehicle density in a rural scenario.

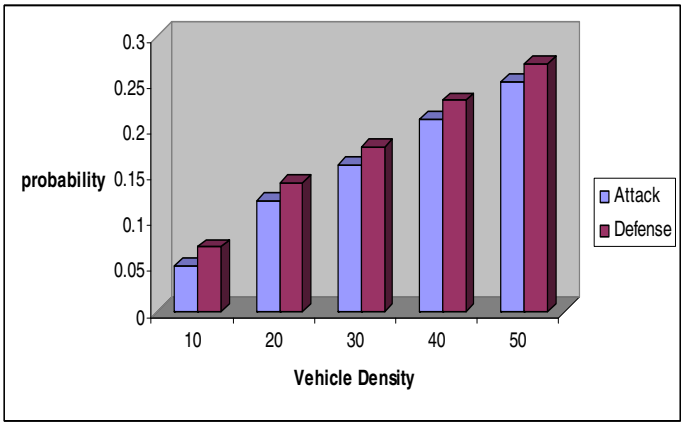


Fig. 2. Attack and Defense rate of VANET security in rural scenario

Fig 2 describes the attack and defense rate of VANET security using NE and MC. In rural scenario, the vehicle density might be low. So when the vehicle density increases, the probability of defense rate for the attack is high to show the better performance. The probability of attack is also being low compared to urban scenarios. The attack and defense rate is measured in terms of probability factor. When using NE and MC for VANET security, the attack rate is very low compared to an existing ACO for VANET security.

Table 2. Attack and Defense rate of VANET security in urban scenario

Vehicle Density	Probability (urban scenario)	
	Attack factor	Defense factor
10	0.05	0.09
20	0.16	0.19
30	0.09	0.1
40	0.14	0.17
50	0.27	0.3

Table 1 describes the probability of attack and defense factor based on vehicle density in an urban scenario.

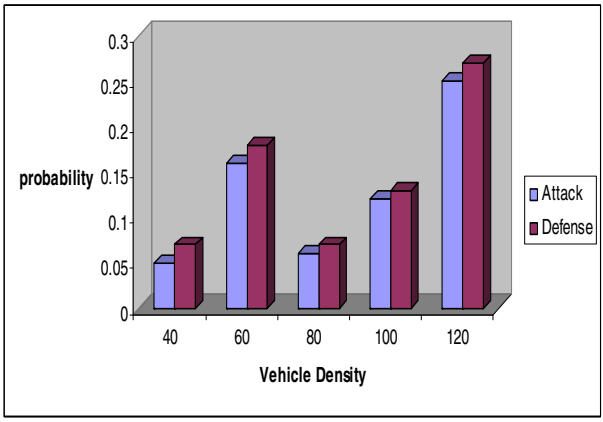


Fig. 3. Attack and Defense rate of VANET security in urban scenario

Fig 3 describes the attack and defense rate of VANET security in an urban scenario using NE and MC. In urban scenario, the vehicle density might be high. So when the vehicle density increases, the probability of attack is high and so the defense rate should also be high to show the better performance. The attack and defense rate is measured in terms of probability factor. When using NE and MC for VANET security in urban scenario, the attack rate is very high and defense rate is also being high compared to an existing ACO for VANET security.

Table 3. Performance Rate of VANET security using NE and NC

Probability of Attack	VANET security using NE and MC
	Performance Rate
0.04	25
0.08	34
0.12	30
0.16	35
0.2	33

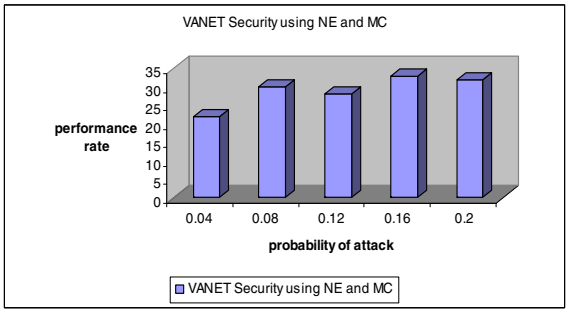


Fig. 4. Performance measure of VANET security

Fig 4 describes the performance measure of VANET security using NE and MC for game theoretic approach. When the probability of attack for VANET security increases, the performance for securing the VANET environment is also be high by using NE and MC. Since the Nash Equilibrium is used for VANET security, the state of the vehicles are maintained stable and it does not allow the intruders to intrude the vehicle in a specific interval of time.

Finally, it is observed that the proposed MC and NE for VANET security efficiently designed using NS2 simulator. The game model is derived based on users' action profiles in the state transition paths using Markov chain model. Based on game model, derive the equilibrium state and identify the stable state of the game model.

6 Conclusion

To improve the game theoretic approach in VANET security, this paper presented a new technique MC (Markov Chains) and NE (Nash Equilibrium). The MC here presented an efficient game model based on the issues raised in the VANET environment. The NE has been computed efficiently based on the game model. By the way, the issues over VANET have been decreased. The security game outperforms the raw approach of protecting locations equivalent to their methods by ignoring attacker behavior. The numerical analysis is based on the sensible simulation data attained from traffic engineering systems. The simulation results showed that the game theoretic approach for VANET security using MC and NE is the best to use.

References

1. Wagan, A.A., Mughal, B.M., Hasbullah, H.: VANET Security Framework for Trusted Grouping using TPM Hardware. In: 2010 Second International Conference on Communication Software and Networks (2010)
2. Moayedi, B.Z., Azgomi, M.A.: A Game Theoretic Approach for Quantitative Evaluation of Security by Considering Hackers with Diverse Behaviors. In: 2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing (2009)
3. Alpcan, T., Buchegger, S.: Security Games for Vehicular Networks. IEEE Transactions on Mobile Computing 10 (2011)
4. Grossklags, J., Christin, N., Chuang, J.: Secure or Insure? A Game-Theoretic Analysis of Information Security Games. In: Proc. 17th Int'l Conf. World Wide Web (WWW 2008, pp. 209–218 (2008)
5. Grossklags, J., Christin, N., Chuang, J.: Predicted and Observed User Behavior in the Weakest-Link Security Game. In: Proc. First Conf. Usability, Psychology, and Security (UPSEC 2008), pp. 1–6 (2008)
6. Do, Y., Buchegger, S., Alpcan, T., Hubaux, J.-P.: Centrality Analysis in Vehicular Ad-Hoc Networks. technical report, EPFL/T-Labs (2008)

7. Wagan, A.A., Mughal, B.M., Hasbullah, H.: VANET Security Framework for Trusted Grouping Using TPM Hardware. In: Second International Conference on Communication Software and Networks, ICCSN 2010, March 25 (2010)
8. Samara, G., Al-Salihy, W.A.H., Sures, R.: Security issues and challenges of Vehicular Ad Hoc Networks (VANET). In: 2010 4th International Conference on New Trends in Information Science and Service Science (NISS), June 17 (2010)
9. Mehraeen, S., Dierks, T., Jagannathan, S., Crow, M.L.: Zero-sum two-player game theoretic formulation of affine nonlinear discrete-time systems using neural networks. In: The 2010 International Joint Conference on Neural Networks (IJCNN), October 14 (2010)

Fast Computation of Image Scaling Algorithms Using Frequency Domain Approach

Prasantha H.S., Shashidhara H.L., and K.N.B. Murthy

PES Institute of Technology, 100 Ft Ring road, BSK 3RD stage, Bangalore-85
Prashanth_34@rediffmail.com,
shashihl@yahoo.com,
principal@pes.edu

Abstract. Image scaling algorithms play important role in many image scaling applications. Image scaling is the process of enlarging an image or reducing the size of an image to make it suitable to display on the given display device. The paper mainly focuses image zooming to fit the given image on a display device to view the details on a bigger display device. When the image is zoomed, artifacts like blurring, jagging and ghosting may arise. The main objective of the paper is to investigate and study the known algorithms for image scaling based on different comparative parameters in frequency domain. The different interpolation techniques such as nearest neighbor and bilinear are studied and compared in both spatial and frequency domain. The paper proposes a novel scheme for fast computation of the different image interpolation techniques.

1 Introduction

An image may be defined as a two-dimensional function f(x, y), where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. The field of digital image processing refers to processing digital images by means of a digital computer. Digital image processing involves the use of computer algorithms to perform image processing on digital images. The usefulness is apparent in many different disciplines covering medicine through remote sensing. The application of digital image processing includes medical applications, restorations and enhancements, digital cinema, image scaling, image compression, image transmission and coding, color processing, remote sensing, robot vision, facsimile, etc.

A fast algorithm for image interpolation is proposed for real-time enlargement of video images is discussed [8]. A novel method of nonlinear scaling is proposed to overcome the compatibility problem [7]. An adaptive resampling algorithm for zooming up images is based on analyzing the local structure of the image and applying a near optimal and least time-consuming resampling function that will preserve edge locations and their contrast. A method is proposed to take into account information about discontinuities or sharp luminance variations while doubling the input picture [5]. Popular methods such as bilinear Interpolation, cubic convolution and cubic convolution are investigated [4]. The different measures can be used to compare the different algorithms of interpolation [11] [6] [5] [10].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 201–208.
springerlink.com

Section 1 discusses the introduction, Section 2 briefs about image scaling, Section 3 explains the different scaling techniques considered, Section 4 discusses the implementation details and Section 5 discusses the test results. Conclusions are discussed in the Section 6.

2 Image Scaling

In computer graphics, image scaling is the process of resizing a digital image. In the case of scaling up (image enlargement), the methods used involve first resizing the image by padding with zeros, followed by convolution with a two-dimensional, spatially invariant linear filter. Since we are dealing with matrices that represent discrete pixel values, we use the filter in the discrete domain. Hence each filter can be represented by the impulse function h, which is also called the mask of the filter. Scaling down (image reduction) increases the incidence of high frequencies and causes several pixels to collapse into one. Hence we need to apply a smoothing filter in order to minimize aliasing problems in the target image. Apart from fitting a smaller display area, image size is most commonly decreased (or sub-sampled or down-sampled) in order to produce thumbnails. There are several methods of increasing the number of pixels that an image contains, which evens out the appearance of the original pixels.

Non-adaptive algorithms include: nearest neighbor, bilinear, spline, sinc, lanczos and others. The more adjacent pixels they include, the more accurate they can become, but this comes at the expense of much longer processing time. Interpolation is the problem of approximating the value for a non-given point in some space, when given some values of points around that point. Nearest neighbor interpolation is the simplest method and basically selects the value of the nearest point, and assigns that value to the output point and does not consider the values of other neighboring points at all, yielding a piecewise-constant interpolate. Bilinear interpolation determines the grey level value from the weighted average of the four closest pixels to the specified input coordinates, and assigns that value to the output coordinates.

3 Image Scaling in the Frequency Domain

Image transforms are extensively used in image processing and image analysis. Transform is a basically mathematical tool, which allows us to move from one domain to another domain (time domain to frequency domain). Image transforms are useful for fast computation of convolution and correlation. The transforms do not change the information content present in the signal. Transforms play a significant role in various image processing applications such as image analysis, image enhancement, and image filtering and image compression as well. The frequency domain representation clusters the image data according to their frequency distribution. In frequency domain filtering, the image data is dissected into various spectral bands, where each band depicts a specific range of details within the image. The process of selective frequency inclusion or exclusion is termed “Frequency domain filtering”.

3.1 Fast Computation of Nearest Neighbor Interpolation

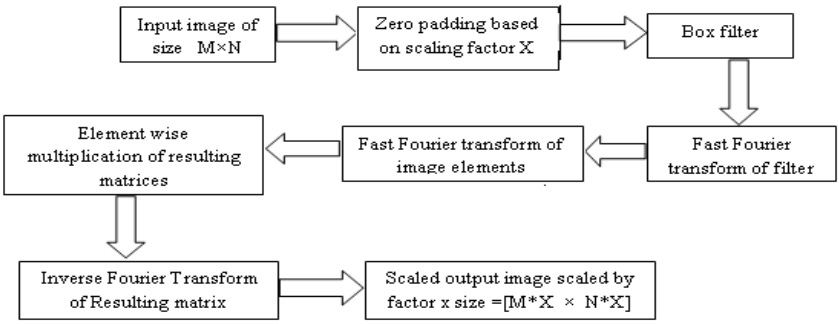


Fig. 1. Nearest Neighbor interpolation using frequency domain approach

In nearest neighbor method, replace each row by a number of identical rows, and likewise for columns. This is known as scaling by pixel replication. In formal terms, this is achieved by applying a box interpolation filter. Hence the rows of zeros have been replaced with a copy of the pixel values directly above them, and the columns of zeros took on the values of the pixels to their immediate left. In general, to scale an image x times in the horizontal direction and y times in the vertical direction, we need a mask of size y rows and x columns with all the elements taking a value of 1. If the original image had n rows and m columns, and we want to scale it up by a factor of x in the horizontal direction and y in the vertical direction, then the zero-padded image will have x*n+1 rows and y*n+1 columns.

3.2 Fast Computation of Bilinear Interpolation

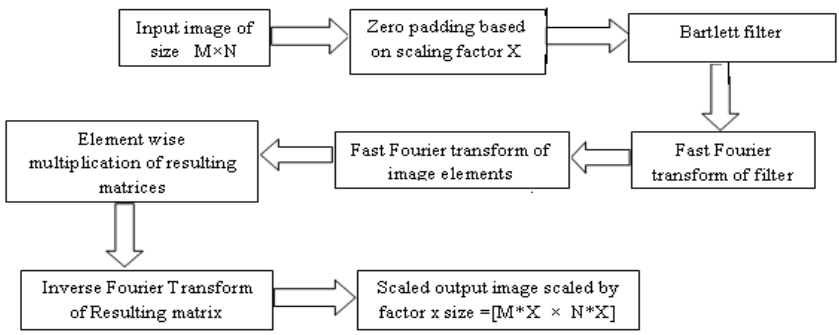


Fig. 2. Bilinear interpolation using frequency domain approach

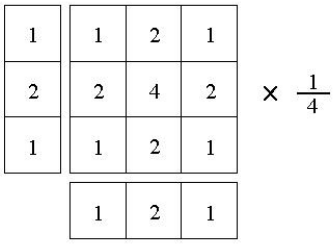


Fig. 3. Filter co-efficient

The figure 3 shows a matrix of a Bartlett filter used to scale an image up by a factor of 2 in both the horizontal and vertical directions. For example, in the case of scaling up 2 in both the horizontal and vertical directions, since each pixel is mapped into 4 pixels in the resulting image, we want the intensity to be 4 times greater than that of the original image in order to maintain the same average intensity level.

4 Implementation Details

The experiments are conducted at PESIT Multimedia Lab for different sets of input images of various file formats. The different interpolation algorithms considered are nearest neighbor, bilinear and spline interpolation for the comparison using different parameters such as MSE, PSNR, and quality index. The digital signal processor multimedia developer kit DM 642 operating at 600 MHZ is considered for the experimentation. All the source files are written in C programming language. The DM642 DSP is capable of executing eight instructions in parallel at the clock speed of 600 MHz If we are able to utilize these execution units then we can benefit from its processing power and obtain a high performance solution. Otherwise, the performance may not be satisfactory. The un-optimized codes for different interpolation techniques are considered for validation.

5 Test Results and Discussions

Experiments are conducted for different set of images of different resolution and file formats. A sample of the result is displayed for the further discussion. Different comparison parameters such as MSE, PSNR, Quality index (QI) and computation time are considered.

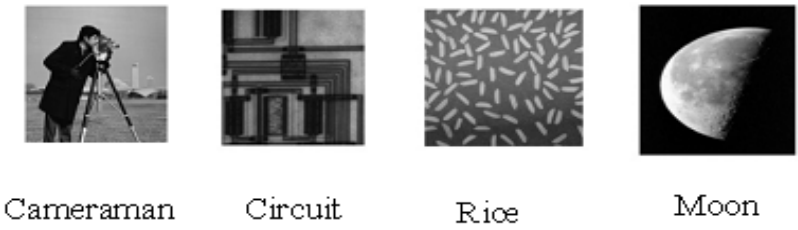


Fig. 4. Input images considered for the discussion

Fig. (a). Input image

Fig. (b). Box filter

Fig. (c). Bartlett filer

Fig. 5. Cameraman image (Scaled by factor 2 in both horizontal & vertical direction)

Fig.(a). Input image

Fig. (b). box filter

Fig. (c). Bartlett filer

Fig. 6. Cameraman image (Scaled by factor 4 in both horizontal and vertical direction)

The figure 4 indicates some of input images considered for the experimentation. The figure 5 and 6 indicates the original image and the subjective quality of the nearest neighbor and bilinear interpolation. The tables 1, 2, 3 and 4 shows the comparison of different parameters such as Mean Square Error (MSE), Peak Signal to Noise Ratio (PSNR) and Quality index (QI) for different scaling factors 2, 4, and 8 for the cameraman, circuit, rice and moon images. The table 5 shows the result for spatial domain computation time and frequency domain computation time. Computation time is the total time involved to obtain the zoomed image.

Table 1. Table showing different comparison parameters for the Cameraman image

Scale factor	Nearest neighbor (Box filter)			Bilinear (Bartlett filter)		
	MSE	PSNR	QI	MSE	PSNR	QI
2	376.8179	22.3695	0.9515	261.2432	23.9604	0.9651
4	729.5694	19.5001	0.9066	497.8059	21.1602	0.9325
8	1193.9	17.3612	0.8454	807.7252	19.0582	0.8865

Table 2. Table showing different comparison parameters for the Circuit image

Scale factor	Nearest neighbor (Box filter)			Bilinear (Bartlett filter)		
	MSE	PSNR	QI	MSE	PSNR	QI
2	206.2694	24.9865	0.9830	133.5975	26.8728	0.9893
4	428.847	21.8078	0.9574	255.7305	24.053	0.9721
8	927.0181	18.4599	0.8910	590.7553	20.4167	0.9255

Table 3. Table showing different comparison parameters for the Rice image

Scale factor	Nearest neighbor (Box filter)			Bilinear (Bartlett filter)		
	MSE	PSNR	QI	MSE	PSNR	QI
2	372.82	22.42	0.943	201.24	25.093	0.961
4	734.564	19.47	0.88	407.805	22.026	0.921
8	98	18.21	0.7463	653.725	19.976	0.801

Table 4. Table showing different comparison parameters for the Moon image

Scale factor	Nearest neighbor (Box filter)			Bilinear (Bartlett filter)		
	MSE	PSNR	QI	MSE	PSNR	QI
2	86.2853	28.7714	0.9973	53.276	30.865	0.9985
4	215.948	24.7873	0.9924	136.1032	26.792	0.9965
8	563.3191	20.6233	0.9817	356.579	22.609	0.9903

Table 5. Comparison of computation time using spatial and frequency domain methods

Scale factor	Spatial domain (in milliseconds)		Frequency domain (in milliseconds)	
	Nearest neighbor	Bilinear	Nearest neighbor	Bilinear
2	19	42	14	40
4	34	80	28	71
8	55	125	49	112

From the experimental results, it is clear that bilinear interpolation algorithmic yields better results than nearest neighbor interpolation. Bilinear interpolation has higher value of PSNR and Quality index for scaling factors 2, 4 and 8 which is shown in the table 1, 2, 3 and 4 for all the input images considered. For any scaling factors, bilinear interpolation has higher value of PSNR and quality index. Also the subjective quality of the image for bilinear interpolation is better than nearest neighbor interpolation which is shown in figure 5 and 6. For the moon image shown in the table 4, PSNR and Quality index for any scaling factor is high which indicates that if the variation in the intensity levels for the given image is less, then it is possible to scale-up or zoom the image without sacrificing the quality. Similarly for cameraman image shown in the table 1, since the variations are more, if we scale the image, quality also suffers. The computation time required using frequency domain methods are comparatively less compared to spatial domain methods which is shown in table 5.

6 Conclusions

Bilinear interpolation can be used to zoom the image since it provides better results. But the main problem with bilinear interpolation is the computation time. Hence it is required to reduce the computation time which can be done by using frequency domain methods. Since filter based techniques are designed using frequency domain approach hence there is comparable reduction in computation time and complexity compared to spatial domain approach. Based on results, it is clear that even in frequency domain box filter based algorithm requires less time than Bartlett filter based algorithm but Bartlett filter usage yielding better quality image than box filter use. Also computation time can be further reduced by using fast computing transforms in place of FFT and IFFT.

References

1. Prasantha, H.S., Shashidhara, H.L., Balasubramanya Murthy, K.N.: Comparative analysis of different interpolation schemes in image processing. In: International Conference on Advanced Communication Systems (ICACS), India, pp. 17–24 (January 2007)
2. Prasantha, H.S., Shashidhara, H.L., Balasubramanya Murthy, K.N.: Image Scaling comparison using Universal Image Quality Index. In: International Conference on Advances in Computing, Control and Telecommunication Technologies (ICACCTT), India, pp. 859–863 (December 2009)

3. Gao, R., et al.: Image zooming algorithm based on partial differential equations technique. International Journal of Numerical Analysis and Modeling 6(2), 284–292
4. Matsuoka, R., et al.: Comparison of Image Interpolation Methods Applied to Least Squares Matching. In: CIMCA (2008)
5. Battiato, S., et al.: A locally adaptive zooming algorithm for digital images. Image and Vision Computing, 805–812 (2002)
6. Yuan, S., et al.: High accuracy WADI image interpolation with local gradient features. In: Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication System, Hongkong (2005)
7. Shi, Z., et al.: A novel nonlinear scaling method for video images. In: International Conference on Computer Science & Software Engineering, pp. 357–360.
8. Mihov, S.G., et al.: Interpolation algorithms for image scaling. Electronics (2005)
9. Lehmann, T.M., et al.: Survey: Interpolation Methods in Medical Image Processing. IEEE Transactions on Medical Imaging 18(11) (1999)
10. Acharya, T., et al.: Computational Foundations of image interpolation algorithms. ACM Ubiquity 8 (2007)
11. Wang, Z., et al.: A universal image quality Index. IEEE Signal Processing Letters XX(Y) (2002)

Word Level Script Identification of Text in Low Resolution Images of Display Boards Using Wavelet Features

S.A. Angadi and M.M. Kodabagi

Department of Computer Science and Engineering
Basaveshwar Engineering College, Bagalkot-587102, Karnataka, India
vinay_angadi@yahoo.com, malik123_mk@rediffmail.com

Abstract. Automated systems for understanding low resolution images of display boards are facilitating several new applications such as blind assistants, tour guide systems, location aware systems and many more. Script identification at character/word level is one of the very important pre-processing steps for development of such systems prior to further image analysis. In this paper, a new approach for word level script identification of text in low resolution images of display boards is presented. The proposed methodology uses horizontal run statistics and wavelet features for distinguishing 5 Indian scripts namely; Hindi, Kannada, English, Malyalam and Tamil. The method works in two phases; In the first phase, the wavelet transform based texture features such as zone wise wavelet energy features, vertical run statistical features of wavelet coefficients and wavelet log mean deviation features of decomposed energy bands at 2 levels are obtained from training word images and knowledge bases are constructed, one for each script/language under study. The second phase is testing, in which test word image is processed to obtain horizontal run statistics to determine whether it belongs to Hindi script. Otherwise, a newly defined descriminant function that measures the city block distance between test sample and pre-constructed knowledge base of every script is used to identify the script of the test sample. The proposed method is robust and insensitive to the variations in size and style of font, number of characters, thickness and spacing between characters, noise, and other degradations. The proposed method achieves an overall identification accuracy of 89.7% and individual identification accuracy of 92% for Kannada Script, 97.67% for English, 82.5% for Malyalam and 87% for Tamil Script.

Keywords: Script Identification, Low Resolution Images, Wavelet Features, City Block Distance, Display Boards.

1 Introduction

In recent years, the camera embedded hand held systems such as smart mobile phones, tablets and PDA's are being widely used and they increasingly provide/exhibit higher computing and communication capabilities. These devices with

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 209–220.
springerlink.com

internet access facilities are being used for wide variety of purposes such as information seeking, mobile commerce and other business and enterprise applications. One such application is to understand written text on display boards in an unknown environment. People who move across different places in the world for field work and business find it difficult to understand written text on display boards particularly in foreign environment. This is especially true in countries like India, which are multilingual. Hence there is a need for a gadget that helps people to understand display boards by detecting and translating written matter while providing localized information.

The written matter on display boards/name boards provides important information for the needs and safety of people, and may be written in unknown languages. The written matter can be street names, restaurant names, building names, company names, traffic directions, warning signs etc. Researchers have focused their attention on development of techniques for understanding written text on such display boards. There is a spurt of activity in the development of web based intelligent hand held systems for such applications and few are summarized in the following and a more elaborate survey of related works is presented in the next section. A mobile context aware tour guide system (*CyberGuide*)**,** which provides route map for the user to navigate and retrieve the information based on user's location and point/object of interest is found in [1]. The context aware communication system (*ComMotion*) for mobile phones is described in [2]. The Image-based Deixis (*IDeixis*) enables hand held device to capture an image and query the web server to retrieve the relevant location information from web [3]. Web based hand held intelligent systems for streaming multimedia, retrieving image based search information, city map navigation to gather information about a point of interest, and mobile learning and many more have been reported and are described in [4-10].

In the reported works on intelligent systems for hand held devices, not many works pertains to understanding of written text on display boards. Therefore, scope exists for exploring such possibilities. The text understanding involves several processing steps; text detection and extraction, preprocessing for line, word and character separation, script identification, text recognition and language translation. In the Indian context, the written text on display board may contain multilingual information. Therefore, recognition and language translation tasks require script identification at word level. Hence, script identification at word level is one of the very important processing steps for development of such systems prior to further analysis. The script identification task also facilitates automation of grouping words of a specific script/language and also separate interlaced words pertaining to other script/languages.

The script identification of text in low resolution images of display boards is a difficult and challenging problem due to various issues such as font size, style, and spacing between characters, skew and other degradations. The reported works on script identification have identified a number of approaches, which are categorized into local and global approaches. The local approaches use connected component analysis process for determining the script of text. The paradox inherent in such approaches is that, the extraction of connected components requires prior knowledge of the script of the document or display board image text. In addition to this, the presence of noise

and other significant degradations in natural scene display board images significantly affects connected component determination and analysis process, thus making such approaches inefficient. In contrast, the global approaches measure the properties of a region/block of text and give sufficient characterization of the underlying script. Hence, texture analysis is good choice for solving such a problem.

In this paper, an approach for word level script identification of low resolution images of display boards using wavelet features is proposed. The method distinguishes input word into five scripts namely; Hindi, Kannada, English, Malyalam and Tamil. The method investigates use of zone wise wavelet energy features, wavelet log mean deviation features and newly obtained properties of wavelet coefficients for solving a practical but hitherto mostly overlooked problem in natural scene image processing-the script identification of text in low resolution images of display boards. The detailed description of the proposed methodology is given in following sections of the paper.

The rest of the paper is organized as follows; the detailed survey related to script/language identification is described in section 2. The proposed method is presented in Section 3. The experimental results and analysis are given in Section 4. Section 5 concludes the work and lists future directions.

2 Related Works

The script identification in a low resolution image of display board is a necessary step for development of various other tasks of display board understanding system. A number of methods for script identification have been published in recent years and are categorized into local and global approaches. The local approaches perform connected component analysis and use statistic based features for script identification. Few such methods are summarized in the following; An approach for determining the script and language of document images is proposed in [11]. Initially, the algorithm determines connected components and locates upward concavities in the connected components. It then classifies the script into two broad classes Han-based (Chinese, Japanese and Korean) and Latin-based (English, French, German and Russian) languages. The Han-based languages are later differentiated using statistics of optical densities of connected components. And Latin-based languages are identified based on the most frequently occurring word shape characteristics.

An automatic technique for the identification of printed Roman, Chinese, Arabic, Devanagari and Bangla text lines from single document image is found in [12]. A method for script and language identification of noisy and degraded document images is described in [13]. This method identifies script based on document vectorization technique that converts each image into vertical cut vector and character extremum points that characterizes the shape and frequency of contained character or word images. The method is tolerant to the variation in text fonts and styles, noise, and various types of document degradation.

In contrast, the global approaches measure the texture of a region of text to identify the underlying script. Some of the texture based approaches are detailed below; The method describing effectiveness of rotation invariant texture features for automatic script identification is found in [14]. The techniques that investigate use of texture analysis for script/language identification from document images are presented in [15-16]. The effectiveness of features extracted from co-occurrence histograms of wavelet decomposed images and KNN classifier for script identification of 7 Indian languages is discussed in [17].

In recent years, the methods that perform similarity analysis for script identification/text recognition/quick reading of words/image analysis are also described in [18-20]. A statistical script identification technique that determines the script of camera-based images which suffer from perspective distortion is discussed in [21].

After the thorough study of literature, it is found that only few works pertain to script identification of text in low resolution images and there is a scope for new method for script/language identification due to limitations of reported works. First, the performance of local approaches depends upon correct segmentation of connected components. Consequently, they are very sensitive to the segmentation error resulting from noise and various types of document degradation. Second, the global techniques need more time to measure the texture of a region. But, these methods are a good choice for analysis of low resolution images of display boards. Hence, use of textural features for script identification task is further investigated in the proposed work.

It is also noticed that, the global techniques, operate on predefined size text blocks containing matter pertaining to same script for determination of script and language of underlying document. But this is not the case with written text on display boards in the Indian scenario, as text may contain multilingual information. Therefore, it is necessary to identify script and language at word level which is essential for later processing steps such as text understanding and language translation. The task of script identification at word level is difficult and challenging, because distinguishing properties are to be obtained from a small region containing text of variable size and font. Therefore more research is desirable to model texture of small region containing text of variable size and font for better characterization and classification with reduced computational complexity. In the current work new properties of texture using wavelet coefficients are employed for script identification of text in low resolution images of display boards. The detailed description of the proposed methodology is given in the next section.

3 Methodology for Word Level Script Identification

The proposed method uses following processing steps such as Preprocessing, Feature Extraction, Construction of Knowledge Bases and Script Class Identification. The block schematic diagram of the proposed model is given in Fig.1. The detailed description of each processing step is presented in the following subsections;

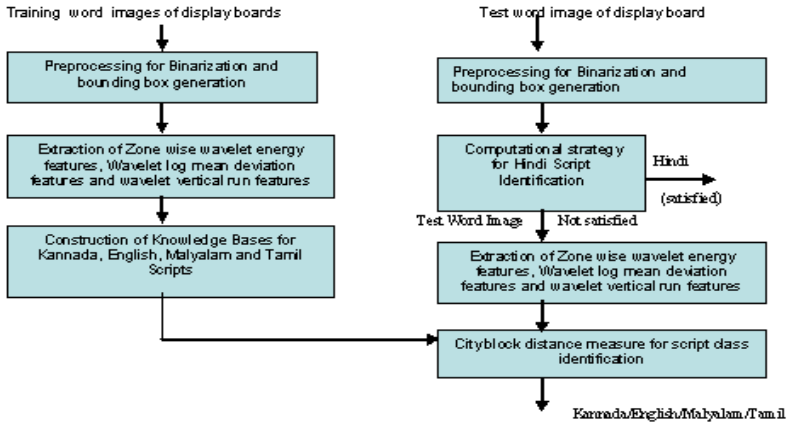


Fig. 1. Block Diagram of Proposed Model

3.1 Preprocessing

The works reported in literature preprocess document image to obtain uniform sized text block, detect and correct skew, and remove uneven spacing between lines, word and characters to obtain optimal texture features for improved classification rate. Because, the presence of noise, skew and uneven spacing and other degradations significantly affect texture features leading to higher classification errors. But the preprocessing task is difficult, computationally expensive and may not be suitable for applications that process small of amount of text containing few lines. Hence, in this work, an attempt is made to evaluate performance of new texture features extracted directly from variable sized word images without preprocessing them. The preprocessing is done to binarize the image and generate bounding box around it.

3.2 Feature Extraction

The wavelet transform has emerged to provide a good framework for multi-scale signal analysis and wavelet coefficients have proved to provide significantly better representation of texture. The reported works on script identification use wavelet band energy features and concurrence signatures at various resolution levels. But these wavelet coefficients have been used for modeling texture of document text. Hence, in this phase, further investigation into the extraction of new texture features from wavelet coefficients suitable for the task of script identification of text in low resolution images of display boards is carried out. Initially, the two dimensional discrete wavelet transform is applied to decompose word image into energy sub-bands at 2 levels as stated in equations (1)-(4).

$$A_j = [\, H_x * [\, H_y * A_{j-1}]\downarrow_{2,1}]\downarrow_{1,2} \tag{1}$$

$$D_{j1} = [\, G_x * [\, H_y * A_{j-1}]\downarrow_{2,1}]\downarrow_{1,2} \tag{2}$$

$$D_{j2} = [\, H_x * [\, G_y * A_{j-1}]\downarrow_{2,1}] \downarrow_{1,2} \tag{3}$$

$$D_{j3} = [\, G_x * [\, G_y * A_{j-1}]\downarrow_{2,1}] \downarrow_{1,2} \tag{4}$$

Where, Aj and Djk are approximation and detail coefficients at each resolution level j, H and G are low and high pass filters, and ↓x,y represents down sampling along each axis by given factors. Then the features are extracted from different zones/regions of detailed sub-bands at 2 levels.

3.2.1 Zone Wise Wavelet Energy Features

The detailed coefficient $\boldsymbol{D_{j1}}$ (Horizontal Energy Band) is divided into three horizontal and four vertical zones at each level $\boldsymbol{j}$ as shown in Fig. 2. The three horizontal zones namely top zone, middle zone and bottom zone covers 30%, 40% and 30% of the region of band and all vertical zones are divided to have equal size.

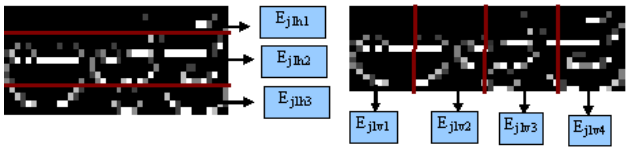


Fig. 2. Three horizontal and four vertical zones of detailed coefficient D_{j1}

Further, $\boldsymbol{3j}$ horizontal and $\boldsymbol{4j}$ vertical zone energy features are obtained, at each level $\boldsymbol{j}$ as described in equations (5) to (8). The method also computes 2 relational features as difference values between top and middle zone, and middle and bottom zone as described in equations (9)-(10), again at each level $\boldsymbol{j}$. Then, the wavelet energy feature is computed from the detailed coefficient $\boldsymbol{D_{j3}}$ (Diagonal Energy Band) at every level $\boldsymbol{j}$ as depicted in equation (11). Hence, totally 10 wavelet energy features are obtained at each level $\boldsymbol{j}$. These features (10 at each level) are stored into feature vector X.

$$E_{j1h1} = \left(\sum_{m=1}^{u1}\sum_{n=1}^{N}(D_{j1}(m,n))\right)/(MxN) \tag{5}$$

$$E_{j1h2} = \left(\sum_{m=u1+1}^{u2}\sum_{n=1}^{N}(D_{j1}(m,n))\right)/(MxN) \tag{6}$$

$$E_{j1h3} = \left(\sum_{m=u2+1}^{M}\sum_{n=1}^{N}(D_{j1}(m,n))\right)/(MxN) \tag{7}$$

$$E_{j1vk} = \left(\sum_{m=1}^{M}\sum_{n=1+(k-1)*step}^{k*step}(D_{j1}(m,n))\right)/(MxN) \tag{8}$$

$$E_{j1h4} = E_{j1h1} - E_{j1h2;} \tag{9}$$

$$E_{j1h5} = E_{j1h2} - E_{j1h3;} \tag{10}$$

$$E_{j3} = (\sum_{m=1}^{M} \sum_{n=1}^{N} (D_{j3}(m,n)) / (MxN) \tag{11}$$

Where,

- E_{j1h1}, E_{j1h2} and E_{j1h3} correspond to energy features of top, middle, and bottom zones of horizontal detailed band D_{j1} at resolution j.
- E_{j1vk} correspond to vertical energy feature of each zone, where k is vertical zone number and varies between **1 to 4**.
- E_{j1h4} and E_{j1h5} represent features that model relation between zone-wise energy features at resolution j.
- j represents resolution level and is varied between **1-2**.
- MxN represents size of detailed band
- (**1:u1, 1:N**) indicates size of top zone
- (***u1+1:u2, 1:N***) indicates size of middle zone
- (***u2+1:u3, 1:N***) indicates size of bottom zone
- ***step*** is the size of each vertical zone and is determined as below;

$$step = N/4;$$

3.2.2 Wavelet Log Mean Deviation Features

Previous work in the field of texture analysis has observed that the logarithmic quantization of wavelet coefficients yields better representation of texture improving overall classification accuracy [16]. Therefore, the effectiveness of wavelet log mean deviation features is evaluated in the current work and an attempt is also made to model relation between detailed energy bands. The method computes totally ***3j*** wavelet log mean deviation features at every resolution level j using equation described in (12), which are stored into feature vector X.

$$LMD_{jp} = \frac{\sum_{m=1}^{M} \sum_{n=1}^{N} \log(\frac{|Djp(m,n)|}{Sj\partial} + 1)}{MN} \tag{12}$$

Where,

∂ is a constant specifying the degree of nonlinearity in the transform.
S_j represents the estimated maximum value of the coefficients at resolution level j.

In the current work, the value of $\partial = 0.001$ is experimented. During experiments it is observed that, the obtained values gives better representation of texture. Further, 2 more additional features that model relation between detailed energy bands are determined as stated in equations (13)-(14). Hence, this step records 10 features (5 at each level) into feature vector X.

$$LMD_{j4} = LMD_{j1} - LMD_{j2;} \tag{13}$$

$$LMD_{j5} = LMD_{j2} - LMD_{j3;} \tag{14}$$

3.2.3 Wavelet Vertical Run Features

A wavelet vertical run $R(\emptyset,d)$ is defined as number of consecutive wavelet coefficients that runs for a distance greater than or equal to a specified value d, in a given

direction *Ø=90 degree (The value 90 is fixed for vertical direction)*. And the wavelet vertical run feature WRF_{j2z} is number of occurrences of wavelet vertical runs in a given area or region and is described in equation (15). These statistical features are obtained from vertical detailed coefficient D_{j2} halved into four equal sized vertical regions/zones (as shown in Fig. 3) leading to a dimension of 8 features at both decomposition levels (4 features for every level ***j***), which are further recorded into feature vector *X*.

$$WRF_{j2z} = \sum_{n=1+(k-1)*zone\_size}^{k*zone\_size} R(Ø, d) \tag{15}$$

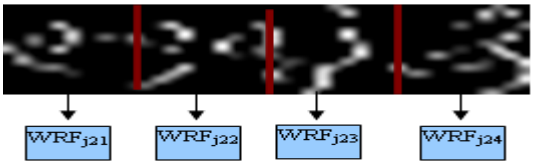


Fig. 3. Four vertical zones of detailed coefficient D_{j2}

Where,

- *R(Ø,d)* is wavelet vertical run for a distance >= *specified value d* (which is empirically chosen as 6), and direction *Ø= 90 degree.*
- *z* indicates zone number of vertical detailed coefficient D_{j2} takes values in the range 1-4.
- *zone_size* is size of each vertical region/zone and is determined as below;

$$zone\_size = = N/4;$$

Finally, the feature vector *X* contains 14 zone wise wavelet energy features, 4 relational features between zones of horizontal detail band, 2 wavelet features of diagonal energy bands, 6 wavelet log mean deviation features, and 4 features that model relation between wavelet log mean deviation features, and 8 wavelet vertical run features leading to a dimension of 38. The feature vector ***X*** at level ***j*** is given below in equation (16);

$$X = [E_{j1h1}\, E_{j1h2}\, E_{j1h3}\, E_{j1h4}\, E_{j1h5}\, E_{j1v1}\, E_{j1v2}\, E_{j1v3}\, E_{j1v4}\, E_{j3}\, LMD_{j1}\, LMD_{j2}\, LMD_{j3}\, LMD_{j4}\, LMD_{j5}\, WRF_{j1}\, WRF_{j2}\, WRF_{j3}\, WRF_{j4},\ j=1,2\,] \tag{16}$$

The data set of such features obtained from training samples is further used for construction of knowledge base for every script.

3.3 Knowledge Bases for Script Identification

For the purpose of knowledge base construction, the images were captured from display boards of government offices in India. The image database consists of 300 Kannada, 300 English, 200 Malyalam, 200 Tamil and 200 Hindi script word images of varying resolutions. The mages in the database are characterized by variable number

of characters, variable font size and style, uneven thickness and spacing between characters, minimal information context, small skew, noise and other degradations.

Then, 70% of the different samples from each script are chosen to train the system. During training, features are extracted from all training samples and knowledge base is organized as a data set of all features. The stored information in the knowledge base sufficiently characterizes all variations in input and script class separation. It is also noticed that, training system with more samples will improve the performance of the system. At the end of training four knowledge bases WD_IMKB_KAN, WD_IMKB_ENG, WD_IMKB_MAL, and WD_IMKB_TAM for Kannada, English, Malyalam and Tamil Scripts are generated. And testing is carried out for all word images of database containing 70% trained and 30% test samples.

3.4 Script Class Identification

The script identification task consists of 2 processing stages. In stage1, the test word image is processed to determine whether it belongs to Hindi Script. Otherwise, stage 2 measures city block distance between test data instance and knowledge base of every script to determine whether it belongs to Kannada, English, Tamil or Malyalam Script. The functionality in both stages is described in the following sections;

3.4.1 Computational Strategy for Hindi Script Identification

In this stage, horizontal run statistics of test word image are used to determine whether the written word in display board image belongs to Hindi or other scripts. Initially, the horizontal runs of length greater than 6 are computed for every row of word image and are stored into a run feature vector HRV. The vector records row number and run length count of all runs for all rows. These run length values are thresholded to classify word image into two classes' w1 and w2. Where, w1 corresponds to Hindi script and w2 corresponds to other scripts category. The classified word image into class w2 is further processed as in stage 2 to determine whether it belongs to Kannada, English, Tamil or Malyalam Script.

3.4.2 City Block Distance Measure

In this stage, test data instance is processed to obtain wavelet features, and a feature vector $\boldsymbol{X_t}$ is constructed as depicted in equation (17).

$$X_t = [tE_{j1h1}\ tE_{j1h2}\ tE_{j1h3}\ tE_{j1h4}\ tE_{j1h5}\ tE_{j1v1}\ tE_{j1v2}\ tE_{j1v3}\ tE_{j1v4}\ tE_{j3}\ tLMD_{j1}\ tLMD_{j2}\ tLMD_{j3}\ tLMD_{j4}\ tLMD_{j5}\ tWRF_{j1}\ tWRF_{j2}\ tWRF_{j3}\ tWRF_{j4},\ j=1,2\ ] \quad (17)$$

Then, the smallest city block distance between test data instance $\boldsymbol{X_t}$ and data set of each knowledge base is determined to obtain distances $\boldsymbol{d_1, d_2, d_3,}$ and $\boldsymbol{d_4}$ as described in equations (18) to (21).

$$d_1 = min\left(\sum_z |\ X_t(\ A_z\ ) - \text{WD\_IMKB\_KAN}(X_i(\ A_z\ ))|\right) \quad (18)$$

$$d_2 = min\left(\sum_z | X_t(A_z) - \text{WD\_IMKB\_ENG}(X_i(A_z))|\right) \quad (19)$$

$$d_3 = min\left(\sum_z | X_t(A_z) - \text{WD\_IMKB\_MAL}(X_i(A_z))|\right) \quad (20)$$

$$d_4 = min\left(\sum_z | X_t(A_z) - \text{WD\_IMKB\_TAL}(X_i(A_z))|\right) \quad (21)$$

Where, A_z represents value of z^{th} attribute of feature variable $\boldsymbol{X_i}$ in the corresponding knowledge base and the value of z lies in the range 1<=z<=38.

The smallest distance between test word image and knowledge base is used to identify the script class. The proposed methodology performs well for variability in font size, style and image resolution. The approach also identifies script of nonlinear text in the image and results are presented in the next section. However, the method requires sufficient training of all variations in font size, style and other degradations.

4 Results and Analysis

The effectiveness of proposed methodology for script identification using wavelet features has been evaluated for 1200 low resolution images of display boards. The images were captured from display boards of government offices in India. The image database consists of 300 Kannada, 300 English, 200 Malyalam, 200 Tamil and 200 Hindi script word images of varying resolutions. The mages are characterized by variable number of characters, variable font size and style, uneven thickness and spacing between characters, minimal information context, small skew, noise and other degradations.

The proposed methodology has produced good results for low resolution word images containing text of different size, font, and alignment with varying background. The approach also identifies script of small skewed text regions. Hence, the proposed method is robust and achieves an identification accuracy of 92% for Kannada Script, 97.67% for English, 82.5% for Malyalam and 87% for Tamil Script. A closer examination of results revealed that misclassifications arise due to minimal information context, noise and larger skew, which affect the texture of region of text and performance of the texture based approach. It is also found that, if the knowledge bases are trained for all variations and degradations, better performance can be obtained.

5 Conclusions and Future Works

In this paper, an approach for word level script identification of low resolution images of display boards employing wavelet features is proposed. The method identifies script of word image without applying techniques for removal of noise and other degradations. This aspect of work makes it more robust and efficient. The proposed set of new texture features are tend to *better* model the texture of a region of text and thus provide sufficient characterization for improving classification accuracy. The testing

of methodology for 1200 low resolution word images containing text of different size, font, and alignment with varying background has yielded an average classification accuracy of 89.7%. The system is found to be resilient to the presence of small skew and degradations. This is a significant result, which makes this work suitable for text understanding and translation systems especially in the Indian context. The method can be extended for script identification of images belonging to other scripts. Only modification needed is construction of new knowledge bases. And further investigations can focus on language identification of word images.

References

1. Abowd, D.G., Atkeson, C.G., Hong, J., Long, S., Kooper, R., Pinkerton, M.: CyberGuide: A mobile context-aware tour guide. Wireless Networks 3(5), 421–433 (1997)
2. Marmasse, N., Schamandt, C.: Location aware information delivery with comMotion. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 157–171 (2000)
3. Tollmar, K., Yeh, T., Darrell, T.: IDeixis - Image-Based Deixis for Finding Location-Based Information. In: Proceedings of Conference on Human Factors in Computing Systems (CHI 2004), pp. 781–782 (2004)
4. Leetch, G., Mangina, E.: A Multi-Agent System to Stream Multimedia to Handheld Devices. In: Proceedings of the Sixth International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 2005 (2005)
5. Premchaiswadi, W.: A mobile Image search for Tourist Information System. In: Proceedings of 9th International Conference on Signal Processing, Computational Geometry and Artificial Vision, pp. 62–67 (2009)
6. Ma, C.-J., Fang, J.-Y.: Location Based Mobile Tour Guide Services Towards Digital Dunhaung. In: International Archives of Phtotgrammtery, Remote Sensing and Spatial Information Sciences, Beijing, vol. XXXVII, Part B4 (2008)
7. Wu, S.-H., Li, M.-X., Yanga, P.-C., Kub, T.: Ubiquitous Wikipedia on Handheld Device for Mobile Learning. In: 6th IEEE International Conference on Wireless, Mobile and Ubiquitous Technologies in Education, pp. 228–230 (2010)
8. Yeh, T., Grauman, K., Tollmar, K.: A picture is worth a thousand keywords: image-based object search on a mobile platform. In: Proceedings of Conference on Human Factors in Computing Systems, pp. 2025–2028 (2005)
9. Fan, X., Xie, X., Li, Z., Li, M., Ma: Photo-to-search: using multimodal queries to search web from mobile phones. In: Proceedings of 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (2005)
10. Hwee, L.J., Chevallet, J.P., Merah, S.N.: SnapToTell: Ubiquitous information access from camera. In: Mobile Human Computer Interaction with Mobile Devices and Services, Glasgow, Scotland (2005)
11. Spitz, A.L.: Determination of Script and Language Content of Document Images. IEEE Trans. Pattern Analysis and Machine Intelligence 19(3), 235–245 (1997)
12. Pal, U., Chaudhury, B.B.: Identification of Different Script Lines from Multi-Script Documents. Image and Vision Computing 20(13-14), 945–954 (2002)
13. Shijian, L., Tan, C.L.: Script and Language Identification in Noisy and Degraded Document Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(1) (January 2008)
14. Tan, T.N.: Rotation Invariant Texture Features and Their Use in Automatic Script Identification. IEEE Trans. Pattern Analysis and Machine Intelligence 20(7), 751–756 (1998)

15. Peake, G.S., Tan, T.N.: Script and Language Identification from Document Images. In: Proc. Eight British Mach. Vision Conf., vol. 2, pp. 230–233 (September 1997)
16. Busch, A., Boles, W.W., Sridharan, S.: Texture for Script Identification. IEEE Trans. Pattern Analysis and Machine Intelligence 27(11), 1720–1732 (2005)
17. Hiremath, P.S., et al.: Script identification in a handwritten document image using texture features. In: IEEE 2nd International Advance Computing Conference, pp. 110–114 (2010)
18. Hochberg, J., Kerns, L., Kelly, P., Thomas, T.: Automatic Script Identification from Images Using Cluster-Based Templates. IEEE Trans. Pattern Analysis and Machine Intelligence 19(2), 176–181 (1997)
19. Vikram, T.N., Chidananada Gowda, K., Shalini, R.: "Symbolic representation of Kannada characters for recognition. In: IEEE Conference on.., pp. 823–826
20. Angadi, S.A.: Intelligent Integrated Automation System for Efficient processing of Postal Mail, PhD thesis submitted to Department of Studies in Computer Science. University of Mysore (2007)
21. Li, L., Tan, C.L.: Script identification of camera-based images. In: 19th International Conference on Pattern Recognition, ICPR 2008, December 8-11, pp. 1–4 (2008), doi:10.1109/ICPR.2008.4760965

Analytical Study Using Data Mining for Periodical Medical Examination of Employees

Kiran Waghmare and Anusha R. Pai

Department of Computer Engineering, Padre Conceicao College of Engineering,
Verna Goa 403722
anusha.pai@gmail.com

Abstract. In the recent years, intelligent systems proved to be a useful and successful tool which is often been used for decision support, data mining and knowledge discovery in medicine. It is of great practical significance to conduct a scientific, accurate, and complete analysis, assessment or even early warning on the safety of current and future medicines. In this paper, we present the results of intelligent data analysis used for determining trends in lifestyle related diseases like Diabetes, Hypertension, Cardio-vascular diseases, Backache, Dyslipidaemia, Hearing impairment and other disease category. The study included Periodical Medical Examination (PME) data on 300 employees of public sector enterprise dealing with oil exploration in India.

Keywords: Intelligent system, Data mining, Decision tree, PME analysis.

1 Introduction

Medical artificial intelligence is primarily concerned with the construction of Artificial Intelligence (AI) programs that perform diagnosis and make therapeutic recommendations. Unlike medical applications based on other programming methods, such as purely statistical and probabilistic methods, medical AI programs are based on symbolic models of disease entities and their relationship to patient factors and clinical manifestations [8]. For example, knowledge-based systems are the most common support systems for diagnosis in medical field. It is also known as an expert system which contain clinical knowledge, usually about a very specifically defined task, and able to produce reasoned conclusion from patient's data.

Occupational Health is primarily concerned with promotion, prevention and protection of health of workers against any hazards arising at the work place and includes preventive health check-ups like Periodical Medical Examination (PME) and mitigation of any occupational diseases arising in occupational environment. Preventive health check-ups like PME not only reduce sickness absenteeism but also increase productivity of employees. The study is undertaken at the occupational health center of the enterprise and includes PME dataset of 300 employees (from 2009-2011) working at various offshore production platforms and rigs in India.

The use of information technology in Occupational health domain is increasing as it improves quality, accessibility and confidentiality of medical care for doctors as well as employees. Moreover, database of PME can be stored and retrieved for future

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 221–227.
springerlink.com

references (as employees are transferred/shifted from one workplace to another). The PME database involves complete medical profile of individual employees like physical examination, blood tests, X-rays, Ultrasonography etc. This test covers all basic parameters to exclude any pathological deviation from normal. The individual is examined by in-house doctor and is appraised about relevant findings and advised medications, if needed and lifestyle changes like balanced nutrition and exercise. Thus, relevant models are constructed depending on various parameters using data mining techniques. The database contains information such as Login user, Employee Information (Personal, Family, Employment, Any disorder/ailment etc.), Physical Examination information (general health, disease category and diagnosis information, Any surgery/operation etc), Investigation information (Blood test, X-rays, USG etc.), Report generation, Analysis.

2 Related Work

Data mining approaches are used for classification and prediction. These two main objectives description and prediction can be easily distinguished in the data mining process when integrated into the medical system. c4.5 classification algorithm is used to predict survival of burn patients [6]. The machine learning algorithm c4.5 was used to classify the patients using WEKA tool. The performance of the algorithm was examined by using the classification accuracy, sensitivity, specificity and confusion matrix. In ref. [2] data mining and knowledge discovery have been applied to hospital management, and a modified data mining Method has been proposed that would be appropriate for mass data in Hospital Information System. The nature of anxiety disorders and an approach toward assisting the personalized treatment of patients suffering from anxiety disorders was proposed in ref. [5]. In ref. [10] the authors have proposed an efficient scalable decision tree construction algorithm with less processing time and most suitable for large datasets. Using Decision-tree theory and related analysis a fairly scientific and rational data mining structure has been worked out which was then built and pruned to get the classification rules and then comprehensive assessment, ultimately providing GSP decision-makers with more effective management decision-making basis [11]. A framework for dynamic evidence based medicine using data mining is proposed which helps in automatically analyzing huge clinical databases and discover pattern behind them [4]. Use of data mining methods for predicting the evolution of patients in an ICU (Intensive Care Unit) is presented in ref. [7].

Thus, the present study was undertaken to analyse dataset of Periodical Medical Examination of employees by using Data Mining technologies. The purpose of the study is to visualize particular ailment/disease in employees at a glance and analyze disease trends in employees working in office/field with its correlated hazards arising in occupational environment.

3 Data Mining

Data mining is an interface between statistics, computer science, artificial intelligence, machine learning, database management and data visualization. The importance of employees health and increasing trend of life-style diseases due to work related stress require not only traditional manual data analysis but also efficient computer assisted analysis for speedy access to information at all levels. In this study, Data analyzing software is developed for OHC, with the objective of viewing employee's health profile at a glance and also analyzing the data set as and when required. It will also help in identifying disease trends in employees working for longer periods in particular vocation.

The term Knowledge Discovery in Databases or KDD refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. KDD refers to the overall process of discovering useful knowledge from data [3, 9]. In this software, we are analyzing various lifestyle related diseases like diabetes, hypertension etc. Once the disease trends are identified, preventive measures can be taken to mitigate these common diseases by adopting healthy practices and health education. Thus ,medical expenditure can considerably reduce in long term by undertaking Health Education Programs like Hearing conversation program, provision of Gymnasium, Sports activities etc. Benefits of Health promotion activities can be enumerated as sickness absenteeism is reduced, accident & injuries are reduced, healthy work force will lead to increase productivity and improves morale of employees.

4 Design of Mathematical Modeling System

In this work, a data mining framework whose steps are discussed below, is proposed to visualize the particular ailment or disease in employee at a glance and to analyze the disease trend and correlate to hazard arising in occupational environment.

Problem definition: The identification of early signs of a disease, its proper diagnosis and treatment has become a serious health concern in all industries.

Data set selection: The data set for the current study includes 303 employees' health examination data including physical examination and investigation reports.

Characteristics of Clinical Metadata: PME record consists of different parameters like personal history, physical examination, investigation, diagnosis, treatment and follow-up. Thus, this PME data has different characteristics as enumerated below:

Polymorphism: Different clinical examination methods and measurements technique lead to different data sources variation such as text, pure data, signal (ECG), image (MRI) and so on. It is most obvious that clinical data differ from one another leading to variation and increase difficulty in data processing and mining.

Integrity: As clinical medicine data exist in databases certain biases are present like uncertainty of description, blur records, sampling inadequacy and inaccuracy of measuring equipments which can make data incomplete.

Redundancy: The clinical medicine database is huge one and is multi-dimensional, nonlinear and incessant which may lead to redundancy.

Sequential: Medical activity is a dynamic process related to time and also emergence of new disease trends cannot be predicted due to introduction of new technologies and new hazards at workplace.

Privacy: Such data inevitably involves patient privacy, so users should do data mining with confidentiality and provide adequate safety precautions so that data cannot be assessed by unauthorized person.

Data Collection: Periodic Medical Examination (PME) is health check-up of employees done at periodic intervals. In ONGC, periodicity of PME is 5yrs upto 45 yrs, 3 yrs from 45 to 55 years and every 2 years above 55 years till superannuation. The dataset of total 303 employees from 2009 to 2011 is used in this study. The individual proforma consists of detailed medical history of employees including personal history, job rotation, personal habits, accident/operation history, physical examination, investigation, diagnosis, recommendation and follow-up.

Cleaning and preprocessing: After deleting abnormal and incomplete data, leaving a final dataset of 303 pieces used in data analysis. The analyzed dataset contains 125 attributes.

Data Preprocessing: Preprocessing method of medical data includes cleaning, integration, transformation, reduction and discretization. Normalization is the most important step in data preprocessing. Traditional normalization methods mainly used are Maximum-minimum Method and Z-score method. While normalizing, the characteristic would not change. Moreover, whole data's topology structure is quite important to further application of manifold learning algorithms

Data analysis: Adopting data mining algorithm, ID3 used to generate a decision tree developed to investigate the implicit meaningful rules from the health examination data [1].

Classification using Decision Tree: Classification is one of the primary data mining tasks used to classify data into predefined classes, described by a set of attributes. The sample data file format is shown in table 1. The most apparent application of classification in clinical care is the process of diagnosis of diseases. The decision tree method is increasing in popularity for both classification and prediction. Once data is pre-processed, training and testing phase are performed. Next, classification using Decision Trees is implemented. These trees can be easily transformed into classification rules that can be conveyed in common language. The results obtained would be compared and discussed. Decision Tree is capable to speed up the classification and deals better with discrete or categorical features.

Table 1. Sample Data File Format

Sr. no.	CPF	pme DATE	Name	Age	Gender	DM	HT	CVS	HI	BK	DL	Others
1	78510	04-03-09	THAKUR D	47	M	NO	YES	NO	NO	NO	NO	NO
2	47892	02-04-09	P N SHREE	50	M	NO	YES	NO	NO	NO	NO	NO
3	45080	27-01-09	RAMESH F	51	M	NO	YES	NO	NO	YES	YES	NO
4	66360	11-02-07	JAYANTA	47	M	NO	YES	NO	NO	YES	YES	NO
5	58368	24-02-09	D N RAI	52	F	YES	NO	NO	NO	NO	YES	NO
6	57805	20-02-09	FRANCIS A	52	M	NO	NO	NO	NO	NO	YES	NO
7	56048	02-03-09	KISHOR D/	51	M	NO	YES	NO	NO	NO	YES	NO
8	56647	17-02-09	I H.MIRKA	51	M	NO	NO	NO	NO	YES	NO	YES
9	52418	02-03-09	HARISH KU	52	M	YES	NO	NO	NO	NO	YES	NO
10	44666	02-08-09	PRATAP C	51	M	NO	NO	NO	NO	NO	NO	YES
11	56056	05-02-09	C R AHME	47	F	NO	NO	NO	NO	NO	YES	NO
12	56061	27-02-09	RAJENDRA	49	M	NO	NO	NO	NO	YES	NO	NO
13	39165	05-01-09	PARADKA	53	M	NO	YES	NO	NO	NO	NO	NO
14	57908	17-02-09	T K ROY	52	M	YES	NO	NO	NO	NO	NO	NO
...	...	...	...	...	...	...	...	...	...	...	...	...

Implementation mechanism of Decision Tree algorithm using ID3: The decision tree algorithm used herein is a hybrid method mainly including two steps, i.e. building and pruning. In the decision tree method, information gain approach is generally used to determine suitable property for each node of a generated decision tree. Thus, we can select the attribute with the highest information gain (entropy reduction in the level of maximum) as the test attribute of current node. In this way, the information needed to classify the training sample subset obtained from later on partitioning will be the smallest. Therefore, the use of such an information theory approach will effectively reduce the required dividing number of object classification.

Set S is set including s number of data samples whose type attribute can take m potential different values corresponding to m different types of c_i, ,i (1,2,3, ..., m). Assume that s_i is the sample number of C_i . Then, the required amount of information to classify a given data is

$$I(s_1, s_2, ..., s_m) = -\sum_{i=1}^{m} P_i \log(P_i) \tag{1}$$

where $P_i = S_{ij} / |S_j|$ is the probability that any subset of data samples belonging to categories C_i. Suppose that A is a property which has v different values $\{a_1, a_2, ..., a_v\}$ .Using the property of A, S can be divided into v number of subsets $\{S_1, S_2, ...S_V\}$, in which S_j contains data samples whose attribute A are equal a_j in S set. If property A is selected as the property for test, that is, used to make partitions for current sample set, suppose that S_{ij} is a sample set of type C_i in subset s_i , the required information entropy is

$$E(A) = \sum_{j}^{v} \frac{S_{1j} + S_{2j} + ...S_{mj}}{s} I(S_{1j}, ..., S_{mj}) \tag{2}$$

Such use of property A on the current branch node corresponding set partitioning samples obtained information gain is:

$$Gain(A) = I(s_1, s_2, ..., s_m) - E(A) \tag{3}$$

ID3 algorithm traverses possible decision-making space using top-down greedy search strategy, and never trace back and reconsider previous selections.

Table 2. PME data table for 2009-2011

Year	No of employees (n=303)	Percentage
2009	101	33.50%
2010	119	39.20%
2011	83	27.30%
Total	**303**	**100%**

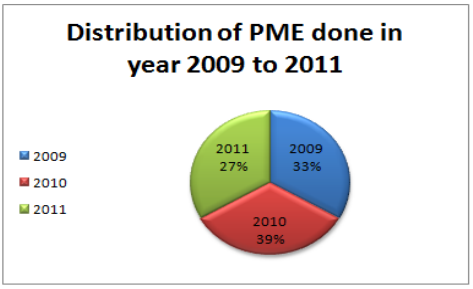


Fig. 1. Pie Diagram for PME data

In the study group, PME data from 2009-2011 till date is analyzed. All employees due for PME have undergone complete PME check-up as per company policy at empanelled hospitals in Mumbai. PMEs done in respective years is tabulated in table 2 and its pie diagram is shown in figure 1.

Data regarding disease profile is tabulated in table 3 and accordingly its pie diagram is shown in figure 2. It was observed that prevalence of Diabetes (23.4%), Hypertension(23.1%), Dyslipidemia (21.4%), Cardiovascular Disorders (13.2%) which are all life style related diseases are same as prevalence in general population in urban India. This can be attributed to aging working population, high work demand, psychological stress, sedentary life style, lack of physical exercise and dietary habits, addiction to smoking, alcohol, backache(15.1%) and hearing impairment (8.9%) can be attributed to employees previous job profile at different work centers.

Evaluation and application: Employ the extracted classification If-Then rules from data set and provide useable knowledge to support for the PME diagnosis and treatment.

Table 3. Employee Disease Profile

Major Disease	No of PME (n=303)	Percentage
Diebetes	71	23.40%
Hypertension	70	23.10%
CardioVascular Disorder	40	13.20%
Hearing Impairment	27	8.90%
Backache	46	15.10%
Dyslipidimia	65	21.40%
Others	86	28.30%

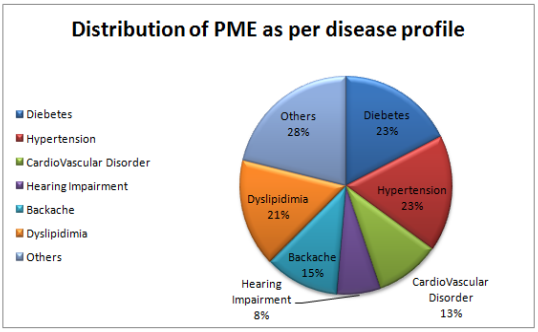


Fig. 2. Pie Diagram for Disease profile

5 Conclusion

Through the construction of decision tree, classification rules for predicting lifestyle related diseases are explored from the health examination data set. This study provides new scientific and quantitative information. The prevalence of Diabetes, Hypertension, Cardiovascular disorders were observed comparatively more in employees, which are life style related diseases which corresponds to prevalence in general population. Doctors would not only benefit from better understanding of cause-effect relationship between work related hazards and diseases but can also take preventive action to reduce their incidence in employees. Moreover, there are tangible benefits to company like reduced medical expenditure, reduced sickness absenteeism, thereby increasing productivity in long run and hidden intangible benefits for employees by improving their quality of life.

References

1. Chae, Y.M., Kim, H.S., Tark, K.C., Park, H.J., Ho, S.H.: Analysis of healthcare quality indicator using data mining and decision support system. Expert Systems with Applications 24(2), 167–172 (2003)
2. Hai-Yan, Y., Jing-Song, L., Xiong, H., Zhen, H., Zhou, T., Jie, C., Tao, Z.: Data Mining Analysis of Inpatient Fees in Hospital Information System. In: Healthcare Informatics Engineering Research, pp. 82–85 (2009)
3. Kantardzic, M.: Data mining concepts, models, methods and algorithms. Wiley, NY (2002)
4. Masuda, G., Sakamoto, N., Yamamoto, R.: A Framework for Dynamic Evidence Based Medicine using Data Mining. In: Proc. of IEEE Symposium on Computer-Based Medical Systems, CBMS 2002, pp. 117–122 (2002)
5. Panagiotakopoulos, T.C., Lyras, D.P., Livaditis, M., Sgarbas, K.N., Anastassopoulos, G.C., Lymberopoulos, D.K.: A Contextual Data Mining Approach Toward Assisting the Treatment of Anxiety Disorders. IEEE Transactions on Information Technology and Biomedicines 14(3), 567–581 (2010)
6. Patil, B.M., Toshniwal, D., Joshi, R.C.: Predicting Burn Patient Survivability Using Decision Tree In WEKA Environment. In: IEEE International Advance Computing Conference, IACC 2009, India, pp. 1353–1356 (2009)
7. Ramon, J., Fierens, F., Guiza, F., Meyfroidt, G., Blockeel, H., Bruynooghe, M., Berghe, G.V.D.B.: Mining Data from Intensive Care Patients. J. Advanced Engineering Informatics 21(3), 243–256 (2007)
8. Shortliffe, E.H.: Computer Programs to Support Clinical Decision Making. Journal of American Medical Association 258, 61–66 (1987)
9. Soukup, T., Davidson, I.: Visual data mining: Techniques and tools for data visualization and mining. Wiley, NY (2002)
10. Thangaparvathi, B., Anandhavalli, D.: An Improved Algorithm of Decision Tree for Classifying Large Data Set Based on RainForest Framework. In: Proc. of Intl. Conf. on Communication Control and Computing Technologies, India, pp. 800–805 (2010)
11. Ye, L., Lu, R., Shao, Q., Dong, R.: Application of Decision-tree Algorithm to GSP Analysis and Assessment of Drug Safety Situation. In: Proc. of Second International Symposium on Knowledge Acquisition and Modeling, China, pp. 103–106 (2009)

Syntactic Representation of Shape of Object Using Regular Grammar

Saket Jalan[1], Pinaki Roy Chowdhury[2], and K.K. Shukla[1]

[1] Department of Computer Engineering, IT-BHU, India
{saket.jalan.cse07,kkshukla.cse}@itbhu.ac.in
[2] Defence Terrain Research Laboratory (DRDO) India
rcpinaki@yahoo.com

Abstract. This paper presents a scheme to represent shape of an object using formal rules describing regular grammar. We use bicubic Bezier curves as the geometric primitives of the boundary of object. Production rules describing regular grammar are then extracted based upon the relation among the Bezier curves obtained. Rules thus obtained represent the knowledge of shape of object.

1 Introduction

Image understanding systems describe the features of object in image as well as it can be used to recognize objects in image. Representation and recognition of objects using its features has attracted the researchers for past several years. One such feature is the shape of object which is determined by its boundary. To represent the shape of objects we need to represent its boundary using primitives. There are various ways to represent the boundary of object using primitives like Chain Codes, Polylines, B-Splines, Fourier Descriptors, etc. [1], [12].

Syntactic approach to pattern recognition use primitives and their relation in the form of production rules to represent complex patterns [3]. Production rules can unambiguously describe how each of the primitive is related to each other.

In the current work, we use syntactic approach to represent the shape of object. We first extract the boundary coordinates of the object and represent it using bicubic Bezier curves. This helps us in describing the complete shape using smaller set of boundary primitives. Then using the relation among the boundary primitives (Bezier curves) obtained, we generate the production rules of the boundary of object.

Syntactic approach to pattern recognition, in past, has used object representation using primitives to form grammar of object [3]. Since then applicability of this approach has been improved by researchers. Feder described the method to extract context-free rules from the object primitives [2]. Syntactic representation of 3-D objects has also been studied where the primitives are the surfaces of object [7]. Syntactic pattern recognition uses the rules to construct syntax trees which are then compared to recognize object. Also, automata matching are performed to recognize objects. Recent research has focused on representing the object or even the complete scene using its

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 229–236.
springerlink.com

various parts to form parse trees [5], [11] notwithstanding the fact that the applicability of this approach has been limited due to the difficulty in extraction of the primitives.

In contrast to the work done so far, we use bicubic Bezier curves to represent the object which has better capability in describing even the complex shapes with less number of primitives. We use Plex Grammar to extract the grammatical rules, and the rules obtained are regular.

In further sections, we describe the rule extraction process. We first give a small overview of Plex grammar in Section 2. Then, we discuss the boundary detection approaches in Section 3. In Section 4, we describe the boundary using Bezier curves. In Section 5, we describe the algorithm to extract the production rules describing a regular grammar of object. Finally, we discuss the results and present our conclusions in Section 6.

2 Plex Grammar

In this Section we give a short overview of Plex grammar given by Feder which is used to obtain context-free rules of the object [2], [3].The context-free rules are based on the relation of its primitives among themselves i.e., the way in which they connect to each other. Further, we show the rules obtained using an example.

A primitive here can be any geometric primitive like a line, a quadratic curve, a Bezier curve etc. An attaching point of a primitive is a point on the primitive by which it attaches itself to some other primitive. There can be multiple attaching points to an entity. An entity with N attaching points is called a NAPE (N Attaching Point Entity) and it can attach itself to other primitives using N points. We number each of these N points from 1, 2 .., N. 0 is the null identifier which is used when there are no points which are taking part for a particular NAPE. The context-free rules are then given by (1).

$$A\Delta_A \rightarrow X\Gamma_x\Delta_x, \quad X \neq NULL,\ X\Gamma_x = connected \tag{1}$$

According to this rule, A appearing in some context can be replaced by $X\Gamma_x$. Following these rules for the complete object we can obtain all the context-free rules of the object. X is known as the definition control list and it's a string of the form $a_1, a_2 \ldots a_n$. This list contains NAPE's which can be a single primitive or could consist of group of primitives attaching to each other. Γ_x is known as the joint list. It describes how at each of the joints NAPEs attach each other. They are divided into field with each field describing about a joint. Field contain for each of the NAPEs values as q_{ij} which indicates that attaching point j of the i^{th} component takes part at the joint specified by the current field.

Δ_A and Δ_x are called the definitum and definition tie point list. They describe how the points of A are related to the points of NAPEs in X. On both the sides lists are ordered and in case a particular component does not take part at a point, we denote it by the null identifier 0.

Let us consider an example of letter A as shown in Figure 1. Non-terminal NAPEs for the letter A are <SIDE> and <A> while let us say <T> is the terminal NAPE. The context-free rules for this object are then given as:

1. <A>{} → <SIDE><SIDE><T>{201;022}{}
2. <SIDE>{1;2;3} → <T><T>{21}{10;21;02}

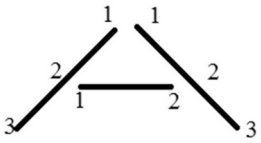


Fig. 1. Letter A showing its primitives

Rule 1 above states that letter A consists of two non-terminals <SIDE> and a terminal in between. Joint list describes that attaching point 2 of first attaches with attaching point 1 of terminal while second side does not take part at this joint. Also, at the second joint, second non-terminal attaches with the terminal.

Similarly, in rule 2 above, its states that sides can be broken down into two terminals where attaching point 2 of first and attaching point 1 of second make a joint. Tie lists here state that point 1 of side is point 1 of first terminal. Similarly, point 2 of side is point 2 of first and point 1 of second terminals respectively. And point 3 of side is point 2 of second terminal.

3 Boundary Detection

Shape of an object is given by its boundary. The boundary in turn is described by the coordinates. To be able to represent shape of object we need to determine the boundary of object. In this section we discuss the boundary detection.

Boundary is determined by the coordinates of edge of object. So, we need to determine the edge of object first. Edge detection algorithms work on 2D image. So, we first convert the image to 2D by converting it to grayscale image. For a pixel with RGB components, equation (2) gives equivalent grayscale pixel [8].

$$gray = R * 0.2989 + G * 0.5870 + B * 0.1140 \tag{2}$$

We apply the equation 2 for all the pixels and obtain the corresponding grayscale image. Many edge detection algorithms exist that can be used to extract the edges of the object like Canny, Sobel etc [4]. Most of these algorithms are based on first finding the derivative using gradients and then comparing with a pre-defined threshold. For example, (3) depicts Sobel masks used in Sobel edge detection algorithm.

$$\begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array} \tag{3}$$

The edge detection algorithm which needs to be applied depends upon the kind of image. Applying edge detection gives us boundary along with internal edges and noise. We need to refine our image to remove these and obtain the boundary. We apply the boundary tracing algorithm which traverses the complete boundary of object and give the coordinates from the image.

Boundary detection procedure described here is very basic; however, it serves our purpose of putting across the idea. In fact we may need a lot of operation while removing the noise and internal edges from the image. Also, there may be other interventions like shadow. A lot of boundary detection techniques exist in literature, many of which are described in [4]. Some of the operations that may be required are dilation of edges, flood fill etc. For example, for the cup shown in Figure 2-(a), Figure2-(b) shows the result of edge detection. We then fill all the holes using flood fill and apply Sobel edge detection again to obtain the boundary of cup as shown in Figure2-(c). Finally we then apply the boundary tracing algorithm to extract all the coordinates of the boundary in order.

(a) (b) (c)

Fig. 2. (a) A sample cup. (b) Result after applying Sobel edge detection on 2-(a). (c) Extracted boundary of the object in 2-(a) after de-noising 2-(b).

4 Representation Using Bezier Curves

In the previous Section, we have discussed a method to obtain the boundary coordinates of the object, whose number for an object is also quite high. Because of this it becomes impractical to describe the object using its boundary coordinates. Therefore, one such representation technique available in the literature is for representing the object with comparatively less number of primitives. Some of the examples of such primitives are Fourier descriptors, Polylines, B-Splines etc., [12]. Any of these, never represent the exact boundary, but they are a close approximation of the original. This indeed, satisfies our objective.

Bezier curves are a powerful means to represent even the complex curves. The degree of Bezier curves can vary and its effectiveness increases with increasing degree. In the present work we use bicubic Bezier curves to represent our boundary. Later, we

will see that using them reduces the number of values we need to handle. Equation (4) gives the equation of a bicubic Bezier curve.

$$B(t) = (1-t)^3 P_0 + 3(1-t)^2 tP_1 + 3(1-t)t^2 P_2 + t^3 P_3, \; t \in [0,1] \tag{4}$$

In the equation described by (4), P_0, P_1, P_2, and P_3 are called the control points. They alone can represent the complete curve. The intermediate points are determined by varying the value of t from 0 to 1. P_0 is the starting point and P_3 is the ending point for each curve. To be able to represent our boundary using Bezier curves, we determine the curves and their control points.

To find these curves, a fitting algorithm is used which determines the curve minimizing the error between the actual boundary and the obtained curves. We use least square fitting algorithm to obtain our curves [6]. Algorithm takes the boundary coordinates and a value called the maximum permissible error as input. Maximum permissible error is the maximum total error which is allowed for each curve. There are various ways for calculating this value [10].

Least square fitting Algorithm initially takes all the coordinates with P_0 as first and P_3 as last point. For each P_0 and P_3, we first find P_1 and P_2 by interpolating the curve, and then determine equal number of intermediate points as the number of points in the original object between P_0 and P_3. We now calculate the least square error between them and compare with the maximum permissible error. For P_1, P_2 … P_n as the original points and Q_1, Q_2 …, Q_n as the generated points, (5) gives the least square error.

$$least\ square\ error = \sum_{i=1}^{n} \left| P_i - Q_i \right|^2 \tag{5}$$

If the error is less than the limit, we take the next set of P_0 and P_3 and repeat, otherwise we break the curve at the point where error is maximum keeping it as P_3 and repeat the process. Finally, we obtain all the Bezier curves. For example, for the object in Figure 2-(a), Figure 3 shows the obtained Bezier curves. Each of the starting and end points are shown with crosses on the object.

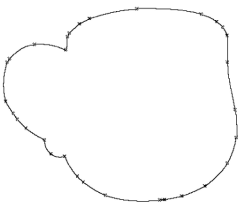


Fig. 3. Bezier curves for the object in Fig 2-(a). Crosses on the curves are the starting and ending points.

In the Figure 3, we obtain 31 curves. The control points for the first Bezier curve that we obtained are:

1. $P_0 = 45.00 + 242.00i$
2. $P_1 = 44.66 + 278.00i$

3. $P_2 = 44.55 + 314.00i$
4. $P_3 = 54.00 + 350.00i$

Thus, we see that the bicubic Bezier curves can effectively represent the complete boundary using less number of primitives. We also see from the Figure 3 and the algorithm, that break points occur at points where curve is not smooth. Thus, object with smooth shape will have lesser number of curves compared to object with rough shape or more corners. Least Square Fitting is a basic algorithm, and better algorithms exist in literature [9].

5 Extraction of Production Rules Describing Regular Grammar

In the previous Section we obtained bicubic Bezier curves of the boundary of object. We also know how these curves attach to each other in order to make the object. Having done this, we exploit this information and utilize it in creating production rules of the object. In this Section, we first extract the formal rules, which are context-free by using the Plex grammar described in Section 2. Then, we convert these rules to another form by which we can describe the object using regular grammar.

Each of the boundary curves has two attaching points and we number them as 1 and 2. 0 is the null identifier as described in Section 2. We can also take a sequence of curve which is a subset of original boundary. This also has two attaching points numbered as 1 and 2. We identify a curve as C_{ij}, while a sequence of curves as S_{ij}. C_{ij} is the curve between points i and j while S_{ij} is the sequence of curves starting from point i to point j. We then apply equation (1) described in Section 2 to obtain the context-free rules.

We can form the rules in number of ways for a particular object depending upon how we take the syntax tree to be. For simplicity, let us suppose another object as shown in Figure 4 whose Bezier curves are marked. This particular object has 9 Bezier curves and they are marked on the Figure.

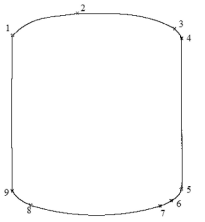
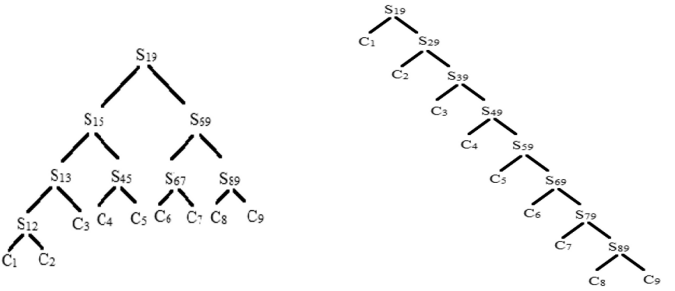


Fig. 4. A cylinder showing its 9 bicubic Bezier curves

One method of forming the syntax tree is to form it like a complete binary tree using the identifications we gave before for curves and sequence of curves. The corresponding syntax tree is shown in Figure 5-(a). The context-free rules in this case are given by:

1. <Object>{} → <S_{15}><S_{69}>{21;12}{}
2. <S_{15}>{1;2} → <S_{13}><S_{45}>{21}{10;02}

3. $<S_{13}>\{1;2\} \rightarrow <S_{12}><C_3>\{21\}\{10;02\}$
4. $<S_{12}>\{1;2\} \rightarrow <C_1><C_2>\{21\}\{10;02\}$
5. $<S_{45}>\{1;2\} \rightarrow <C_4><C_5>\{21\}\{10;02\}$
6. $<S_{69}>\{1;2\} \rightarrow <S_{67}><S_{89}>\{21\}\{10;02\}$
7. $<S_{67}>\{1;2\} \rightarrow <C_6><C_7>\{21\}\{10;02\}$
8. $<S_{89}>\{1;2\} \rightarrow <C_8><C_9>\{21\}\{10;02\}$



(a) (b)

Fig. 5. Syntax trees for object in Fig 4. (a) A complete tree (b) Tree such that each level as at least one curve.

Another way of making our syntax tree is to go on extracting a single curve each time till the end. This way we always have at least one curve at each level and one non-terminal except at the last level which has two terminals. The corresponding syntax tree is shown in Fig 5-(b).

The context-free rules in this case are given as:

1. <Object>{} → <C1><S29>{21;12}{}
2. <S29>{1;2} → <C2><S39>{21}{10;02}
3. <S39>{1;2} → <C3><S49>{21}{10;02}
4. <S49>{1;2} → <C4><S59>{21}{10;02}
5. <S59>{1;2} → <C5><S69>{21}{10;02}
6. <S69>{1;2} → <C6><S79>{21}{10;02}
7. <S79>{1;2} → <C6><S89>{21}{10;02}
8. <S89>{1;2} → <C8><C9>{21}{10;02}

We see that having only the boundary while applying Plex grammar is an advantage as we can form the tree as in Fig 5-(b). Each of the production rules obtained in this case has one terminal and one non-terminal except in rule 8 which has two terminals. We now modify the rule 8 and add an extra rule as below:

8. $<S_{89}>\{1;2\} \rightarrow <C_8><S_{99}>\{21\}\{10;02\}$
9. $<S_{99}>\{1;2\} \rightarrow <C_9>\{\}\{1;2\}$

We thus have the production rules capable of describing a regular grammar of object. Similarly, we can form the production rules for the cup in Fig 2-(a) or any object following the procedure.

6 Discussion and Conclusion

We have described in previous Sections that it is possible to represent any object using the formal rules which are context-free. We also described how we can form the production rules which can describe the regular grammar of object. They can be used for generative representation of an object. Bezier curves are powerful means for representing curves and can represent even the complex curves using just the four control points. Thus comparing to handling the boundary with coordinates, Bezier curves present us a way to practically describe the complete boundary. Number of Bezier curves obtained depends upon the smoothness of surface. Thus for object with smooth shape, we have less number of curves.

Syntactic approach to pattern recognition has used production rules, syntax trees and automata to compare objects. The rules we obtained in Section 5 can be used similarly in such applications. We obtained context-free rules of the boundary by applying the Plex grammar. We also obtained in Section 5 production rules that can represent regular grammar, using which we can form a finite state automata representing the object. The number of rules obtained depends upon the number of curves.

References

1. Ballard, D.H., Brown, C.M.: Computer Vision. Prentice-Hall, Englewood Cliffs (1982)
2. Feder, J.: Plex Languages. Information Sciences 3, 225–241 (1971), doi:10.1016/S0020-0255(71)80008-7
3. Fu, K.S.: Syntactic Pattern Recognition and Applications. Prentice-Hall, Englewood Cliffs (1982)
4. Gonzalez Rafael, C., Woods Richard, E.: Digital Image Processing. Prentice-Hall (2002)
5. Feng, H., Chun, Z.S.: Bottom-Up/Top-Down Image Parsing with Attribute Grammar. IEEE Trans. Pattern Anal. Mach. Intell. 31(1), 59–73 (2009), doi:10.1109/TPAMI.2008.65
6. Khan, M.: Approximation of data using cubic Bezier curve least square fitting (2007), http://130.235.212.213/trac/raw-attachment/wiki/BezierPaths/cubicbezierleastsquarefit.pdf (accessed April 30, 2012)
7. Lin, W.C., Fu, K.S.: A Syntactic Approach to 3-D Object Representation. IEEE Trans. Pattern Anal. Mach. Intell. 6(3), 351–364 (1984), doi:10.1109/TPAMI.1984.4767528
8. Pratt William, K.: Digital Image Processing, 4th edn. John Wiley & Sons Inc. (2007)
9. Lejun, S., Hao, Z.: Curve Fitting With Bezier Cubics. CVGIP: Graphical Model and Image Processing 58(3), 223–232 (1996), doi:10.1006/gmip.1996.0019
10. Sun, Z., Wang, W., Zhang, L., Liu, J.: Sketch Parameterization Using Curve Approximation. In: Liu, W., Lladós, J. (eds.) GREC 2005. LNCS, vol. 3926, pp. 334–345. Springer, Heidelberg (2006)
11. Yao, B., Yang, X., Wu, T.: Image parsing with stochastic grammar: The Lotus Hill dataset and inference scheme. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2009, p. 8 (2009), doi:10.1109/CVPR.2009.5204331
12. Zhang, D., Lu, G.: Review of shape representation and description techniques. Pattern Recognition 37(1), 1–19 (2004), doi:10.1016/j.patcog.2003.07.008

Message Overhead Analysis of Quorum Protocols

Parul Pandey and Maheshwari Tripathi

Institute of Engineering and Tecnology, Lucknow, India
parul_pandey123@rediffmail.com, mt@ietlucknow.edu

Abstract. Data replication is a vital component in the design of any parallel information system. Specifically, the widespread use of clustered architecture often requires replication of data for performance and availability reasons. A problem that must be solved when using replication is how to maintain the copies in a consistent state when multiple accesses are being made. There must exist a control protocol responsible for synchronizing the accesses so that the logical data is consistent. To address this limitation, Quorums are often suggested as a way to reduce the overall overhead of replication. Several quorum based replica control protocols, have been proposed in past few years. In this paper, we analyze several quorum types in terms of message overhead.

Keywords: Replica-control, distributed database, quorum, message overhead.

1 Introduction

In a distributed database system, data is replicated [1], [15], [10] to achieve fault-tolerance. One of the most important advantages of replication is that it masks and tolerates failures in the network gracefully and increases availability. Specifically the system remains operational and available to the users despite failures. In case of multiple accesses a problem that must be solved while using replication is, how to maintain the copies in a consistent state [8]. To keep logical data consistent, there must exist a control protocol responsible for synchronizing the accesses. A popular method for maintaining consistency of replicated data is weighted voting [7] which is a generalization of the majority consensus method presented in [17]. In the quorum consensus (QC) [11] algorithm, we assign a non-negative weight [6] to each copy x_A of x. We then define a read threshold RT and write threshold WT for x, such that both 2WT and (RT + W T) are greater than the total weight of all copies of x. A read (or write) quorum of x is any set of copies of x with a weight of at least RT (or WT). For better performance, some logical structure is imposed on the network, and the quorums are chosen under the consideration of such structures. Such logical structures include the tree [4], diamond [5], ring [12], grid [14], and wheel [13]. A new logical arrangement of sites for replication, known as Wheel was introduced in [13]. In Replication, sites coordinate their activities by exchanging messages. In practice, this can have a significant impact on the overall behavior of the protocol. On the one side, CPU cycles are lost in dealing with the messages (flattening, sending, receiving, and unflattening). On the other side, the network bandwidth might be exceeded, resulting in additional delays. The message overhead

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 237–245.
springerlink.com

affects not only scalability but also availability. Analysis in this paper, focuses on the number of messages exchanged to measure the impact on CPU time (message overhead at sending and receiving sites) and network load. It is difficult to capture message size without knowing the application domain. Hence, we do not consider the size of the messages because this information is strongly application dependent in most situations.

For the purposes of our study, we consider only the best possible implementation (the one with the least number of messages). Shadow copy approach is used for executing write operations. Updates are executed locally on shadow copy and then sent at the end to the replicas in a single message. All participating sites follow 2PC protocol in an update transaction. In contrast to write operations, read operations cannot be delayed until the end of the transactions or executed on a local shadow copy. Hence, a read operation needs a read request to be send to each member of the read quorum, by the originator of the transaction. In response to this read request, each participating site must reply with read value and its version. For read-write quorum systems (R, W) in which R is 1-uniform (i.e., all read quorums have size rq = 1), we will assume that the read is performed locally and no messages are needed for read operations. Furthermore, transactions that only consist of read operations do not require any message overhead.

The proportion of update operations in the load is represented by *w*. A small w indicates that there are generally few write operations in the system. For instance, the workload can have many queries (read-only transactions) and few update transactions. Each of these few update transactions, however, can have many write operations. We have used same message overhead relations and model as in [8].

With this, assuming that a transaction contains on average O_w write operations, the message overhead for point-to-point messages is defined by:

Definition 1 (Message Overhead, Point-to-Point): The message overhead per operation performed in a quorum system using point-to-point communication is given by the expression:

$$msg = w\frac{3(wq-1)}{Ow} + 2(1-w)(rq-1) \tag{1}$$

If a multicast primitive is available, the number of exchanged messages varies. For all updates, one message is needed for the vote request, $wq - 1$ messages are needed to get the vote message of each participant, and one more message are required to commit or abort the transaction. For each read operation, one needs a message to request the read, and $rq - 1$ messages to get the responses from the participants. Thus:

Definition 16 (Message Overhead, Multicast): The message overhead per operation performed in a quorum system using multicast communication is given by the expression:

$$msg = w\frac{(wq+1)}{Ow} + (1-w)rq \tag{2}$$

In this paper, we will analyze different quorum based protocols in terms of message overhead. Quorums reduce the number of copies involved in reading or updating data. Hence, quorums reduce the overall cost of replication in terms of performance penalties, communication overhead, and systems availability. Comparison between different existing protocol and the new Wheel protocol will be done in this paper.

The paper is organized as follows. In Section 2 we describe the system model. Section 3 discusses different protocols and their message overhead. In Section 4, we present performance evaluation. We conclude the paper in Section 5.

2 Model

A distributed system consists of a set of distinct sites that communicate with each other by sending messages over a communication network. No assumptions are made regarding the speed, connectivity, or reliability of the network. It is assumed that sites are fail-stop [16] and communication links may fail to deliver messages.

Replication of data is achieved by storing copies of the same logical data item at di erent nodes. Read and write operations can be performed on replicated data. A node needs to obtain permission from a number of copies (quorum) before performing the operation using a control protocol.

In a replicated database, copies of an object may be stored at several sites in the network. Multiple copies of an object must appear as a single logical object to the transaction. This is termed as one-copy equivalence [3] and is enforced by the replica control protocol. The correctness criteria for replicated databases is one-copy serializability [3], which ensures one-copy equivalence and serializable execution of transactions. In order to ensure one-copy equivalence, a replicated object z may be read by reading a read quorum of copies, and it may be written by writing a write quorum of copies. If a transaction contains write operations, a 2-phase-commit protocol (2PC) at the end of the transaction is executed among all sites. The following restriction is placed on the choice of quorum assignments:

Quorum Intersection Property: For any two operations $o[Z]$ and $ó[z]$ on an data item x, where at least one of them is a write, the quorums must have a nonempty intersection.

A client submits a transaction and with it all the operations of this transaction to any of the sites in the system. This site is called the originator of the transaction and its operations, and coordinates with the rest of the system. A transaction and its operations are called local at the site it is submitted to, and remote at the other sites. We consider in the study two kinds of transactions: queries, which contain only read operations, and update transactions.

3 Message Overhead Analysis

In this section, we briefly discuss different quorum protocols, their quorum sizes, and derive message overhead expression by using equation 1 and 2.

3.1 The Grid Protocol

Maekawa [11] proposed arranging copies in a logical grid. A read quorum group contains exactly one node from each column. A write quorum group consists of nodes in a read group and all nodes in a column of the grid. Nodes are arranged in grid topology only conceptually, which is used to describe the protocol. If we consider a √Nx√N grid, the size of read and write quorum is √N and 2√N-1, respectively. Message overhead for point-to-point is given as:

$$msg = 2(\sqrt{n} - 1)\left(\frac{3w}{Ow} + (1 - w)\right) \tag{3}$$

Message overhead for multicast communication is represented as:

$$msg = \sqrt{n}\left(\frac{2w}{Ow} + (1 - w)\right) \tag{4}$$

3.2 The Tree Protocol

Tree protocol [2], logically organizes the copies of an object to form a complete binary tree, i.e., if k is the level of the tree, then it has 2^{k+1} - 1 copies, where the root is at level 0. The standard tree terminology, i.e., root, child, parent, leaf, etc., is used. A path in the tree is defined to be a sequence of copies s_1, s_2, . . s_i, s_{i+1}, s_n such that s_{i+1} is a child of s_i. Informal description of the algorithm for constructing a quorum for a binary tree is as follows. A quorum is constructed by selecting any path starting from the root and ending with any of the leaves. If successful, this set of copies constitutes a quorum. If a path cannot be constructed due to the inaccessibility of a copy *c*, residing on a failed or inaccessible site (due to partitioning failures), then the algorithm must substitute for that copy with two paths, both of which start with the children of copy c_i and terminate with leaves. Note that each path must terminate with a leaf, hence if the last copy in the path is inaccessible, the operation must be aborted. Considering a tree with degree of 3, read and write quorum sizes in best case are 1 and $2^{1+\lfloor \log_3 n \rfloor}$ respectively. Based on these quorum sizes message overhead for point-to-point is given in equation 5 and for multicast communication in equation 6 .

$$msg = \frac{3w2^{1+\lfloor \log_3 n \rfloor} - 1}{Ow} \tag{5}$$

$$msg = \frac{w2^{1+\lfloor \log_3 n \rfloor} + 1}{Ow} + (1 - w) \tag{6}$$

3.3 The Hierarchical Protocol

The hierarchical quorum consensus protocol [9] logically organizes a set of copies of an object in a database into a multilevel tree with the root as level 0. Higher level nodes of the tree correspond to logical groups and leaves store physical copies of an object. A node at level *i*, where *i* varies from 0 to m-1 is viewed as a logical group which in turn consists of subgroups at level *i+1*. A quorum is associated with each

level and to access a logical group at a certain level, a quorum consisting of its subgroups must be first assembled. A read (write) quorum at level *i* is defined as the number of subgroups of a level i-1 group l_{i+1} that must be locked by a read (write) operation to obtain read (write) access to the group. The read (write) quorum at level *i* is denoted by $r_i(w_i)$. Note that this is a recursive definition. Therefore, each level i group must in turn assemble r_{i+1} of its subgroups at level $i + 1$, and so on. This would eventually translate into a quorum consisting of physical copies of the object. To perform read (write) operations on the replicated object, a read (write) quorum at level 0 must be obtained first. The quorum size of this protocol is $N^{0.63}$. Message overhead for point-to-point network is given as :

$$msg = (n^{0.63} - 1)\left(\frac{3w}{Ow} + 2(1 - w)\right) \tag{7}$$

Message overhead for multicast network is given as :

$$msg = w\frac{n^{0.63} + 1}{Ow} + (1 - w)n^{0.63} \tag{8}$$

3.4 The Ring Protocol

In the ring protocol [12] copies are organized into a ring structure. It uses the adjacency property to reduce the read and write quorums. There are two protocols - The flat ring protocol and the hierarchical ring protocol. Flat ring arranges nodes in a single ring and achieves a read quorum of two copies (constant), and a write quorum equal to the majority of copies. The hierarchical ring protocol uses a multi-level ring structure and is a generalization of the at ring protocol. For the special case taken in [12], best and worst quorum sizes are given by $q_r = n^{\log_d 2}$ and $q_r = (\lfloor \frac{d}{2} \rfloor + 1)log_d n$. Message overhead for point-to-point and multicast are given below in equations 9 and 10 respectively.

$$msg = w\frac{3\left(\left(\lfloor\frac{d}{2}\rfloor + 1\right)\log_d n - 1\right)}{Ow} + (1 - w)2(n^{\log_d 2} - 1) \tag{9}$$

$$msg = w\frac{\left(\left(\lfloor\frac{d}{2}\rfloor + 1\right)\log_d n + 1\right)}{Ow} + (1 - w)\, n^{\log_d 2} \tag{10}$$

3.5 Diamond Quorum Consensus

The sites in the network are logically organized into a 2- dimensional diamond structure [5]. Diamond is actually as a specialized version of grid protocol because it is a grid with holes. To form a write quorum, we can choose all nodes of any one row plus an arbitrary node for each remaining rows. Read quorum can be formed by choosing any entire row of nodes, or by using an arbitrary node of each row. Minimum read quorum can be obtained by choosing the whole top and bottom row of

nodes plus a node for each remaining row. This protocol can achieve high read capacity, low quorum size, and other desirable features for replicated data management. For diamond quorum consensus optimal read quorum size is 2 and is independent of the total number of sites. Worst case read quorum size is ⌈√(2N)⌉. Optimal and worst quorum sizes for diamond write quorum are ⌈√(2N)⌉ and $2\lceil\sqrt{(2N)}\rceil - 2$ respectively. For best case message overhead for point-to-point communication is given below:

$$\mathbf{msg} = \mathbf{w}\frac{3(\lceil\sqrt{2n}\rceil - 1)}{0\mathbf{w}} + 2(1 - \mathbf{w}) \tag{11}$$

For multicast communication, message overhead is given as:

$$msg = w\frac{(\lceil\sqrt{2n}\rceil + 1)}{0w} + 2(1 - w) \tag{12}$$

3.6 The Wheel Protocol

Nodes are arranged in a logical wheel structure, with one node called HUB in the middle and other nodes around it, forming a cycle [13].Informally, read quorum can be obtained by reading only HUB and write quorum by reading HUB plus alternate spokes. Read quorum size of one is the minimum among all other proposed protocols. In case of failure of HUB, an election algorithm can be used to elect a new hub. Thus making HUB always available and keeping read quorum size minimum. Even in case of no reconfiguration, read quorum can be obtained by reading any two adjacent spokes in the wheel, which is also a smaller read quorum size. Optimal read quorum size is 1 and Write quorum size is given as ⌈(n-1)/2⌉+1.

Message overhead for point-to-point is given as:

$$msg = \frac{3w\lceil\frac{n+1}{2}\rceil}{0w} \tag{13}$$

$$msg = w\frac{\left(\lceil\frac{n+1}{2}\rceil + 2\right)}{0w} + (1 - w) \tag{14}$$

4 Protocols Comparison

Comparison among different protocols message overhead for point-to-point is shown in figure 2. Minimum communication overhead is achieved by tree and wheel protocols. Overhead increases with increasing value of *w*. Tree shows better performance than wheel protocol, whereas, its quorum size is larger than wheel protocol. Wheel protocol, ensures read quorum size of 1 even in worst case, which becomes as big as $(d + 1)^h$ in tree. So, for read intensive applications, wheel gives both the advantages of smallest read quorum and smaller communication overhead.

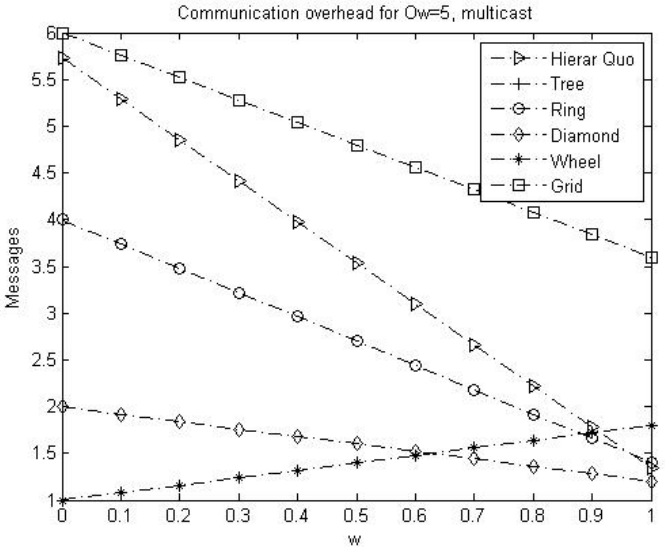


Fig. 1. Message overhead, multicast

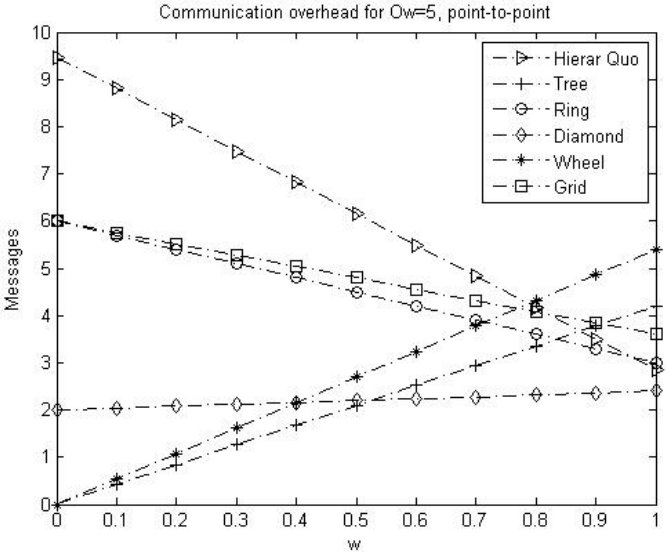


Fig. 2. Message overhead, point-to-point

In case of multicast figure 1, wheel and tree incur minimum message overhead than other protocols. Overhead increases with increasing number of writes in transactions (i.e. *w*).

5 Conclusion

In this paper, we have compared replica control protocols using different logical structures. Among all protocols wheel and tree protocols has shown minimum message overhead. Wheel protocol gives smallest possible quorum size of one, which makes it best of read intensive applications. Wheel protocol provides smallest quorum size with minimum message overhead. Message overhead for wheel protocol multicast never exceeds 2, in fact for lesser number of writes its smaller than 1.5. For point-to-point also, overhead is smaller than most of them. Thus, we conclude that wheel protocol is best choice for read intensive applications, as it gives smallest read quorum and minimum communication overhead.

References

[1] Understanding replication in databases and distributed systems. In: Proceedings of 20th International Conference on Distributed Computing Systems, pp. 464–474 (2000)
[2] Agrawal, D., El Abbadi, A.: An efficient solution to the distributed mutual exclusion problem. In: Proceedings of the Eighth ACM Symposium on Principles of Distributed Computing, pp. 193–200 (August 1989)
[3] Bernstein, P.A., Goodman, N.: A proof technique for concurrency control and recovery algorithms for replicated databases. Distributed Computing 2(1), 32–44 (1987)
[4] Abbadi, A.E., Agrawal, D.: The tree quorum protocol: An efficient approach for managing replicated data. In: Proceedings of the 16th International Conference on Very Large Data Bases, vol. 90, pp. 243–254 (1990)
[5] Fu, A.W.-C., Wong, Y.S., Wong, M.H.: Diamond quorum consensus for high capacity and efficiency in a replicated database system. Distrib. Parallel Databases 8, 471–492 (2000)
[6] Garcia-Molina, H., Barbara, D.: How to assign votes in a distributed system. J. ACM 32, 841–860 (1985)
[7] Gifford, H.: Weighted voting for replicated data. In: Proceedings of 7th Symposium on Operating Systems, pp. 150–162. ACM (1979)
[8] Gray, J., Helland, P., O'Neil, P., Shasha, D.: The dangers of replication and a solution. In: SIGMOD 1996: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, vol. 25, pp. 173–182. ACM, New York (1996)
[9] Kumar, A.: Hierarchical quorum consensus: a new algorithm for managing replicated data. IEEE Transactions on Computers 40(9), 996–1004 (1991)
[10] Lin, Y., Kemme, B., Patiño-Martínez, M., Jiménez-. Peris, R: Consistent Data Replication: Is It Feasible in WANs? (2005)
[11] Liu, M.L., Agrawal, D., Abbadi, E.A.: On the implementation of the quorum concensus protocol. In: Proc. Parallel and Distributed Computing Systems (1995)
[12] Mendona, N.C., Anido, R.O.: The hierarchical ring protocol: An efficient scheme for reading replicated data. Technical Report DCC-93-02. Department of Computer Science. University of Campinas, p. 30 (February 1993) (in English)
[13] Tripathi, M., Pandey, P.: Exploiting logical structures to reduce quorum sizes of replicated databases. Advanced Computing: An Interntional Journal 3, 99–104 (2012)

[14] Ammar, M.H., Cheung, S.Y., Ahamad, M.: The grid protocol: A high performance scheme for maintaining replicated data. IEEE Transactions on Knowledge and Data Engineering 4(6), 582–592 (1992)
[15] Saito, Y., Shapiro, M.: Optimistic replication. ACM Comput. Surv. 37(1), 42–81 (2005)
[16] Schlichting, R.D., Schneider, F.B.: Fail-Stop Processors: An Approach to Designing Fault-Tolerant Computing Systems. Computer Systems 1(3), 222–238 (1983)
[17] Thomas, R.H.: A Majority consensus approach to concurrency control for multiple copy databases. ACM Trans. Database Syst. 4(2), 180–209 (1979)

Modified (Q, r) Policy for Stochastic Inventory Control Systems in Supply Chain

R. Bakthavachalam[1,*], S. Navaneethakrishnan[2], and C. Elango[3]

[1] Raja College of Engineering & Technology, Madurai, TN, India
bakthaa@yahoo.com
[2] V.O.C College, Thoothukudi, TN, India
[3] C P A College, Bodinayakanoor, TN, India
chellaelango@yahoo.com

Abstract. In this paper, we consider a continuous review inventory control for a multi-echelon system, which is a building block for supply chain. The system consists of a warehouse, one distribution center and single retailer. A (s, S) type inventory system with Poisson demand and exponentially distributed lead times is assumed at retailer node and modified (Q, r) type inventory policy is assumed at distribution center. The distribution center replenishes its stocks with exponentially distributed lead times from warehouse which has abundant stocks for supply. Demands occurring at retailer node during the stock out periods are assumed to be lost. The items are supplied from warehouse to distribution center then to retailers in packs of Q (= S-s). The transient, steady state probability distribution of system states and the system performance measures in the steady state are obtained. Numerical examples are provided to illustrate the proposed model.

Keywords: Supply chain, Multi-echelon system, Markov process, Retailer managed inventory control, Optimization.

1 Introduction

The study of supply chain management (SCM) started in the late 1980s and has gained a growing level of interest from both companies and researchers over the past three decades. There are many definitions of supply chain management. Hau Lee, the head of the Stanford Global Supply Chain Management Forum (1999), gives a simple and straight forward definition at the forum website as follows: 'Supply chain management deals with the management of materials, information and financial flows in a network consisting of suppliers, manufacturers, distributors, and customers'. From this definition, we can see that SCM is not only an important issue to manufacturing companies, but is also relevant to service and financial firms. A supply chain may be defined as an integrated process wherein a number of various business entities (suppliers, distributors and retailers) work together in an effort to (1) acquire raw

[*] Part-time Research Scholar - Mathematics - Manonmaniyam Sundarnar University, Tirunelveli.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 247–255.
springerlink.com

materials (2) process them and then produce valuable products and (3)transport these final product to retailers. The process and delivery of goods through this network needs efficient maintenance of inventory, communication and transportation system. The supply chain is traditionally characterized by a forward flow of materials and products and backward flow of information, money, etc.

One of the most important aspects of supply chain management is inventory control. Inventory control models are almost invariably stochastic optimization problems with objective function being either expected costs or expected profits or risks. In practice, a retailer may want an optimal decision which achieves a minimal expected cost or a maximal expected profit with low risk of deviating from the objective.

A complete review of SCM was provided by Benita M. Beamon (1998) [10]. However, there has been increasing attention placed on performance, design and analysis of the supply chain as a whole. HP's (Hawlett Packard) Strategic Planning and Modeling (SPaM) group initiated this kind of research in 1977. From practical stand point, the supply chain concept arose from a number of changes in the manufacturing environment, including the rising costs of manufacturing, the shrinking resources of manufacturing bases, shortened product life cycles, the leveling of planning field within manufacturing, inventory driven costs (IDC) involved in distribution (2005) and the globalization of market economics. With-in manufacturing research, the supply chain concept grow largely out of two-stage multi-echelon inventory models, and it is important to note that considerable research in this area is based on the classic work of Clark and Scarf (1960)[7].

Hadley and Whitin (1963)[8], and among others, present the methods to find the optimal or near optimal solution to minimize the inventory costs at a single stocking point with stochastic demand, based on the continuous review (r,Q), periodic review (R,T), and one-for-one polices. A complete review on this development was recorded by Federgruen (1993) [3]. Recent developments in two-echelon models may be found in Q. M. He, and E. M. Jewkes (2000) [11], S. Axaster (1993) [2], Nahimas (1982) and Antony Svoronos & Paul Zipkin(1991) [9]. A continuous review (s, S) policy with positive lead times in two echelon Supply Chain was considered by K. Krishnan and C. Elango 2007 [6].

This paper deals with a simple supply chain that is modeled as system with a single warehouse, a distribution center and single retailer, handling a single product. In order to avoid the complexity, at the same time without loss of generality, we assumed the Poisson demand pattern at retailer node. This restricts our study to design and analyze as the tandem network of inventory, which is the building block for the whole supply chain system.

The rest of the paper is organized as follows; the model formulation is described in section 2. In section 3, both transient and steady state analysis are done. Section 4 deals the operating characteristics of the system in steady state and section 5, deals with the cost analysis for the operation. Numerical example and sensitivity analysis are provided in section 6. The last section 7 concludes the paper.

2 The Model Description

We consider a three level supply chain system consisting of a warehousing facility, single distribution centre (DC) and one retailer dealing with a single product.

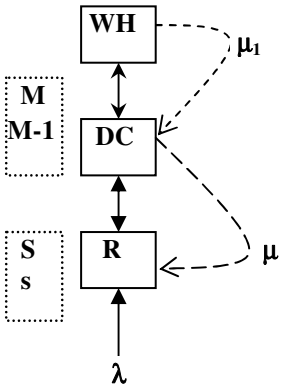


Fig. 1. Tandem Network Inventory Control System in Supply Chain

A finished product is supplied from warehouse to distribution center which adopts modified (Q, r) replenishment policy then the product is supplied from distribution center to retailer who adopts (s, S) policy. The demands at retailer node follows a Poisson process with rate $\lambda>0$. Supply to the retailer in packets of Q = S-s items is administrated with exponential lead time having parameter $\mu>0$. Demands occurring during the stock out periods are assumed to be lost sales at retailer node. A demand at distribution center (Q items) is satisfied by local purchase during the shortage at distribution center. The replenishment of items in terms of pockets are made from warehouse to distribution center with exponential lead time having parameter $\mu_1 > 0$. In our model the maximum inventory level at retailers node S is fixed and the maximum inventory level at distribution centre is a variable M (M = nQ ; Q = S-s , $n \in N$). According to the above definition, on hand inventory levels at both nodes follows a two dimensional random process.

Under the (s ,S) policy, the finished goods inventory is continuously reviewed and a new order is place each time inventory falls to the reorder point 's'. The ordering quantity is Q = S - s. Under the modified (r, Q) policy, finished goods inventory is continuously reviewed and a new order is place each time inventory falls to the reorder point 'r', and the ordering quantity 'Q' is decided only at the time of replenishment. That is Q is equal to maximum inventory level minus current on hand stock (pull back to maximum inventory level).

In our model, the maximum inventory level at distribution center is M (= nQ), the reorder point is M - Q [= (n-1) Q], and the ordering quantity Q is equal to M minus current on hand stock. This is possible when a distribution centre get a truck load of products from the manufacture.

We fix the following notations for the forthcoming analysis part of our paper.

$[R]_{ij}$: the element /sub matrix at $(i,j)^{th}$ position of R

0 : zero matrix of appropriate dimension

I : an identity matrix of appropriate dimension .

e : a column vector of 1's of appropriate dimension

3 Analysis

Let $I_0(t)$ denote the on hand inventory level at retailer node and $I_1(t)$ denote the on hand inventory level at distribution centre node at time t+. From the assumptions on the input and output process, $I(t) = \{(I_0(t), I_1(t)): t \geq 0\}$ is a Markov process with state space $E = \{(j, q) / j = S, S-1 \ldots 3, 2, 1, 0.$ and $q = 0, Q, 2Q \ldots nQ\}$. The infinitesimal generator of this process $A = (a(j,q:k,r))_{(j,q)(k,r)\in E}$ can be obtained from the following arguments.

- The arrival of a demand of an item at retailer node makes a transition in the Markov process from [j, q] to [j-1, q] with intensity of transition $\lambda(>0)$.
- Replenishment of inventory at retailer node makes a transition from [j, nQ] to [j +Q, (n-1)Q] with rate of transition μ (> 0).
- Replenishment of inventory at DC node makes a transition from [j, (n-1) Q] to [j, nQ] with rate of transition μ_1 (> 0).

The infinitesimal generator is given by $R = \begin{matrix} nQ \\ (n-1)Q \\ (n-2)Q \\ \ldots \\ Q \\ 0 \end{matrix} \begin{pmatrix} A & B & 0 & \ldots & 0 & 0 \\ C & D & B & \ldots & 0 & 0 \\ C & 0 & D & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ C & 0 & 0 & \ldots & D & B \\ C & 0 & 0 & \ldots & 0 & E \end{pmatrix}$

The entries of the matrix R can be written as

$$[R]_{q\ x\ r} = \begin{cases} A & \text{if} \quad q = r, \qquad r = nQ \\ B & \text{if} \quad q = r + Q, \quad r = (n-1)Q, (n-2)Q \ldots 2Q, Q \\ C & \text{if} \quad q = 2r - iQ, \quad r = nQ; i = 1, 2 \ldots (n-1) \\ D & \text{if} \quad q = r, \qquad r = (n-1)Q, (n-2)Q \ldots 2Q, Q \\ E & \text{if} \quad q = r, \qquad r = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then the sub matrices A, B, C, D and E are given by

$$[A]_{j\ x\ q} = \begin{cases} \lambda & \text{if} \quad q = j-1, \quad j = S, S-1, S-2, \ldots s, s-1, s-2 \ldots 2, 1. \\ -\lambda & \text{if} \quad q = j, \qquad j = S, S-1, S-2, \ldots (s+1). \\ -(\lambda+\mu) & \text{if} \quad q = j, \qquad j = s, s-1, s-2, \ldots 2, 1. \\ -\mu & \text{if} \quad q = j, \qquad j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$[B]_{j\ x\ q} = \begin{cases} \mu & \text{if} \quad q = j+Q, \quad j = s, s-1, s-2, \ldots 2, 1, 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$[C]_{j\ x\ q} = \begin{cases} \mu_1 & \text{if} \quad q = j, \qquad j = S, S-1, S-2, \ldots s, s-1, s-2, \ldots 2, 1, 0. \\ 0 & \text{otherwise.} \end{cases}$$

$$[D]_{j\,x\,q} = \begin{cases} \lambda & \text{if } q = j-1, \quad j = S, S-1, S-2, \ldots s, s-1, s-2 \ldots 2,1. \\ -(\lambda+\mu_1) & \text{if } q = j, \quad j = S, S-1, S-2, \ldots (s+1). \\ -(\lambda+\mu+\mu_1) & \text{if } q = j, \quad j = s, s-1, s-2, \ldots 2,1. \\ -(\mu+\mu_1) & \text{if } q = j, \quad j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$[E]_{j\,x\,q} = \begin{cases} \lambda & \text{if } q = j-1, \quad j = S, S-1, S-2, \ldots s, s-1, s-2 \ldots 2,1. \\ -(\lambda+\mu_1) & \text{if } q = j, \quad j = S, S-1, S-2, \ldots s, s-1, s-2 \ldots 2,1 \\ -\mu_1 & \text{if } q = j, \quad j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

3.1 Transient Analysis

Let $I_0(t)$ and $I_1(t)$ denote the on hand inventory levels at retailer node and distribution node respectively at time t+. From the assumptions on the input and output process, clearly the vector process $\{I(t) : t \geq 0\}$ where $I(t) = (I_0(t), I_1(t))$ for $t \geq 0$ is a continuous time Markov Chain with state space $E = \{(j, q) / j = 0, 1, 2..., S ; q = 0, Q, 2Q, ..., nQ\}$. Define the transition probability function: $P_{j,q}(k, r : t) = \Pr\{(I_0(t), I_1(t)) = (k, r) \mid (I_0(0), I_1(0)) = (j, q)\}$.The corresponding transient matrix function is given by $P(t) = (P_{j,q}(k, r : t))_{(j,q)(k,r)} \in E$ which satisfies the Kolmogorov forward equation $P'(t) = P(t)A$, where A is an infinitesimal generator. Above equation, together with initial condition $P(0) = I$, yields a solution of the form $P(t) = P(0)e^{At} = e^{At}$ where the matrix expansion in power series form is $e^{At} = I + \sum_{n=1}^{\infty} \frac{A^n t^n}{n!}$.

case(i) :Suppose that the eigen values of A are all distinct. Then from the spectral theorem of matrices, we have $A = HDH^{-1}$, where H is the non-singular (formed with the right eigen vectors of A) and D is the diagonal matrix having its diagonal elements, the eigen values of A. Now 0 is an eigen value of A and if $d_i \neq 0$, i = 1, 2, ...,m, are the distinct eigen values then we have $A^n = HD^nH^{-1}$. Using A^n in P(t) we have the explicit solution of P(t) as $P(t) = He^{Dt}H^{-1}$ with

$$e^{Dt} = \begin{pmatrix} 1 & 0 & .. & .. & 0 \\ 0 & e^{d_1 t} & .. & .. & ... \\ ... & ... & .. & .. & ... \\ 0 & 0 & .. & e^{d_{m-1} t} & 0 \\ 0 & .. & .. & .. & e^{d_m t} \end{pmatrix} \text{ and } D = \begin{pmatrix} 0 & 0 & .. & .. & 0 \\ 0 & d_1 & .. & .. & ... \\ ... & ... & .. & .. & ... \\ 0 & 0 & .. & d_{m-1} & 0 \\ 0 & .. & .. & .. & d_m \end{pmatrix}$$

case(ii): Suppose the eigen values of A are all not distinct, we can find a canonical representation as $A = SZS^{-1}$. From this the transition matrix P(t) can be obtained in a modified form. (Medhi, J).

3.2 Steady State Analysis

The structure of the infinitesimal matrix R reveals that the state space E of the Markov chain {I(t) : t ≥ 0}, is finite and irreducible. Let the limiting distribution of the inventory level process be defined by $P_j^q = \lim_{t\to\infty} \Pr\{(I_0(t), I_1(t)) = (j,q)\}_{(j,q)\in E}$ where P_j^q is the steady state probability for the system be in state (j,q), (Cinlar - 1975[5]).

Let $\nu = \left(\nu^{nQ}, \nu^{(n-1)Q}, \nu^{(n-2)Q} \ldots \nu^{Q}, \nu^{0}\right)$ denote the steady state probability distribution where $\nu^q = \left(\nu_S^q, \nu_{S-1}^q \ldots \nu_0^q\right)$ for the system under consideration. For each (j, q), ν_j^q can be obtained by solving the matrix equation νR = 0. The steady state balance equations are $\nu^{nQ} A + \sum_{i=0}^{n-1} \nu^{iQ} C = 0$ and $\nu^{jQ} B + \nu^{(j-1)Q} D = 0 \; ; j = n, n-1, \ldots 1$, together with normalizing condition $\sum_{j,q} \nu_j^q = 1$.

4 Operating Characteristics

In this section, we derive some important system performance measures.

(1) Mean reorder rate: The mean reorder rate at retailer node and the DC are β0, β1 respectively thus

$$\beta_0 = \lambda \sum_{q=Q}^{nQ} P_{s+1}^q \tag{1}$$

and

$$\beta_1 = \mu \sum_{j=0,q=0}^{s,(n-1)Q} P_j^q \tag{2}$$

(2) Mean inventory level: Let $\overline{I_i}$ denote the mean inventory level in the steady state at node i (i=0,1). Thus,

$$\overline{I_0} = \sum_{q=Q}^{nQ} \left(\sum_{j=0}^{S} j.p_j^q \right) \tag{3}$$

and

$$\overline{I_1} = \sum_{j=0}^{S} \left(\sum_{q=Q}^{nQ} q.p_j^q \right) \tag{4}$$

(3) Expected shortage rate: The shortages occur at retailer node and DC are α_0, α_1, thus

$$\alpha_0 = \lambda \sum_{q=Q}^{nQ} P_0^q \tag{5}$$

and

$$\alpha_1 = \mu \sum_{j=0}^{s} P_j^0 \quad \backslash \tag{6}$$

5 Cost Analysis

In this section we analyze the cost structure for the proposed model by considering the minimization of the steady state total expected cost per unit time. Let k_1 - the fixed ordering costs from warehouse to DC, k_0 - the fixed ordering costs from distribution centre to retailer, h_1 - the holding cost per unit of item per unit time at DC, h_0 - the holding cost per unit of item per unit time at retailer node, g_0 - the shortage cost per unit shortage at retailer node. g_1 – the penalty cost per unit shortage at DC. (Since the demands satisfied by some local purchase during stock out period at DC) The long run expected cost rate C(s, Q) is given by

$$C(s,Q) = h_0\overline{I_0} + h_1\overline{I_1} + k_0\beta_0 + k_1\beta_1 + \alpha_0 g_0 + \alpha_1 g_1 \tag{7}$$

Although we have not proved analytically the convexity of the cost function C(s,Q), our experience with considerable number of numerical examples indicates that C(s,Q) for fixed Q to be locally convex in s. In some cases it turned out to be an increasing function of s. Hence we adopted the numerical search procedure to determine the optimal values s*, consequently we obtain optimal $n^* = \left[M/Q^* \right]$.

6 Numerical Example and Sensitivity Analysis

In this section we discuss the problem of minimizing the steady state total expected cost rate under the following cost structure. The results we obtained in the steady state case may be illustrated through the following numerical example.
For the input, S = 12, M = 3 (3Q), $\lambda = 0.5$, $\mu_0 = 0.75$, $\mu_1 = 1.25$, $h_0 = 3.25$, $h_1 = 1.75$, $k_0 = 0.45$, $k_1 = 0.55$, $g_0 = 1.35$, $g_1 = 2.25$. We get the cost for different reorder levels(S) as follows,

Table 1. The total expected cost rate as a function of C(s,Q)

s	C(s, Q)	Q	result
1	264.3862	11	For each of the inventory capacity S, the optimal reorder level's' and optimal cost C(s,Q) are indicated by the symbol '*'.
2	251.7745	10	
3	239.7739	9	
4	227.8788	8	
5*	215.8550*	7	

Table 2. The total expected cost deviation based on various g_0 and S

S / g_0	9	12	15	observations
0	141.7051	215.7670	288.6348	For the same input with reorder at s = 5. It is observed that if the maximum inventory level S is increased then the total expected cost C(s, Q) also increase depending upon the various shortage rates (g_0).
0.2	141.7800	215.7800	288.6439	
0.4	141.7930	215.7930	288.6530	
0.6	141.8061	215.8061	288.6622	
0.8	141.8191	215.8191	288.6713	
1	141.8322	215.8322	288.6804	

Table 3. The total expected cost deviation based on various g_0 and h_0

h_0 / g_0	2.5	3.5	4.5	observations
0	260.8061	297.9110	335.0159	For the same input with reorder at s = 5. It is observed that the total expected cost C(s,Q) is increasing with the different holding cost and shortage rate of the item. Hence the shortage rate and holding cost are key parameter of this system.
0.2	260.8157	297.9201	335.0250	
0.4	260.8244	297.9293	335.0342	
0.6	260.8335	297.9394	335.0433	
0.8	260.8426	297.9475	335.0387	
1	260.8518	297.9566	335.0615	

7 Concluding Remarks

In this paper, we analyzed a continuous review inventory control system to Multi-echelon system. The structure of the chain allows vertical movement of goods from warehouse to distribution center then to retailer. The model dealing with the supply from warehouse to distribution center then to retailer is in the terms of pockets. We are also proceeding in this multi-echelon stochastic inventory control system with perishable products and also dealing with backlogging. This model deals with only tandem network (basic structure of supply chain). This structure can be extended to tree structure and to be more general.

References

1. Cinlar, E.: Introduction to Stochastic Processes. Prentice Hall, Englewood Cliffs (1975)
2. Axaster, S.: Exact and approximate evaluation of batch ordering policies for two level inventory systems. Operation Research 41, 777–785 (1993)
3. Federgruen, A.: Centralized planning models for multi echelon inventory system under uncertainty. In: Graves, S.C. (ed.) Handbooks in ORMS, vol. 4, pp. 133–173. North-Holland, Amsterdam (1993)
4. Medhi, J.: Stochastic processes. New age international publishers, India (2009)

5. Elango, C.: A continuous review perishable inventory system at service facilities, unpublished. Ph. D., Thesis. Madurai Kamaraj University, Madurai (2000)
6. Krishnan, K.: Stochastic Modeling in Supply Chain Management System, unpublished. Ph.D., Thesis. Madurai Kamaraj University, Madurai (2007)
7. Clark, A.J., Scarf, H.: Optimal Policies for a Multi- Echelon Inventory Problem. Management Science 6(4), 475–490 (1960)
8. Hadley, G., Whitin, T.M.: Analysis of inventory systems. Prentice-Hall, Englewood Cliffs (1963)
9. Svoronos, A., Zipkin, P.: Evaluation of One-for-One Replenishment Policies for Multiechelon Inventory Systems. Management Science 37(1), 68–83 (1991)
10. Beamon, B.M.: Supply Chain Design and Analysis, Models and Methods. International Journal of Production Economics 55(3), 281–294 (1998)
11. He, Q.M., Jewkes, M.: Performance measures of a make-to-order inventory- production system. IIE Transactions 32, 409–419 (2000)
12. Bakthavachalam, R., Elango, C.: Multi-Echelon Stochastic Inventory Control Systems in Supply Chain. Computational and Mathematical Modelling, pp. 191–199. Narosa Publishing House, New Delhi (2011)
13. Bakthavachalam, R., Elango, C.: Perishable Inventory Control System with Partial Backlogging in Supply Chain. The PMU Journal of Humanities and Science 2(2), 31–41 (2011)

Single Input Variable Universe Fuzzy Controller with Contraction-Expansion Factor for Double Inverted Pendulum

Yogesh Kr. Dhanni and M.J. Nigam

IIT Roorkee, India
Yogesh.dhanni@gmail.com, mkndnfec@ iitr.ernet.in

Abstract. To stabilize the double inverted pendulum, a single input variable universe fuzzy controller is designed. In conventional fuzzy controller, input variables are *the error* and *the change-in-error* but in this single input fuzzy controller, input variable is *the signed distance*. The universe of discourse of this single input fuzzy controller is varied with the help of contraction-expansion factor in order to improve the accuracy and respond speed. This control method has a high accuracy as well as improved response time over conventional fuzzy controller which has been observed by experiments.

1 Introduction

The Double Inverted Pendulum (DIP) is a multivariable, nonlinear fast reaction and unstable system [1]. As it is a challenging problem to stabilize a double inverted pendulum, therefore it can also be used to analyze the performance of any control method. Fuzzy control, variable structure control and robust control are some of the methods which commonly used to solve this problem. The performance of Fuzzy Logic Controller (FLC) depends on the number of its inference rules. The performance of the FLC can be easily enhanced by increasing the number of rules. But the large set of rules also requires more computational time [2]. This problem is solved by the introduction of Single input Fuzzy Logic Controller (SFLC) [3]. In conventional fuzzy controllers, the input variables are mostly the error and the change-in-error but in SFLC the input variable is the *signed distance*. This signed distance variable is sole fuzzy input variable in single input fuzzy logic controller.

Traditional fuzzy controller has many advantages, but its control accuracy is low [4]. So this type of method is not appropriate in such applications where highly precise control is required. For high precision, a variable universe adaptive fuzzy controller was proposed by professor Li in 1999 [5]. The controlling power of variable adaptive fuzzy control is verified for effective dealing with nonlinear system [6]. So in this paper a controller is designed using the technique of variable universe fuzzy control and having single input to stabilize the double inverted pendulum.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 257–265.
springerlink.com

2 Modeling of Double Inverted Pendulum

2.1 Double Inverted Pendulum Structure

The structure of DIP is shown in Fig.1. DIP can be simplified as a system of cart and two quality rods, where M is the mass of cart, m_1 is the mass of pendulum1, m_2 is the mass of pendulum2, m_3 is the mass of joint, l_1 is the length of pendulum1, l_2 is the length of pendulum2, θ_1 is the angle between pendulum1 and vertical, θ_2 is the angle between pendulum2 and vertical and F is the external force acting on the system. The main objective is to erect the stable pendulums mounted on the cart, within the limited rail length and to achieve dynamic balance.

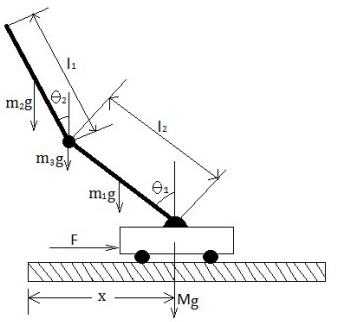


Fig. 1. Schematic diagram of double inverted pendulum system

2.2 Mathematical Model of Double Inverted Pendulum

The mathematical model of the double inverted pendulum system is established using Lagrange equation, taking the state variables [13]:

$$x_1 = x,\ x_2 = \theta_1,\ x_3 = \theta_2,\ x_4 = \dot{x},\ x_5 = \dot{\theta}_1,\ x_6 = \dot{\theta}_2,$$

The equation state is taken around;

$$X = \begin{bmatrix} x & \theta_1 & \theta_2 & \dot{x} & \dot{\theta}_1 & \dot{\theta}_2 \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$$

Table 1 shows the parameter of the DIP used to derive state space model.

Table 1. Parameters of Double Inverted Pendulum[13]

Parameter	Definition	Value	Unit
M	Mass of Cart	1.096	Kg
m_1	Mass of pendulum1	0.05	Kg
m_2	Mass of pendulum2	0.13	Kg
m_3	Mass of joint	0.236	Kg
l_1	Length of pendulum1	0.0775	m
l_2	Length of pendulum2	0.25	m
g	Gravity constant	9.8	N/Kg

By substituting the parameters, the following linear model is obtained:

$$\begin{cases} \dot{X} = AX + Bu \\ Y = CX + Du \end{cases} \tag{1}$$

Where:

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 86.69 & -21.62 & 0 & 0 & 0 \\ 0 & -40.31 & 39.45 & 0 & 0 & 0 \end{bmatrix}, B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 6.64 \\ -0.088 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, D = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

3 Single Input Fuzzy Logic Controller

The Conventional Fuzzy Logic Controller (CFLC) has two inputs, which are mostly the error and the change-in-error. It requires a 2-dimensional rule table for inference. This rule table is in skew-symmetric form, i.e. output membership is same in diagonal direction. Each point on the particular diagonal line has a magnitude that is proportional to the distance from its main diagonal line (L_Z). For any combination of (e,ė), the output membership function will lie in any one of the diagonal line (L_{NB}, L_{NM}, L_{NS}, L_Z, L_{PS}, L_{PM}, L_{PB}). The main diagonal line (L_Z) can be representation as [7]:

$$\dot{e} + \lambda e = 0 \tag{2}$$

Where, λ is the slope magnitude of the main diagonal line L_Z. The distance from any point (e,ė) to the main diagonal line can be written as [7]:

$$d = \frac{\dot{e} + \lambda e}{\sqrt{1 + \lambda^2}} \tag{3}$$

Depending on the distance d, the new rule table can be constructed and given in Table 2. Rule table is one dimensional and contains only seven rules and confirms linear control surface. Number of input for FLC will be one and structure of single input FLC is given in Fig.2 [8]. The calculated distance (d) is the only input to the fuzzy logic controller.

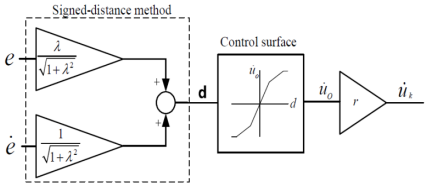


Fig. 2. Single input Fuzzy Logic Controller

Table 2. SFLC Rule Base

d	LNL	LNM	LNS	LZ	LPS	LPM	LPL
U	NL	NM	NS	Z	PS	PM	PL

4 Variable Universe Fuzzy Logic

Let $X_i = [-E_i, E_i](i = 1, 2, ..., n)$ be the universe of the input variable $x_i (i = 1, 2, ..., n)$ and $Y = [-U, U]$ be the universe of the output variable *y*. $\phi_i = \{A_{ij}\}_{(j=1,2,...,m)}$ stand for the fuzzy sets X_i and $\Psi_j = \{B_j\}$ stand for the fuzzy sets *Y*. ϕ_i and Ψ_i can be called the linguistic variables. Fuzzy inference rule set $\{R_s\}_{(s=1,2,...,z)}$ can be formed as:

$$R_s: \text{If } x_1 \text{ is } A_{1j}, \ldots, \text{ and } x_n \text{ is } A_{nj} \text{ then } y = B_j \tag{4}$$

The so-called variable universe means that some universes such as X_i and *Y*, can change along with changing variables x_i and *y* [9]. The transformed universe discourse is denoted as:

$$X_i(x_i) = [-\alpha_i(x_i)E_i, \alpha_i(x_i)E_i] \tag{5}$$

$$Y(y) = [-\beta(y)U, \beta(y)U] \tag{6}$$

Where $\alpha_i(x_i)$ and $\beta(y)$ are contraction-expansion factors [10]. The varying universe is shown in Fig.3 [11]:

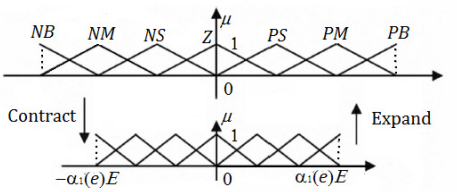


Fig. 3. Variable situation of the universe

Therefore change rule of $\alpha(x)$ is:

$$\Delta\alpha = k.\Delta x.x.(1-\alpha) \tag{7}$$

Integrating both sides to obtain $\alpha(x)$;

$$\alpha(x) = 1-\eta e^{-kx^2} \tag{8}$$

Assume $\beta(t)$ is the universe contraction-expansion factor of output *Y*. $\beta(t)$ is designed with the principle of weighted integral [4].

Where K_I is proportionality constant, $P_n = [p_1, p_2]^T$ is a constant vector.

When $\beta(t) = 1$, $K_I = 1$, $P_n = [1, 1]^T$,

$$\beta(t) = \int_0^t (e+ec)d\tau + 1 \tag{10}$$

5 Control Scheme for Double Inverted Pendulum

The six variables make the double inverted pendulum a six dimensional system. In order to simplify the complexity of the system all three errors (E_1, E_2 & E_3) and change-in-errors (EC_1, EC_2 & EC_3) should be synthesize into only two variables the error (*E*) and the change in error (*EC*). This can be done by the help of Information Fusion Method [12]. After this, the signed distance variable (*d*) is obtained by the help of Signed Distance Method [7]. This signed distance variable is fed to fuzzy controller as sole fuzzy input. Then variable universe technique is used to improve the accuracy and respond time of the system [5].

5.1 Implementation of Information Fusion

Error *E* and change-in-error *EC* can be defined as:

$$E \triangleq \begin{bmatrix} k_1 & k_2 & k_3 \end{bmatrix} \begin{bmatrix} x \\ \theta_1 \\ \theta_2 \end{bmatrix} \tag{11}$$

$$E \triangleq \begin{bmatrix} k_4 & k_5 & k_6 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} \tag{12}$$

Where $(k_1, k_2, k_3, k_4, k_5, k_6)$ are synthesis parameters.

Now a state feedback matrix for the state equation has to be designed. For this, make the quadratic performance index function [12];

$$J = \frac{1}{2}\int_0^\infty (X^T QX + U^T RU)dt \tag{13}$$

Where the positive semi definite matrix Q = *diag* (200, 50, 50, 0, 0, 0) and symmetric positive definite matrix R= 1.

For solving the Riccati equation:

$$A^T P + PA - PBR^{-1}B^T P + Q = 0 \tag{14}$$

The optimal feedback gain matrix values can be obtained:

$$K = R^{-1}B^T P \tag{15}$$

K = (14.1421, 93.1921, -152.13, 14.0841, 2.9657, -24.7386)

The fuzzy controller of the upper pendulum main control variable has best result [1]. So in order to consider pendulum2 as the main control variable, above equations are transformed as:

$$E = \begin{bmatrix} \frac{14.1421}{-152.13} & \frac{93.1921}{-152.13} & 1 \end{bmatrix} \begin{bmatrix} x \\ \theta_1 \\ \theta_2 \end{bmatrix} \tag{16}$$

$$E = \begin{bmatrix} \frac{14.0841}{-24.7386} & \frac{2.9657}{-24.7386} & 1 \end{bmatrix} \begin{bmatrix} \dot{x} \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} \tag{17}$$

5.2 The Signed Distance Variable

The error *E* and the change-in-error *EC* are combined to obtain the signed distance *d* by using (3). The gain factor for error is taken as 3 whereas for change-in-error it is taken as 0.5. The schematic diagram for SFLC is shown in Fig. 2.

5.3 Variable Universe Fuzzy Controller

The universe contraction-expansion factor of *Ds* from (8) is:

$$\alpha(e) = 1 - \eta \exp(-ke^2)$$

On choosing, $\eta = 0.27$, $k = 10^{-2}$;

$$\alpha(e) = 1 - 0.27 \exp(-10^{-2} e^2)$$

Assume $\beta(t)$ is the universe contraction-expansion factor of output *U*. Then $\beta(t)$ from (10) is:

$$\beta(t) = \int_0^t (d)\, d\tau + 1$$

6 Simulation Results

MATLAB SIMULINK is used in this paper for simulation of controller to control double inverted pendulum. The initial fuzzy universe of D_s is taken [-1 1] and for the output *U* it is [-1 1]. The membership functions of input and output variables have seven variables with triangular membership. The control rules designed for double inverted pendulum are described in Table 2.

In control of double inverted pendulum, the stability of inverted pendulum at the given position is highly sensitive to the initial position of cart and the initial angles of both the inverted pendulums. Now following two cases are taken with different initial conditions:

Case A: In this case the initial simulation conditions are set at:

$$x = 0.1m,\ \ \theta_1 = 0.1rad,\ \ \theta_2 = 0.1rad$$

The length of simulation step is taken 1*ms* and simulation time is 5 seconds. Now the cart is required to move at $x = 0$. Simulation results for *case A* are shown in Fig. 4. From the simulation results it can be observed that system reach equilibrium position within 2.5 seconds whereas for conventional fuzzy controller it takes 3 seconds as shown in [12].

Case B: In previous case both the pendulums were tilted in same direction but in this case both pendulums are tilted in opposite directions. In *case B* the initial simulation conditions are set at:

$$x = 0.1m,\ \ \theta_1 = 0.1rad,\ \ \theta_2 = -0.1rad$$

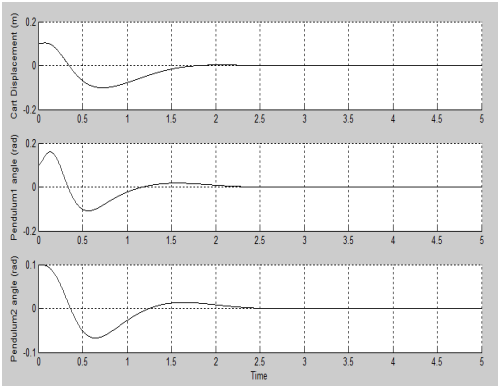


Fig. 4. Simulation results for case A

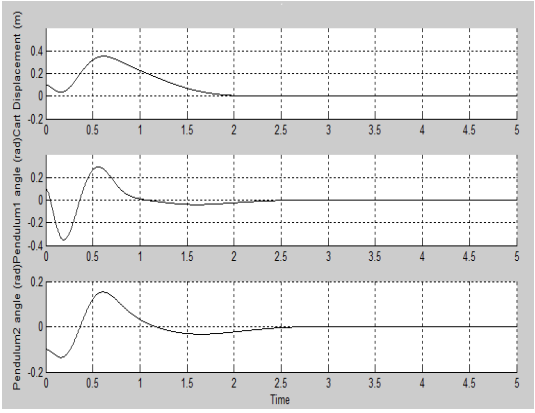


Fig. 5. Simulation results for case B

The length of simulation step is taken 1*ms* and simulation time is 5 seconds. Now the cart is required to move at $x = 0$. Simulation results for *case B* are shown in Fig. 5. From the simulation results it can be observed that system reach equilibrium position within 2.5 seconds whereas for conventional fuzzy controller it takes 3 seconds as shown in [12].

7 Conclusion

A double inverted pendulum model based on Lagrange is taken as controlled object for stabilization. To simplify the controller design, three pairs of *the error* and *the change-in-error* are combined into single pair of *the error (E)* and *the change-in-error (EC)* by the method of information fusion. Then, *E* and *EC* are merged to form the *signed distance (d)* variable with the help of signed distance method. A suitable single input fuzzy controller with variable universe of discourse is designed and the length of universe of discourse is adjusted by universe contraction-expansion factor.

Simulation results are obtained for two different cases with different initial conditions. From the simulation results, it can be observed that performance of the controller is precise in nature and also poses high degree of accuracy to the conventional fuzzy controller.

References

1. Li, Q.-R., Tao, W.-H., Sun, N., Zhang, C.-Y., Yao, L.-H.: Stabilization Control of Double Inverted Pendulum System. In: Proc. IEEE Innovative Computing Information and Control (ICICIC 2008), p. 417 (June 2008)
2. Amjad, M., Kashif, M.I., Abdullah, S.S., Shareef, Z.: A Simplified Intelligent Controller for Ball and Beam System. In: Proc. IEEE Conference on Education Techonology and Computer (ICETC 2010), vol. 3, pp. 494–498 (2010)
3. Choi, B.-J., Kwak, S.-W., Kim, B.K.: Design of a single-input fuzzy logic controller and its properties. Trans. of Fuzzy Sets and Systems 106, 299–308 (1999)
4. Beibei, H., Yan, G.: Variable Universe Fuzzy Controller with Correction Factors for Ball and Beam System. In: Proc. IEEE Intelligent Systems and Applications (ISA), pp. 1–4 (May 2011)
5. Li, H.X.: Adaptive fuzzy controllers based on variable universe. Science in China, Ser. E 42, 10–20 (1999)
6. Xiaofeng, G., Hongxing, L., Guannan, D., Haigang, G.: Variable Universe Adaptive Fuzzy Control on the Triple Inverted Pendulum and Choosing Contraction-Expansion Factor. In: Deng, H., Wang, L., Wang, F.L., Lei, J. (eds.) AICI 2009. LNCS, vol. 5855, pp. 63–67. Springer, Heidelberg (2009)
7. Choi, B.-J., Kwak, S.-W., Kim, B.K.: Design and Stability Analysis of Single-Input Fuzzy Logic Controller. Trans. of Systems, Man and Cybernetics 30, 303–309 (2000)
8. Taeed, F., Salam, Z., Ayob, S.M.: Implementation of Single Input Fuzzy Logic Controller for Boost DC to DC Power Converter. In: Proc. IEEE Power and Energy (PECon), pp. 797–802 (December 2010)
9. Cao, D.-Y., Zeng, S.-P., Li, J.-H.: Variable universe fuzzy expert system for aluminum electrolysis. Trans. of Nonferrous Metals Society of China 21, 429–436 (2011)

10. Wang, J., Qiao, G.-D., Deng, B.: H_∞ Variable universe adaptive fuzzy control for chaotic system. Trans. Chaos, Solitons & Fractals 24, 1075–1086 (2005)
11. Wang, J., Zhang, W.-W.: Chaos control via variable universe fuzzy theory in auto gauge control system. In: Proc. IEEE Computer Application and System Modeling (ICCASM), vol. 1, pp. V1-363–V1-369 (October 2010)
12. Wang, L., Zheng, S., Wang, X., Fan, L.: Fuzzy control of a double inverted pendulum based on information fusion. In: Proc. IEEE Intelligent Control and Information Processing (ICICIP), pp. 327–331 (August 2010)
13. Googol Technology Inverted pendulum Experiment Manual, 2nd edn. Googol Inc. (July 2006)

Performance Evaluation of A Novel Most Recently Used Frequency Count (MRUFC) List Accessing Algorithm

Rakesh Mohanty[1] and Ashirbad Mishra[2]

[1] Department of Computer Science and Engineering
Indian Institute of Technology, Madras
Chennai, India-600036
rakesh.iitmphd@gmail.com
[2] Department of Computer Science and Engineering
Veer Surendra Sai University of Technology, Burla
Sambalpur, Odisha, India – 768018
ashirbadm@gmail.com

Abstract. List Accessing Problem (LAP) is a problem of significant theoretical and practical interest in the context of linear search. Move-To-Front (MTF) Transpose(TRANS) and Frequency Count(FC) are the three most primitive list accessing algorithms developed in the literature. FC is the static optimal offline algorithm, but it has not been studied extensively in the literature till date. In this paper, a comprehensive study of FC algorithm has been done. After our analysis we have explored the limitation of FC algorithm which is a scope to improve its performance. Based on an idea of combining the concepts of Most Recently Used (MRU) of paging algorithm and FC List Accessing algorithm, we have proposed a novel hybrid list accessing algorithm which we call as Most Recently Used Frequency Count(MRUFC) algorithm. We have evaluated the performance of FC list accessing algorithm and our proposed MRUFC algorithm by using Calgary Corpus as the input dataset. Our experimental results show that MRUFC performs better than FC for all the request sequences generated from the input dataset.

Keywords: Algorithm, Data Structure, Linear Search, Linked List, List Accessing Problem, Frequency Count Algorithm, Empirical study.

1 Introduction

List Accessing Problem (LAP) is a computational problem of significant theoretical and practical interest in the context of linear search for the last four decades since the pioneering work of McCabe in 1965[1]. In this problem, the inputs are an unsorted linear list l of finite items and a sequence of requests σ where each request is an access or search operation on an item of the list. A request sequence σ is said to be served on the list l, when each item of the σ is linearly searched one by one in the list by incurring some access cost based on a cost model. After each access the list is rearranged by incurring some reorganization cost. A List accessing algorithm

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 267–276.
springerlink.com

minimizes the total access and reorganization cost while serving a request sequence σ on a list. The most widely used cost model for a LAP is the standard full cost model introduced by Sleator and Tarjan[2]. In this cost model, the access cost is i^{th} item of the list from the front is i i.e the position of the item from the front of the list.

1.1 Applications and Motivation

List accessing algorithms are widely used in Data Compression. Some other important applications of list accessing algorithms are computing point maxima in computational geometry, resolving collisions in hash table and dictionary maintenance. The majority of research work in the literature is based on theoretical analysis of three primitive list accessing algorithms- Move-To-Front(MTF), Transpose(TRANS) and Frequency Count(FC). The empirical studies of list accessing algorithms have gained importance due to its practical applications in various real life situations.

1.2 FC Algorithm

FC is considered to be the static optimal algorithm for the list accessing problem. In FC, a frequency counter is maintained for each item of the list to count the number of accesses of each item from the request sequence. After each access, the counter of the accessed item is incremented by 1. The items are rearranged in non-increasing order their frequency counts in the list. In FC, the accessed items which have same frequency count are arranged in First Come First Serve (FCFS) order among themselves in the list, thereby maintaining the relative order of these items in the list as before. If some accessed items appear far from the front of the list and are more frequently requested in future, the access cost of serving these items becomes more. To overcome this limitation and to reduce the future access cost, most recently used items need to be kept towards the front of the list.

1.3 Our Contribution

In this paper, we have proposed a new variant of the FC algorithm by using the Most Recently Used (MRU) concept of paging. Our proposed hybrid algorithm is named as Most Recently Used Frequency Count (MRUFC). We have experimentally evaluated the performances of our proposed MRUFC algorithm by using the Calgary Corpus as input data set. We have also evaluated the performance of FC algorithm using the same data set. Our comparative performance evaluation shows that MRUFC always performs better than the FC.

1.4 Literature Review

In their seminal paper Sleator and Tarjan[2] have shown that the well-known MTF algorithm is 2-competitive. Irani[3] has provided a matching upper bound for MTF

algorithm. Albers [4] have developed an algorithm called TIMESTAMP and proved that it is 2-competitive. For randomized algorithms, so far the best known upper bound of 1.6 is obtained by an algorithm called COMB due to Albers, von Stengel and Werchner [5]. A well-known lower bound for any randomized algorithm is 1.5 due to Teia[6] . Few comprehensive surveys of list accessing algorithms with associated results can be found in [7], [8], and [9]. Various empirical studies of List accessing algorithms and associated results are mentioned in [10], [11], [12], [13], [14], [15].

1.5 Organization of Paper

This paper is organized as follows: Introduction is presented in Section 1. The analysis of FC algorithm and scope for its improvement are presented in section 2. Section 3 contains our new proposed MRUFC algorithm. Section 4 contains our experimental results of performance evaluation obtained from the comparison of access costs of FC and MRUFC. Section 5 provides some concluding remarks.

2 Our Proposed Algorithm

2.1 Uniqueness of Our Approach

The Most Recently Used (MRU) concept has been extensively used in other areas of computer science, such as Caching and Paging. In MRU, the item that has been most recently used has greater probability of access in near future. Hence, we have used the concept of MRU in the FC algorithm to develop a novel hybrid algorithm with better performance.

2.2 Assumptions and Terminologies

We assume that the request sequence σ is given as input and the list l is generated from σ. The size of σ is assumed to be greater than the size of l. We use doubly linked list as the data structure for our experimentation. As the list rearrangement is achieved through a change of constant number of pointers, the reorganization cost is neglected. Hence, the total cost is computed based on only access costs. Here we use the standard full cost model for computing the access cost. The list configuration is denoted by *L*. *i, j, k* are used as variables for items of the list *L*. *i[count]* is the counter field of item *i* which shows the number of times the item has been requested. Let *i[usage]* is the usage field of item *i,* which stores the last usage value of the item. *current_usage* is a variable which stores the current usage value. It also stores the number of items that have been requested before the current request. The pseudo code of our proposed algorithm is presented below..

```
Procedure MRUFC (L)
{
        set current_usage = 0;
        for all items k in L
                k[count] = 0 and k[usuage]=0;
        while (item i is requested)
        {
                set flag = 0
                i[count]++
                for all items j in L
                        if (i==j)
                                flag =1
                                exit from the 'for' loop
                        if (i[count] > j[count])
                                insert item i before item j in list
                                flag = 1
                                exit from the 'for' loop
                        else if ((i[count]==j[count]) && (i[usage]>j[usage]))
                                insert item i before item j in list
                                flag=1
                                exit from the 'for' loop
                if (flag!=1)
                                insert item i at end of L
                                i[count]=1
                current_usage ++
                i[usage] = current_usage
        }
}
```

Fig. 1. Pseudo code for MRUFC

2.3 Illustration of MRUFC

Suppose we have a list sequence LS as 1 2 3 4, and the request sequence RS is 2 2 3 1 4 3 3 1 1 2 4 4 4 2 3. The count and usage value of each of the items in LS is initialized to 0. We represent the service of the each request for item in table 1. For

FC, Initially the list configuration is 1 2 3 4 and the RS 2 2 3 1 4 3 3 1 1 2 4 4 4 2 3 is served one by one. After each request of the RS the list configuration may change as shown in the table 1. Finally, the list configuration becomes 4 2 3 1. And the frequency count of items 1 2 3 4 are 3, 4, 4 and 4 respectively. The sum of access costs is 45. For MRUFC, initially the list configuration is 1 2 3 4, the same RS is served one by one. After each request of the RS the list configuration may change as shown in the table. Finally, the list configuration becomes 4 2 3 1; and the count of each item is 3, 4, 4 and 4 respectively. The total cost becomes 41.

Table 1. Illustration of MRUFC and FC

Requests	Configuration of list for FC	Configuration of list for MRUFC	Count of items 1 2 3 4	Usage of items 1 2 3 4	current_usage	Cost of FC	Cost of MRUFC
	1 2 3 4	1 2 3 4	0 0 0 0	0 0 0 0	0	0	0
2	2 1 3 4	2 1 3 4	0 1 0 0	0 1 0 0	1	2	2
2	2 1 3 4	2 1 3 4	0 2 0 0	0 2 0 0	2	1	1
3	2 3 1 4	2 3 1 4	0 2 1 0	0 2 3 0	3	3	3
1	2 3 1 4	2 3 1 4	1 2 1 0	4 2 3 0	4	3	3
4	2 3 1 4	2 3 1 4	1 2 1 1	4 2 3 5	5	4	4
3	2 3 1 4	3 2 1 4	1 2 2 1	4 2 6 5	6	2	2
3	3 2 1 4	3 2 1 4	1 2 3 1	4 2 7 5	7	2	1
1	3 2 1 4	3 1 2 4	2 2 3 1	8 2 7 5	8	3	3
1	3 1 2 4	1 3 2 4	3 2 3 1	9 2 7 5	9	3	2
2	3 1 2 4	1 3 2 4	3 3 3 1	9 10 7 5	10	3	3
4	3 1 2 4	1 3 2 4	3 3 3 2	9 10 7 11	11	4	4
4	3 1 2 4	4 1 3 2	3 3 3 3	9 10 7 12	12	4	4
4	4 3 1 2	4 1 3 2	3 3 3 4	9 10 7 13	13	4	1
2	4 2 3 1	4 2 1 3	3 4 3 4	9 14 7 13	14	4	4
3	4 2 3 1	4 2 3 1	3 4 4 4	9 14 15 13	15	3	4

3 Experiment and Results

In our work, we have implemented FC and MRUFC list accessing algorithms using C language and Linux operating system. An empirical test of performances of FC and MRUFC list accessing algorithms was done with respect to the request sequences generated from Calgary Corpus as the input dataset. The access costs of both FC and MRUFC were computed by serving the request sequences generated from the Calgary Corpus data set. The access costs of both algorithms have been compared for each request sequence.

3.1 Input Dataset

The Calgary corpus is a collection of (mainly) text files that serves as a popular bench mark for testing the performance of (text) compression algorithm and can also be used for access cost performance testing. The corpus contains 9 different types of files and overall 17 files. In particular it contains books, papers, numeric data, pictures, programs and object files. Each file was used to generate 2 different request sequences. The first sequence was generated by parsing the files into "words" ('word' parsing). A word is defined as the longest string of non space characters. For some of the non text files in the corpus (e.g. pic), the parsing does not yield a meaningful sequence, hence results are ignored. The second sequence is generated by reading the file as a sequence of bytes (Byte Parsing).

3.2 Experiment Performed

The input to each algorithm is a byte or word parsing of each of the file. The LS is created when the RS are parsed. Two tables are created for each type of the parsing. Each table contains the number of items in the request sequence and the number of items in the list sequences that are generated for each file. The cost of FC and MRUFC are computed and recorded for each file as shown in table.

3.3 Experimental Results

Let |LS| be the size of LS and |RS| be the size of RS. L-R ratio is the ratio between |LS| and |RS|. Let C(FC) and C(MRUFC) be the access cost of FC and MRUFC respectively. Then gain (G) is defined as follows.

G = {(C(FC) - C(MRUFC)) / C(FC)} * 100. Each of the table generated is given below:

SL NO.	File name	|LS|	|RS|	L-R ratio	C(FC)	C(MRUFC)	G
1	bib	81	111261	1373	1629147	1629087	0.0036829
2	book1	82	768771	9375	7083638	7082759	0.0124089
3	book2	96	610856	6363	6735481	6733270	0.0328262
4	geo	256	102400	400	3884942	3880969	0.1022666
5	news	98	377109	3848	5610662	5608535	0.0379100
6	obj1	256	21504	84	841982	830464	1.3679627
7	obj2	256	246814	964	9099891	9072157	0.3047729
8	paper1	95	53161	559	669590	668107	0.2214788
9	paper2	91	82199	903	796112	795241	0.1094067
10	paper3	84	46526	553	470042	469226	0.1736015
11	paper4	80	13286	166	139755	139164	0.4228829
12	paper5	91	11954	131	149921	148814	0.7383889
13	paper6	93	38105	409	495080	493530	0.3130807
14	pic	159	513216	3227	1431185	1430511	0.0470938
15	progc	92	39611	430	612530	610859	0.2728030
16	progl	87	71646	823	792346	791547	0.1008398
17	progp	89	49379	554	639502	638267	0.1931190

Fig. 2. Byte Parsing Table

SL NO.	File name	|LS|	|RS|	L-R ratio	C(FC)	C(MRUFC)	G
1	bib	4282	20020	4.675	12639329	12595190	0.349219
2	book1	21075	142176	6.746	339256782	337730106	0.450006
3	book2	14974	103389	6.904	189641620	186466720	1.674157
4	geo	13423	37501	2.794	111530897	111423175	0.096584
5	news	14974	66411	4.435	158270466	156886302	0.874556
6	obj1	1760	8908	5.061	1844411	1833979	0.565601
7	obj2	11370	72981	6.419	90372271	88870636	1.661610
8	paper1	2537	8853	3.489	4868522	4800301	1.401267
9	paper2	3298	13912	4.218	8454144	8365525	1.048231
10	paper3	2602	7255	2.788	4545714	4513738	0.703432
11	paper4	873	2253	2.581	526058	521081	0.946093
12	paper5	816	2224	2.725	469264	464271	1.064007
13	paper6	1782	7345	4.122	2617529	2568258	1.882348
14	pic	4903	466140	95.072	14416540	14388939	0.191454
15	progc	1937	9656	4.985	2727389	2689352	1.394631
16	progl	2140	16900	7.897	4166116	4032726	3.201783
17	progp	1530	13182	8.616	1653465	1631909	1.303686

Fig. 3. Word Parsing Table

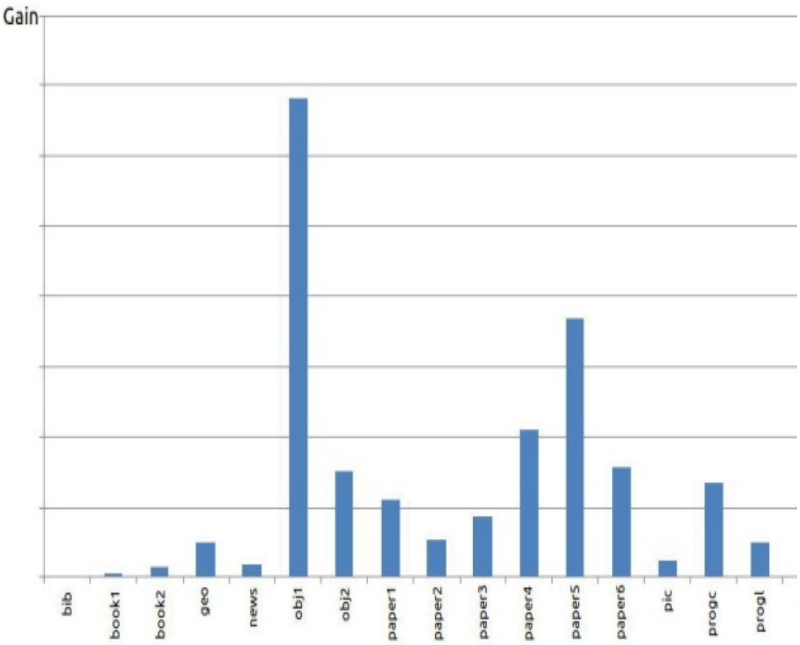


Fig. 4. Byte Parsing Histogram

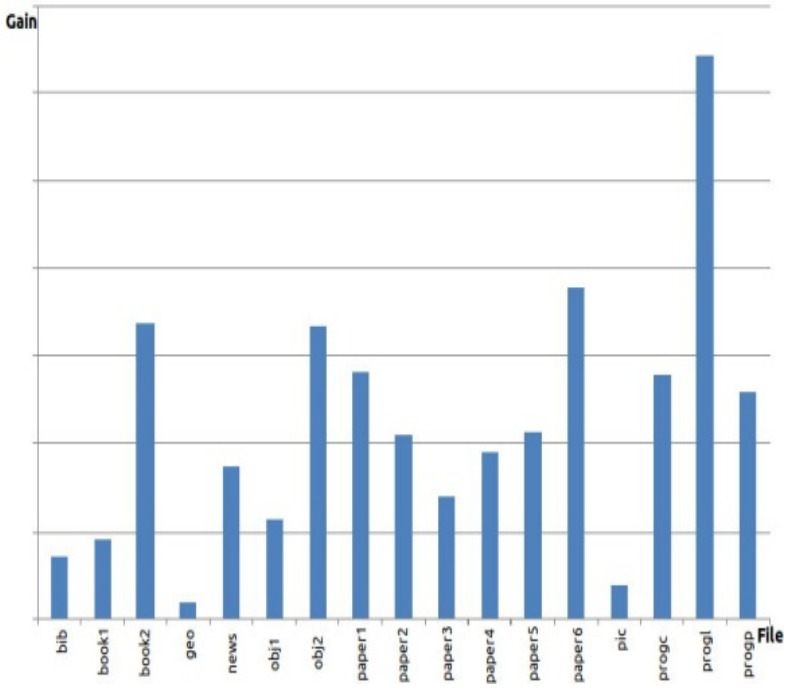


Fig. 5. Word Parsing Histogram

A representation of the above results has been done graphically using a histogram. The two histograms for both type of parsing represent the percentage gain in cost reduction for each file. The gain is a measure of how much MRUFC is better than FC algorithm.

It can be noticed by looking at the two histograms that average gain is more in word parsing than byte parsing. This is because of the L-R ratio associated with each parsing.

3.4 Observations

From our experimental results we have observed that the L-R ratio is greater for byte parsing table than word parsing table. These ratios can be used to identify the significance of input request sequences. For example, the (word) sequences generated from obj1 and obj2 are of minor interest since on an average almost every word in the request sequence is new. According to this measure we would expect the four word-level sequences corresponding to the files progp, progl, book2 and book1 to be of greater significance than the rest of the word-level sequences. In general, as can be seen in the tables, the two kinds of request sequences (word- and byte-based) are considerably different, with the word-based sequences resulting in very long lists and relatively short sequences (compared to the list length). On the other hand, the byte-based sequences correspond to very short lists and very long sequences. The access cost for MRUFC is observed to be less than the access cost of FC algorithm in all the input files. The gain(G) is found to be more for input files in the data set having higher L-R ratio than the input files having lower L-R ratio. Therefore, it can be inferred that for quite large value of L-R ratio gain will be increased, as a result of which performance of MRUFC will be significantly better than FC.

4 Concluding Remarks

In this paper we have proposed a new variant of the FC algorithm by using the Most Recently Used (MRU) concept. We have experimentally evaluated the performances of our proposed MRUFC algorithm by using the Calgary Corpus as input data set. We have also evaluated the performance of FC algorithm using the same data set. Our experimental results obtained with the Calgary Corpus as the input data set have shown that our proposed MRUFC algorithm performs better than FC algorithm. Based on the experimental results and our intuition, it can be inferred that for all the request sequences the access cost of MRUFC will be either same or less than the access cost of FC. It is a challenging research issue to conduct a theoretical analysis to validate the above fact.

References

[1] McCabe, J.: On serial files with relocatable records. Oper. Res. 12, 609–618 (1965)
[2] Sleator, D.D., Tarjan, R.E.: Amortized efficiency of list update and paging rules. Commun. ACM 28(2), 202–208 (1985)

[3] Irani, S.: Two results on the list update problem. Information Processing Letters 38, 301–306 (1991)
[4] Albers, S.: Improved randomized on-line algorithms for the list update problem. In: SODA 1995: Proceedings of the Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, Philadelphia, PA, USA. Society for Industrial and Applied Mathematics, pp. 412–419 (1995)
[5] Albers, S., von Stengel, B., Werchner, R.: A combined BIT and TIMESTAMP algorithm for the list update problem. Information Processing Letters 56(3), 135–139 (1995)
[6] Teia, B.: A lower bound for randomized list update algorithms. Information Processing Letters 47, 5–9 (1993)
[7] Hester, J.H., Hirschberg, D.S.: Self-organizing linear search. ACM Comput. Surveys, 295–312 (1985)
[8] Albers, S., Westbrook, J.: Self Organizing Data Structures. In: Fiat, A. (ed.) Online Algorithms 1996. LNCS, vol. 1442, pp. 31–51. Springer, Heidelberg (1998)
[9] Mohanty, R., Narayanaswamy, N.S.: Online Algorithms for Self Organizing Sequential Search - A Survey. Electronic Colloquium on Computational Complexity, Report No. 97, 1–13 (2009)
[10] Tenenbaum, A.: Simulations of dynamic sequential search algorithms. Commun. of the ACM 21(9), 790–791 (1978)
[11] Bentley, J.L., McGeoch, C.C.: Amortized analysis of self organizing sequential search heuristics. CACM 28, 404–411 (1985)
[12] Bachrach, R., El-Yaniv, R.: Online list accessing algorithms and their applications: recent empirical evidence. In: SODA 1997: Proceedings of the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 53–62. Society for Industrial and Applied Mathematics, Philadelphia (1997)
[13] Bachrach, R., El-Yaniv, R., Reinstadtler, M.: On the competitive theory and practice of online list accessing algorithms. Algorithmica 32(2), 201–245 (2002)
[14] Angelopoulos, S., Dorrigiv, R., López-Ortiz, A.: List Update with Locality of Reference. In: Laber, E.S., Bornstein, C., Nogueira, L.T., Faria, L. (eds.) LATIN 2008. LNCS, vol. 4957, pp. 399–410. Springer, Heidelberg (2008)
[15] Albers, S., Lauer, S.: On List Update with Locality of Reference. In: Aceto, L., Damgård, I., Goldberg, L.A., Halldórsson, M.M., Ingólfsdóttir, A., Walukiewicz, I. (eds.) ICALP 2008, Part I. LNCS, vol. 5125, pp. 96–107. Springer, Heidelberg (2008)

Automatic Test Case Generation Using Sequence Diagram

Vikas Panthi and Durga Prasad Mohapatra

Computer Science & Engineering Department,
National Institute of Technology, Rourkela
vpanthi@gmail.com, durga@nitrkl.ac.in

Abstract. Software Testing plays an important role in Software development because it can minimize the development cost. We Propose a Technique for Test Sequence Generation using UML Model Sequence Diagram.UML models give a lot of information that should not be ignored in testing. In This paper main features extract from Sequence Diagram after that we can write the Java Source code for that Features According to ModelJunit Library. ModelJUnit is an extended library of Junit Library. By using that Source code we can Generate Test Case Automatic and Test Coverage. This paper describes a systematic Test Case Generation Technique performed on model based testing (MBT) approaches By Using Sequence Diagram.

1 Introduction

UML models are intended to help to reduce the complexity of a problem, with the increase in product sizes and complexities. Still, the UML models themselves become large and complex involving thousands of interactions across hundreds of objects [1]. It is cumbersome for generating test models like control flow graph from source code. The UML sequence diagrams are used for modelling discrete behaviour of an object through sequence graph. Such states and transitions are critical to decide the specific operation invocations that would be made based on the conditions arising during a scenario execution. For unit level testing, we can derive tests from UML state machine diagrams, which embody the behavioural description of each component [2]. The information about a system is distributed across several model views of a system, captured through a large number of diagrams.

2 Some Basic Definitions

We first provide some basic definitions of relevant test coverage criteria. After that we will defined our proposed approach to the generation of test cases.

D1*:*Test case: A test case is the triplet [I, S, O], where I is the initial state of the system at which the test data is input, S is the test data input to the system and O is the expected output of the system [from our paper]. The output produced by the execution of the software with a particular test case provides a specification of the actual software behavior.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 277–284.
springerlink.com

D2:**Sequence Diagram:** A sequence diagram is a tuple (L, O, E, M, <, $R_{o,l}$, $R_{o,e}$, R_o,m) where L is a set of lifelines, O is a set of Occurrence Specifications, E is a set of Execution Specifications and M is a set of messages, < is a total ordering on O, $R_{o,l}$ is a relation- ship between O and L indicating lifelines covered by Occurrence Specifications, $R_{o,e}$ is a relationship between O and E indicating initial and terminal Occurrence Specifications of every Execution Specification, $R_{o,m}$ is a relationship between O and M indicating end points of every message.

D3*:* **Sequence graph:** A Sequence diagram can be viewed as a graph called a Sequence graph G= (N, T), where N is the set of nodes (vertices) of G and T is the set of edges or massage. In G, nodes of edges. In *G*, nodes represent object and edges represent massage between object. Since every node of a Sequence graph represents an Object, we shall use the terms 'node' and 'Object' interchangeably when no confusion arises. Without any loss of generality.

D4*:* **Path:** The number of predecessors of a node is its in-degree, and the number of successors of the node is its out-degree. A path from a node x_1 to a node x_k in a graph G = (V, E) is a sequence of nodes (x_1, x_2..., x_k) such that (x_i, x_{i+1}) ∈ E for every i, $1 \leq i \leq k-1$.

D5*:* ***Extended Finite State Machine (EFSM)*****:** An Extended Finite State Machine (EFSM) is defined as a 7 tuples
M= (I, O, S, D, F, U, T) Where I= set of input symbols. O= set of output symbols.

Some basic coverage Criteria. In this section, we discuss some of the relevant coverage criteria which are used in our approach.

1) *State Coverage*: It covers every state in every state chart for basic test generation. State coverage is a test adequacy criterion that requires tests to check programs' output variables [6]. All variables still defined when executing in test scope (even those which are not visible, such as private fields of objects) are considered by state coverage.
2) Message path coverage: A test set *TS* is said to achieve transition path coverage if given a state machine graph *G*, *TS* causes each possible transition path in *G* to be taken at least once [3].

3 ATGSD: Our Proposed Approach to Generate Test Cases

In this section we, discuss our proposed approach Automatically Test Sequence Generation from Sequence Diagram *(*ATGSD*)*. Our approach for generating test cases is schematically shown in Figure 1. The first step is constructing the Sequence diagram. The next step is to convert the Sequence diagram into Sequence graph. Then, the graph is traversed to select the predicate functions. In fourth step, we transform the predicate into source code.

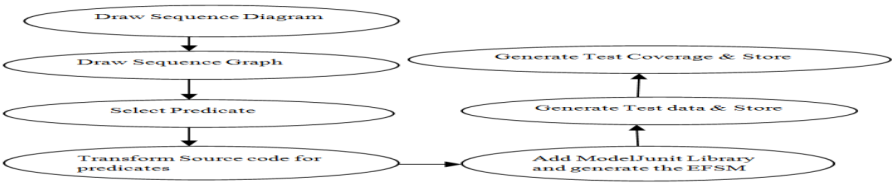


Fig. 1. Te**Fig. .** st Case Generation Process

Then, we construct the Extended Finite State Machine (EFSM) from the code. Finally, we generate the test data corresponding to the transformed predicate functions and store the generated test data for future use.

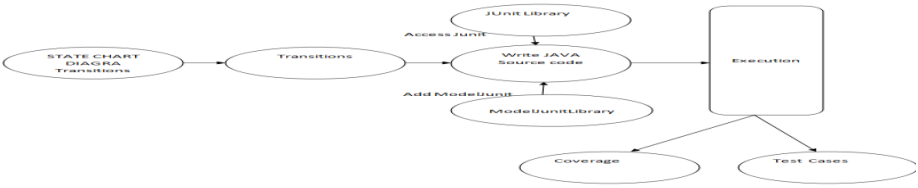


Fig. 2. Architecture of ModelJunit

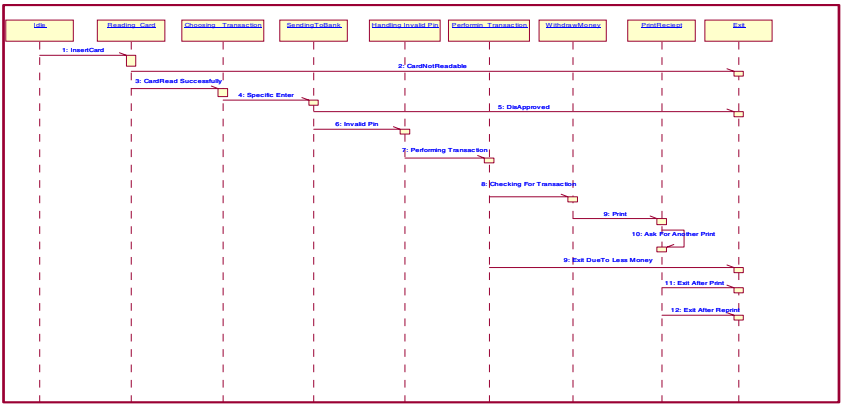


Fig. 3. Sequence Diagram of Bank ATM System

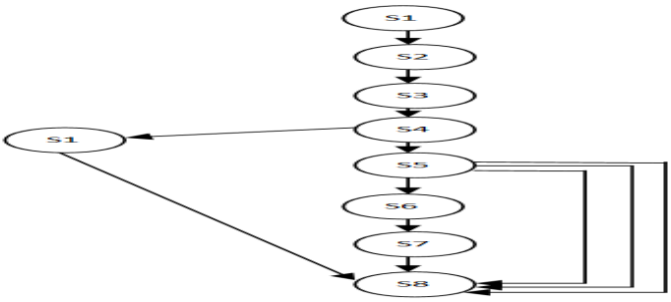


Fig. 4. Sequence Graph for Bank ATM Sequence Diagram

4 Pseudo Code of ATGSD Algorithm for Bank ATM System

Input: *Sequence Graph, Pin, Your_Balance, Withdraw_Money, Card_read, CustomerWantsAnotherPrint, Print ()}*
Output*: TSi(Test Sequence), SC (State Coverage),*
TC (Transition Coverage), ACC (Action Coverage), TPC (Transition Pair Coverage), EFSM Graph
Begin
State enum {Idle, Reading_Card, CardReadSuccessfully, Reading_Pin, PinReadSuccessfully, Choosing_Transaction, SendingToBank, HandlingInvalidPin, Performing_Transaction, WithdrawMoney, LessMoneyExit, PrintingReciept, AskForAnotherPrint, Exit}

```
If (state=Idle) then
    Print (TSi, Current State, Final State)
      State← Reading_Card
End if
If (state= Reading_Card AND card_read = false) then
    Print (TSi, Current State, Final State)
    Print ("Card Not Readable Please Check")
   State ← Exit
End if
If (state= Reading_Card AND card_read = true) then
     Print (TSi, Current State, Final State)
       State ← Choosing_Transaction
End if
If (state= Choosing_Transaction AND card_read=true) then
     Print (TSi, Current State, Final State)
        State ← SendingToBank
End if
If (state= SendingToBank AND pin !=1234) then
     Print (TSi, Current State, Final State)
      state← HandlingInvalidPin
End if
```

```
End if
    If (state= SendingToBank AND pin !=1234 ) then
        Print (TSi, Current State, Final State)
        Print("You are given invalid Pin No.");
        State ← Exit
    End if
  If (state= SendingToBank AND pin =1234) then
      Print (TSi, Current State, Final State)
      State ← Performing_Transaction
  End if
  If (state= HandlingInvalidPin AND pin =1234)) then
      Print (TSi, Current State, Final State)
        State ← Performing_Transaction
  End if
 If (state =Performing_Transaction AND pin =1234 AND card_read=true AND
Your_Balance >=withdraw_Money AND withdraw_Money <= max_limit_money)
then
  If((withdraw_Money %100)= null)
      Print("Withdraw Money")
      Print("After Withdraw Your Balance in Account")
      State ← withdraw_Money
      Print (TSi, Current State, Final State)
 Else
      Print("please give withdraw money multiple of 100")
      Print (TSi, Current State, Final State)
       state ←Exit
   End if
End if
     If (state= Performing_Transaction AND pin =1234 AND card_read=true AND
Your_Balance >=withdraw_Money AND withdraw_Money > max_limit_money) then
          Print("Your maximum limit in one transaction is over");
          Print (TSi, Current State, Final State)
            state=Exit;
 If (state = withdraw_Money AND Print=true) then
     If(Print = true)
         Print (TSi, Current State, Final State)
          state←PrintReciept
 End if
   End if
 If (state= Performing_Transaction AND pin =1234 AND card_read=true AND
Your_Balance < withdraw_Money AND withdraw_Money <= max_limit_money)
then
   Print("You have insufficient balance for withdraw Money");
        Print (TSi, Current State, Final State)
          state← Exit
 End if  Endif
```

5 Working of ATGSD for with Bank ATM System

In this Section, we are explaining the working of our ATGSD algorithm using Bank ATM example.

The Bank ATM is a Money dispenser Machine in which we can withdraw Money from machine. The sequence diagram of a Bank ATM object for various events of interest is shown in figure 3.

The objects first enter into *idle state*, after those objects insert the ATM card. After that machine will enter into *ReadingCard State* which *read* the card and store the information about customer for one transaction. If there is some problems for reading it will enter in *Exit State.* If machine haven't any problem for reading the card then it will enter into next state *Transaction State*. After that all the information of customer send to bank in *SendToBank State* because by the using this state all the personal information about customer will be secure. After that customer insert his Pin and Object will go to *Performing Transaction State*. If Pin is not match with original Pin then object enters to Exit State due to invalid Pin number. After that if Pin is match to original Pin. Then will machine display Amount Window for Customer in this state customer will have condition for withdrawing money.

1) WihtdrawMoney = 100 × n
   Means customer can withdraw money multiple of 100.
2) WithdrawMoney <= 40000
3) WithdrawMoney <=Your_Balance

If any condition will false then object can't withdraw money and go to *Exit State*

6 An Implementation of Our Approach

ModelJUnit Library allows us to write simple Sequence diagram as Java classes, then generate tests from those models and measure various model coverage metrics as well as EFSM. Model-based testing allows us to automatically generate test suites from a model of a system under test. ModelJUnit is a Java library that extends JUnit to support model-based testing [6]. The set of test cases generated corresponding to our *ATGSD* algorithm with the test coverage achieved is shown in Figure 4. In Figure 5, the initial node, the last node and the test data corresponding to each predicate are also shown. The percentage of test coverage which is achieved by implementing the case study of Bank ATM object is shown in the Table 1.

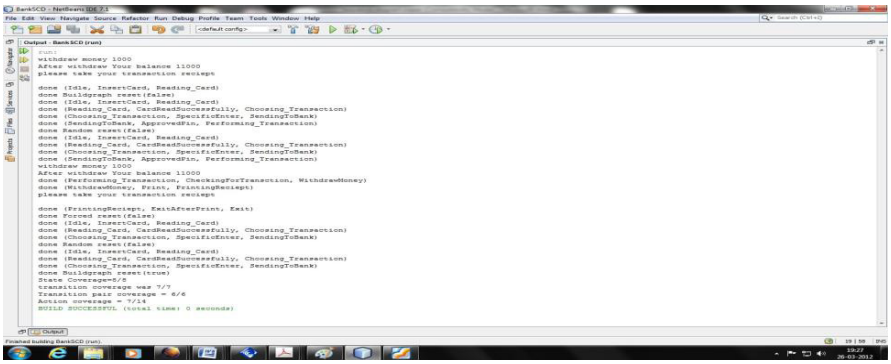


Fig. 5. Screenshot of generated test data with test coverage

Table 1. Test Coverage (YB= Your_Balance, WM = Withdraw_Money, CR = Card_read, CWA= customerWantsAnother, NS:No.of States, NT: No.of Transitions, SCP:% of State Coverage , TCP:% of Transition Coverage, TPCP: % of Transition Pair Coverage, AC: % of Action Coverage)

Sl NO.	PIN	YB	WM	CR	Print	CWA	NS	NT	SCP	TCP	TPCP	AC
1	1234	15000	5000	T	T	T	8	8	100%	100%	88.8%	57.1%
2	1235	12000	2000	T	T	T	6	5	100%	100%	100%	33.3%
3	0000	11000	4000	T	T	T	6	5	100%	100%	100%	33.3%
4	1234	10000	11000	T	T	T	6	5	100%	100%	100%	33.3%
5	1234	9000	9000	T	T	T	8	8	100%	100%	88.9%	53.4%
6	-1234	5000	4000	T	T	T	6	5	100%	100%	100%	33.3%
7	1234	20000	15000	F	T	T	3	2	100%	100%	100%	13.4%
8	1234	90000	15000	T	T	F	6	5	100%	100%	100%	33.4%
9	1234	60000	12000	F	T	F	3	2	100%	100%	100%	13.4%
10	1234	50000	10000	T	F	T	8	7	100%	100%	100%	46.7%
11	1234	50000	12102	T	T	F	6	5	100%	100%	100%	33.4%
14	1234	50000	40000	T	T	T	6	5	100%	100%	100%	33.4%

7 Conclusion

We have proposed a methodology to generate test cases from UML sequence diagrams. Our technique achieves many important coverage like state coverage, transition coverage, and action coverage. Our future work is to generate test cases using other UML diagrams and combination diagrams.

References

[1] Abdurazik, A., Offutt, J.: Using UML Collaboration Diagrams for Static Checking and Test Generation. In: Evans, A., Caskurlu, B., Selic, B. (eds.) UML 2000. LNCS, vol. 1939, pp. 383–395. Springer, Heidelberg (2000)

[2] Blanco, R., Fanjul, J.G., Tuya, J.: Test case generation for transition-pair coverage using Scatter Searc. International Journal of Software Engineering and Its Applications 4(4) (October 2010)

[3] Lorenzoli, D., Mariani, L., Pezzè, M.: Automatic generation of soft-ware behavioral models. In: Proc. 30th Int'l. Conf. on Softw. Eng (ICSE 2008), Leipzig, Germany, pp. 501–510 (May 2008)

[4] Fraikin, F., Leonhardt, T.: SeDiTeC–Testing Based on Sequence Diagrams. In: ASE 2002 (2002)
[5] http://www.sequencediagrameditor.com/uml/sequence-diagram.htm
[6] http://www.cs.waikato.ac.nz/~marku/mbt/modeljunit
[7] Koster, K., Kao, D.C.: State coverage: A structural test adequacy criterion for be-havior checking. In: ESEC/FSE (2007)

Normalized Wavelet Hybrid Feature for Consonant Classification in Noisy Environments

T.M. Thasleema and N.K. Narayanan

Department of Information Technology
Kannur University
thasnitm1@hotmail.com, nknarayanan@gmail.com

Abstract. This paper investigates on the use of Wavelet Transform (WT) to model and recognize the utterances of Consonant – Vowel (CV) speech units in noisy environments. The peculiarity of the proposed method lies in the fact that using WT, non stationary nature of the speech signal can be accurately considered. A hybrid feature extraction namely Normalized Wavelet Hybrid Feature (NWHF) using the combination of Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition (WPD) along with z-score normalization technique are studied here. CV speech unit recognition tasks performed for noisy speech units using Artificial Neural Network (ANN) and k – Nearest Neighborhood (k – NN) are also presented. The result indicates the robustness of the proposed technique based on WT in additive noisy condition.

1 Introduction

Speech recognition research has a history of more than 50 years. With the advancement of powerful computers and robust algorithms, Automatic Speech Recognition (ASR) has gone through a great amount of develop over the last few years. Generally a speech recognition system tries to identify the basic unit in any language, phonemes or words which can be compiled into text [1]. The present research work is motivated by the knowledge that only little attempts were rendered for the automatic speech recognition of CV speech unit in Indian languages like Hindi, Tamil, Bengali, Marathi Chinese etc and very less works have been found to be reported in the literature on Malayalam CV speech unit recognition, which is the principal language of South Indian state of Kerala. Very few research attempts were reported so far in the area of Malayalam vowel recognition. So more basic research works are essential in the area of Malayalam CV speech unit recognition.

Malayalam is one of the major languages from Dravidian language family. Malayalam is the principal language of the South Indian state of Kerala and also of the Lakshadweep Islands off the west coast of India spoken by about 36 million peoples [2]. Malayalam language now contains 51 V/CV units includes 15 long and short vowel sounds and the remaining 36 basic consonant sounds. For the present work, all the experiments are carried out using 36 Malayalam CV speech unit database uttered by 96 different speakers. For the recognition experiments, database is divided into five different phonetic classes based on the manner of articulation of the consonants as given in table 1.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 285–290.
springerlink.com

Table 1. Malayalam CV unit classes

Class	Sounds
Unaspirated	/ka/,/ga/,/cha/,/ja/,/ta/,/da/,/tha/,/d_ha/,/pa/,/ba/
Aspirated	/kha/,/gha/,/chcha/,/jha/,/tta/,/dda/,/ththa/,/dha/, /pha/, /bha/
Nasals	/nga/,/na/,/nna/,/na/,/ma/
Approximants	/ya/,/zha/,/va/,/lha/,/la/
Fricatives	/sha/,/shsha/,/sa/,/ha/,/ra/,/rha/

Since human speech is highly dynamic in nature, in order to achieve a reliable representation of the speech signal in the time – frequency plane a multi resolution approach is needed. Wavelet Transform (WT) is a tool for Multi Resolution Analysis (MRA) which can be used to efficiently represent the speech signal in the time – frequency plane. There have been lots of works reported in the literature using WT for the feature extraction process [3][4][5].The objective of the present work is to model Malayalam CV speech unit waveforms using WT based Normalized Wavelet Hybrid Feature (NWHF) extraction technique in speaker independent environments. Classifications are carried out using Artificial Neural Networks (ANN) and k – Nearest Neighborhood (k-NN) . The performance of the present method is discussed in noisy environments.

2 Wavelet Transform

Certain ideas of wavelet theory appeared quite a long time ago [6][7]. Wavelet transform can be defined as the transformation of the signal under analysis into another representation which presents the signal in a more useful form [8]. The Discrete Wavelet Transform (DWT) has been treated as a Natural Wavelet Transform (NWT) for discrete time signals by different authors [9][10]. For computing the wavelet coefficients several discrete algorithms have been established [11]. As Daubechies mentioned in his work DWT can be interpreted as a discretization of Continuous Wavelet Transform (CWT) through sampling specific wavelet parameters. In the present work we utilize the characteristics of wavelet transform using two major wavelet decomposition techniques namely Classical Wavelet decomposition (CWD) and Wavelet Packet Decomposition for CV speech unit recognition. CWD and WPD based approximation signal Plot of fifth level decomposition for the speech sound /ka/ is plotted in figure 1.

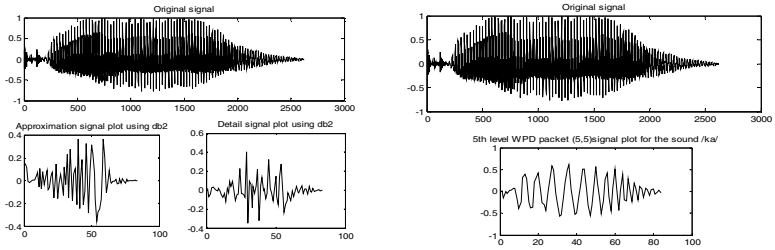


Fig. 1. CWD and WPD plot for the sound /ka/

3 Normalized Wavelet Hybrid Feature Extraction

Normalized Wavelet Hybrid Feature (NWHF) vector for the present work is generated using CWD and WPD method for representing CV speech unit recognition. The process for extracting NWHF feature vector is described below.

In the first step, the sound signal is made to undergo recursively to decompose into k^{th} level of resolutions; therefore the approximation coefficient matrix at this level is a sufficiently small representative of the original sound signal and carries enough information content to describe sounds characteristics coarsely. Let A_k represents this approximation matrix at decomposition level k, which can be written as

$$A_k = \begin{bmatrix} A_{1,1} & A_{1,2} & .... & A_{1,n} \\ A_{2,1} & A_{2,2} & .... & A_{2,n} \\ ... & ... & .... & .... \\ A_{k,1} & A_{k,2} & .... & A_{k,n} \end{bmatrix} \tag{1}$$

Then, the first component of NWHF feature vector $\boldsymbol{v_1}$ is,

$$v_1 = \bigcup_{i=1}^{k} \bigcup_{j=1}^{n} \{A_{i,j}\} \tag{2}$$

In the second step, for Wavelet Packet Decomposition, decompose each sound using k^{th} level of resolutions for the best level of wavelet packet decomposition tree. The first coefficient matrix at the best level tree contains enough information to represent the given input consonant CV speech unit without loss of much speech features. Let *m* represent mean of one row vector in the coefficient matrix then the WPD feature vector $\boldsymbol{v_2}$ is given by

$$v_2 = m_i \text{ , i = 1,2,.....m} \tag{3}$$

Where m is the number of rows in the best level coefficient matrix

In the third step we combined $\boldsymbol{v_1}$ and v_2 to fusion CWD and WPD coefficients.

$$V = \bigcup_{i=1}^{2} \{v_i\} \tag{4}$$

Then the final feature vector ***F*** after z-score normalization is given by

$$F = \frac{V - \mu(V)}{\sigma(V)} \tag{5}$$

Where μ and σ represents mean and variances with respect to the feature vector ***V.***

The feature vector of size 20 is estimated from NWHF vectors. The NWHF for different speaker shows the identity of the same sound so that an efficient feature vector can be formed using the proposed feature vectors. The graph obtained for different sounds seems to be distinguishable.

4 Classification Using Artificial Neural Network and k – Nearest Neighbor Classifiers

Pattern recognition can be defined as a field concerned with machine recognition of meaningful regularities in noisy or complex environments [12]. Nowadays pattern recognition is an integral part of most intelligent systems built for decision making. In the present study two widely used approaches for pattern recognition problems namely statistical pattern classifier (k – NN) and connectionist approaches (ANN) are applied.

Using k – NN, trial and error technique is applied to get better recognition accuracy and is obtained for the value of k=7.

Present work investigates the recognition capabilities of the Feed Forward Multi Layer Perceptron (FFMLP) based Malayalam consonant recognition system using Multi Layer Feed Forward Neural Network (MLFFNN) and Back Propagation (BP) algorithm. The number nodes in the input layer are fixed to 20 according to NWHF vector size. The number of nodes in the output layer is 36 for 36 Malayalam consonants. The experiment is repeated by changing the number of hidden layers. After trial and error experiments the number of hidden layer is fixed as 8 and the number of epochs as 10,000 for obtaining the successful architecture for the present study. The simulation experiments and the results obtained using these two pattern recognition (ANN and k – NN) approaches are explained in the next section.

5 Simulation Experiment and Results

All the simulation experiments are carried out using Malayalam CV speech unit database, uttered by 96 different speakers. We used 8kHz samples speech signal which is low pass filtered to band limit to 4kHz. Then each speech signal is corrupted additive white Gaussian noise of different Signal to Noise Ratio (SNR) levels. A fourth order Daubechies (db4) wavelet is used for this work.The classification is conducted for 36 Malayalam CV speech unit using Malayalam CV speech database uttered by 96 different speakers. We divide the dataset into training and test set which contains first 48

samples for training and next 48 for testing. Thus training and test set contains total of 1728 samples each. The recognition accuracies obtained for Malayalam CV speech database for the five different phonetic classes at various SNR levels are tabulated in Table 2.

Table 2. Experimental results using 5 differerent phonetic classes

	Recognition Accuracy							
	Classifier	k – NN			ANN			
SNR in dB	0	3	10	20	0	3	10	20
Unaspirated	25.7	28.9	49.5	62.4	44.7	49.9	59.1	68.9
Aspirated	29.1	36.4	45. 9	60.7	46.9	50.3	63.3	74.2
Nasala	46.1	50.9	54.6	61.1	54.1	61.4	69.1	76.7
Approximants	41.7	49.3	55.2	64.1	50.4	57.6	67.3	79.2
Fricatives	43.3	49.1	56.5	63.6	48.5	52.6	64.5	72.2

Experimental results using NWHF feature vector implies that ANN can be considered to be a good classifier for Malayalam CV speech database compared with k – NN in additive noisy environments. Results indicate that the NWHF vectors are able to improve the recognition accuracies at low level of SNRs.

6 Conclusion

A Multi Resolution Analysis (MRA) approach to Malayalam Consonant – Vowel (CV) speech unit recognition using Wavelet Transform (WT) has been studied. Two decomposition algorithms namely Classical Wavelet Decomposition (CWD) and Wavelet Packet Decomposition are combined to extract Normalized Wavelet Hybrid Feature (NWHF) vector along with z-score normalization technique for the present classification study. The recognition accuracies are calculated and compared using Artificial Neural Network (ANN) and k – Nearest Neighborhood (k – NN) at different levels of Signal to Noise Ratio (SNR) values and it is observed that the NWHF parameters improve their recognition accuracies at low level of SNR. More effective implementation of wavelet features in combination with frequency domain features and the development of multiple classifiers would be some of our future research directions.

References

1. Gold, B., Morgan, N.: Speech and Audio Signal Processing. John Wiley & Sons Inc., New York (2000)
2. Ramachandran, H.P.: Encyclopedia of Language and Linguistics. Pergamon Press, Oxford

3. Bourlard, H., Dupont, S.: Subband-based speech recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP Signal Processing (ICASSP 1997), Munich, Germany, vol. 2, pp. 1251–1254 (April 1997)
4. Gupta, M., Gilbert, A.: Robust speech recognition using wavelet coefficient features. In: IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2001), Madonna di Campiglio, Trento, Italy, pp. 445–448 (December 2001)
5. Sarikaya, R., Pellom, B.L., Hansen, J.H.L.: Wavelet packet transform features with application to speaker Identification. In: 3rd IEEE Nordic Signal Processing Symposium (NORSIG 1998), Vigsø, Denmark, pp. 81–84 (June 1998)
6. Grossman, A., Morlet, J., Gaoupillaud, P.: Cycle octave and related transforms in seismic signal Analysis. Geoexploration 23, 85–102 (1984)
7. Mallat, S.: A wavelet Tour of Signal Processing, The Sparse Way. Academic, New York (2009)
8. Soman, K.P., Ramachandran, K.I.: Insight into Wavelets, from Theory to Practice. Prentice Hall of India (2005)
9. Vetterly, M., Herley, C.: Wavelets and Filter banks: Theory and Design. IEEE Trans. on Signal Processing 40(9), 2207–2232 (1992)
10. Shensa, M.J.: Affine Wavelets: Wedding the Atrous and mallat Algorithms. IEEE Trans. on Signal Processing 40, 2464–2482 (1992)
11. Mallat, S.: Multi frequency Channel Decomposition of Images and Wavelet Models. IEEE. Trans on Acoustics, Speech and Signal Processing 37, 2091–2110 (1989)
12. Duda, R.O., Hart, P.E.: Pattern Classification and Scene Analysis. Wiley Inter Science, New York (1973)

Cuckoo Search for Inverse Problems and Topology Optimization

Xin-She Yang[1] and Suash Deb[2]

[1] University of Cambridge, Department of Engineering, Cambridge CB2 1PZ, UK
Also: Xi'an Engineering University, School of Science, Xi'an, China
xy227@cam.ac.uk
[2] International Neural Network Society (INNS), India Regional Chapter, c/o Mahura, Janla, Bhubaneswar 752054, India
Suash.deb@gmail.com

Abstract. Many inverse problems in engineering can be considered as constrained optimization, while topology optimization is usually very challenging due to its intrinsic link with inverse problems. Under nonlinear complex constraints, it is very time-consuming to solve such topology optimization problems. Now we use cuckoo search algorithm to carry out topology optimization, and results show that distribution of different topological characteristics can be achieved efficiently.

1 Introduction to Inverse Problems

The primary aim of an inverse problem is to estimate important parameters of structures and materials, given observed data which are often incomplete, so that the differences between observations and predictions can be minimize. To improve the quality of the estimates, we have to combine a wide range of known information, including any prior knowledge of the structures, available data etc. To incorporate all useful information and carry out the minimization, we have to deal with a multi-objective optimization problem. In the simplest case, we have to deal with a nonlinear least-squares problem with complex constraints [1, 6].

Under various complex constraints, we have to deal with a nonlinear, constrained, global optimization problem. In principle, we can then solve the formulated constrained problem by any efficient optimization techniques [3, 11, 12, 9]. However, as the degree of freedom in inversion is typically large, data are incomplete, and non-unique solutions or multiple solutions may exist; therefore, meteheuristic algorithms are particularly suitable.

Therefore, in this paper, we will first outline the basic formulation of inversion and also outline the fundamental ideas of cuckoo search (CS), and we will the use the cuckoo search to solve a case study in heat transfer. We also discuss some the implications of the proposed approach.

For an inverse problem with unknown parameters q and observed data $d_i (i = 1, ..., m)$, we have a generalized least-squares problem [6, 7, 8]

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 291–295.
springerlink.com

$$\min f = \| d - \phi(x, q) \|^2, \tag{1}$$

or

$$\min \sum_i [d_i - \phi(x_i, q)]^2, \tag{2}$$

which is equivalent to a nonlinear, constrained optimization problem:

$$\min f(x, q, d) \tag{3}$$

subject to

$$\begin{aligned} h_j(x, q) &= 0, \quad (j = 1, ..., J), \\ g_k(x, q) &\le 0, \quad (k = 1, ..., K). \end{aligned} \tag{4}$$

where J and K are the numbers of equality and inequality constraints, respectively. The main task now is to find an optimal solution to approximate the true parameter set **q***. In principle, such optimization can be solved using any efficient optimization algorithm. However, as the number of free parameters tends to be very large, and as the problem is nonlinear and possible multimodal, conventional algorithms such as hill-climbing usually do not work well. More sophisticated metaheuristic algorithms have the potential to provide better solution strategies [5, 11].

2 Cuckoo Search

Cuckoo search (CS) is one of the latest nature-inspired metaheuristic algorithms, developed in 2009 by Xin-She Yang and Suash Deb [13, 14]. CS is based on the brood parasitism of some cuckoo species. In addition, this algorithm is enhanced by the so-called Lévy flights, rather than by simple isotropic random walks. Recent studies show that CS is potentially far more efficient than PSO and genetic algorithms [14].

For simplicity in describing the Cuckoo Search, we now use the following three idealized rules: a) Each cuckoo lays one egg at a time, and dumps it in a randomly chosen nest; b) The best nests with high-quality eggs will be carried over to the next generations; c) The number of available host nests is fixed, and the egg laid by a cuckoo is discovered by the host bird with a probability p_a. In this case, the host bird can either get rid of the egg, or simply abandon the nest and build a completely new nest.

As a further approximation, this last assumption can be approximated by a fraction p_a of the host nests that are replaced by new nests (with new random solutions). For the implementation point of view, we can use the following simple representations that each egg in a nest represents a solution, and each cuckoo can lay only one egg (thus representing one solution), the aim is to use the new and potentially better solutions (cuckoos) to replace a not-so-good solution in the nests. Obviously, this algorithm can be extended to the more complicated case where each nest has multiple

eggs representing a set of solutions. For this present work, we will use the simplest approach where each nest has only a single egg. In this case, there is no distinction between an egg/a solution, a nest or a cuckoo, as each nest corresponds to one egg which also represents one cuckoo.

There are two key branches or types of generating new solutions in cuckoo search. Once type is to generate solutions by Lévy flights [13]

$$x_i^{(t+1)} = x_i^{(t)} + \alpha\, s_L, \tag{5}$$

where s_L is a vector drawn from the Lévy distribution

$$L(s,\lambda) = \frac{\lambda\Gamma(\lambda)\sin(\pi\lambda/2)}{\pi}\frac{1}{s^{1+\lambda}}, \quad (s \gg s_0), \tag{6}$$

where $s_0 > 0$ is the minimum step size and Γ is a Gamma function. Here $\alpha > 0$ is the step size scaling factor which should be related to the scales of the problem of interest. Here s is the step size drawn from a Lévy distribution.

The other branch of solution generation is that new solutions are generated by using the similarity between the existing eggs/solutions and the host eggs with a discovery rate p_a. This can be represented mathematically as

$$x_i^{t+1} = x_i^t + s \otimes H(p_a - \varepsilon) \otimes (x_j^t - x_k^t), \tag{7}$$

where x_i, x_j and x_k are three different solutions. $H(u)$ is a Heaviside function of u, and ε is a random number drawn from a uniform distribution in $[0,1]$. Again s is the step size vector.

3 Topology Optimization for Microdevice

Inverse problems and shape/topology optimization can occur in many applications. In the rest of this paper, we use the cuckoo search algorithm to two case studies. In the rest of the simulations, we have used $n = 20$, $\beta_0 = 1$, $\gamma = 1.5$. In addition, the total number of iterations is set to 1000.

Heat management, basically heat transfer modelling, is very important for many electronic device, especially those using large-scale integrated circuits. In fact, nanoscale heat transfer is an interesting area, and topological optimization for the design of a nanoscale device is even more challenging [15, 10, 2]. For example, Evgrafov et al. proposed a topology optimization benchmark for a nanoscale heat-conducting system with a size of 150 nm by 150 nm. Since heat transfer can occur at many different scales, though smaller scales may be more difficult to control. Now we extend this to a unit area of 1 mm by 1 mm, and the aim is to distribute two different materials so as to maximize the temperature difference $|T_A - T_B|$ at two points A and B under the boundary conditions given in [2]. Two materials used in the design

of the unit area have heat diffussivities of K_1 and K_2, respectively. In addition, $K_1 \gg K_2$. For example, Si and Mg_2Si, $K_1 / K_2 \approx 10$. The domain is continuous under boundary heat flux conditions and the objective is to distribute the two materials such that the difference $|T_A - T_B|$ is as large as possible.

By dividing the domain into 40×40 small grids and using CS to search the possible design solutions, an optimal shape and distribution of materials are show in Fig. 0 where Si is shown in blue and Mg_2Si is shown in red.

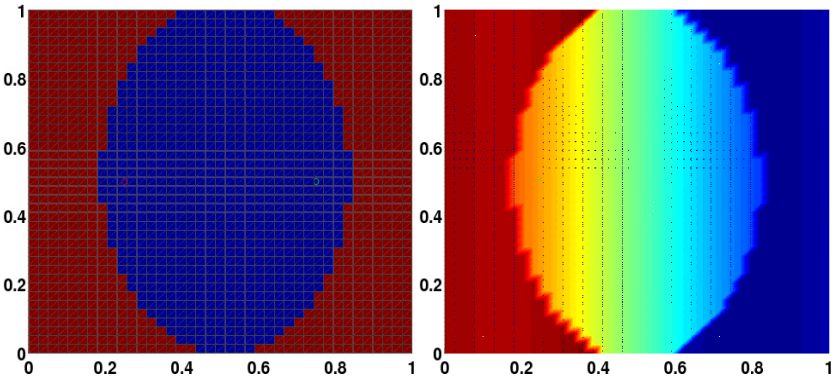


Fig. 1. Optimal topology and distribution of two different materials (left) and the temperature distribution (right)

For each configuration generated during the search process, the temperature distribution is estimated using the finite-difference method by solving the heat conduction equation with varied material conductivities so that the temperature difference at the two fixed points should be as large as possible. The final temperature distribution after 1000 iterations is shown in Fig. 1.

4 Summary

Topology optimization problems are often linked with inverse problems, while inverse problems can be very challenging due to nonlinearity in objectives and complex constraints as well as the large degrees of freedom. In this paper, we have highlighted that these problems can be in fact solved using an optimization framework and have thus demonstrated the effectiveness of this approach using cuckoo search algorithm.

It is worth pointing out that the problem must be well-posed so that unique solutions exist. Otherwise, whatever the solution techniques in use will not solve the non-uniqueness problem if the data are incomplete or the problem is not well-posed. Obviously, more studies should address how these issues may be approached in a feasible way so that better insight can be achieved.

References

1. Bendsøe, M.P.: Optimization of Structural Topology. Shape and Material. Springer (1995)
2. Evgrafov, A., Maute, K., Yang, R.G., Dunn, M.L.: Topology optimization for nano-scale heat transfer. Int. J. Num. Methods in Engrg. 77(2), 285–300 (2009)
3. Greenhalgh, S.A., Zhou, B., Green, A.: Solutions, algorithms and inter-relations for local minimization search geophysical inversion. J. Geophys. Eng. 3, 101–113 (2006)
4. Kar, C.L., Yakushin, I., Nicolosi, K.: Solving inverse initial-value, boundary-value problems via genetic algorithms. Engineering Applications of Artificial Intelligence 13, 625–633 (2000)
5. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proc. of IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942–1948 (1995)
6. Sambridge, M.: Geophysical inversion with a neighbourhood algorithm–I. Search a parameter space. Geophys. J. Int. 138, 479–494 (1999)
7. Sambridge, M., Mosegaard, K.: Monte Carlo methods in geophysical inverse problems. Reviews of Geophysics 40(3), 1–29 (2002)
8. Scales, J.A., Smith, M.L., Treitel, S.: Introductory Geophysical Inverse Theory. Samizdat Press (2001)
9. Talbi, E.G.: Metaheuristics: From Design to Implementation. John Wiley & Sons (2009)
10. Yang, X.S.: Modelling heat transfer of carbon nanotubes. Modelling Simul. Mater. Sci. Eng. 13, 893–902 (2005)
11. Yang, X.S.: Nature-Inspired Metaheuristic Algorithms. Lunver Press, UK (2008)
12. Yang, X.S.: Engineering Optimization: An Introduction with Metaheuristic Applications. John Wiley & Sons (2010)
13. Yang, X.S., Deb, S.: Cuckoo search via Lévy flights. In: Proc. of World Congress on Nature & Biologically Inspired Computing (NaBic 2009), pp. 210–214. IEEE Publications, USA (2009)
14. Yang, X.S., Deb, S.: Engineering optimization by cuckoo search. Int. J. Math. Modelling Num. Optimisation 1(4), 330–343 (2010)
15. Zhirnov, V.V., Cavin, R.K., Hutchby, J.A., Bourianoff, G.I.: Limits to binary logic switch scaling-a gedanken model. Proc. of the IEEE 91(11), 1934–1939 (2003)

A Lock Management Framework for a Class Hierarchy Tree

Arvind Mohan, Gaurav Singhal, and Bhaskar Biswas

Department of Computer Engineering, IT-BHU, Varanasi
arvind.mohan.cse08@itbhu.ac.in,
gaurav.singhal.cse08@itbhu.ac.in,
bhaskar.cse@itbhu.ac.in

Abstract. In case of a hierarchical system (say an N-ary object tree), critical sections consisting of node or nodes in the tree structure are inter-dependent due to the ancestor-successor relationship – depending on the position of the node in the hierarchy. In this paper, we present an approach to achieve higher concurrency and better performance in a multi-threaded system that has such a hierarchical object tree to deal with

1 Introduction

A perfect binary tree of height h will have minimum 2^h nodes. If h is a large number, the trade-off on the concurrency due to M waiting threads (where M approximately equal to h) is huge. A locking mechanism should not be a cause for indefinite waits on the object structure. A locking mechanism for such type of structure should provide highly concurrent transactions [1] or operations in a hierarchy, e.g. an XML document tree.

The reason for designing a new LockManager is to avoid granting sequential access to threads on a hierarchical structure; in other words, to increase parallelism [2]. Suppose that, on whole of the structure only one thread, regardless of its action (Read or Write) and the place in structure, wants access. This situation can be resolved at any instance of time by taking a lock on root of the tree and any incoming thread will have to check for the lock on the root and will go to wait until the lock is released.

The aim behind such type of strategy was to make sure that whole structure does not get deleted; hence we must keep this aim in our mind as we explore our new locking mechanism.

A possible solution can be distinguishing the threads according to their actions, so granting multiple Reads and exclusive Write on the structure. But this solution cannot be beneficial when multiple Read access and few write access are requested, because Write access needs to be exclusive. Thread asking for Write access will have to wait until all the threads acquiring Read access on the structure complete their tasks [7]. This will be a bad solution as thread asking for Write access may have to wait for infinite time if threads requesting for Read access keep on coming. Ultimately the Write thread will starve.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 297–302.
springerlink.com

We present an approach to achieve higher concurrency and better performance in a multi-threaded system that has such a hierarchical object tree to deal with. This system works through a lock handler/wrapper class that uses a mix of classical locks and reference counting to achieve better results when various operations are taking place at different levels/nodes of the object tree.

2 LockManager

A locking mechanism that allows different threads to have access on different nodes concurrently, is referred here as LockManager [3]. Access on a node given to any thread can be characterized as follows:

(i) Read Access: Getting the data stored in the node

(ii) Write Access: Modifying the data stored in the node [including deletion of the node itself]

In this mechanism multiple threads can read the data stored at a node simultaneously but can modify it only mutually exclusively. This is quite intuitive since reading the data stored in the node can be done by several threads simultaneously because they are not going to change properties of the node, unlike modifying the node [including deletion] cannot be done simultaneously by different threads [8]. Here we are proposing several locks in order to maintain properties of our hierarchical structure.
(i) RLock: This is a basic Read Lock and is granted whenever a thread wants to read the data stored in any node.

(ii) WLock: This is a basic Write Lock and is granted whenever a thread wants to write at any node at any level.

(iii) xRLock: xRLock at 'this' signifies that somewhere in the sub-tree, whose root is 'this' node, some thread has definitely acquired a Read lock. When a thread has to acquire read lock (RLock) on a node, expected read lock (xRLock) is provisioned for all the ancestors.

(iv) xWLock: xWLock at 'this' signifies that somewhere in the sub-tree, whose root is 'this' node, some thread has definitely acquired a Write lock. When a thread has to acquire write lock (WLock) on a node, expected write lock (xWLock) is provisioned for all the ancestors.

3 Properties of Nodes of Structure under Lockmanager

Integrity of a node (and its ancestors) is maintained while it is under an operation by the combination of read/write locks on the node and references on its ancestors [5]. To manage the lock modes and to increase the performance of Lock Manager described above we define these counts:

(i) ReadRef
Value of Readref at any node indicates the number of threads that have been granted Read access on the descendant nodes including 'this' node.

(ii) WriteRef
Value of Writeref at any node indicates the number of threads that have been granted Write access on the descendant nodes including 'this' node.

(iii) xRref
xRref is just the reference count of expected read lock at any node. It can be viewed as Boolean variable since at any node xRref can be TRUE (if there is xRLock) or FALSE (if there is no xRLock). It indicates that on any of the descendant node a Read access is granted to some thread.

(iv) xWref
xWref is also a reference count of expected write lock at any node. It can be viewed as Boolean variable since at any node xWref can be TRUE (if there is xWLock) or FALSE (if there is no xWLock). It indicates that on any of the descendant node a Write access is granted to some thread.

(v) RLock
All the above stated reference counts are general for any sub-tree but the node, at which a Read lock is granted, has to maintain some specific count in order to be distinguished. RLock is the count for a Read lock on a node. It will store the number of threads currently holding the read access, since multiple threads can acquire Read lock simultaneously.

(vi) WLock
Similar to RLock, WLock is specific to the node at which a Write lock is granted. It is the count for a Write lock on a node. Since Write access is granted to a thread exclusively hence value of WLock will be either 0(False) or 1(True).

4 Locking

4.1 Acquiring a Read Lock

Before the Read lock is granted, all the ancestors are traversed and checked whether there is a lock present on any of the ancestors or not [5].

Following algorithm can be used to implement Read Lock on any Node:

ReadLock (node NODE)

- Traverse from NODE to ROOT to check a READ/WRITE Lock on any node.
  1. If no READ/WRITE lock present
     a. Grant READ lock to NODE [Assign TRUE to RLock of NODE]
     b. Grant xR lock to all nodes in the path from NODE to ROOT before finding any xW/xR lock on a node or ROOT.
     c. If an xR/xW lock is found or ROOT is reached, just do nothing and terminate the checking.

i. Increase the Readref of every node on this path to the ROOT. [locking done successfully].

2. Else if READ lock is found
   a. Grant READ lock to NODE. [Assign TRUE to RLock of NODE]
   b. Increase the Readref of every node on this path to the ROOT. [No need to grant xR lock, locking done successfully].
3. Else if WRITE lock is found
   a. READ lock cannot be granted.

4.2 Acquiring Write Lock

Before the Write lock is granted, all the ancestors are traversed and checked whether there is a lock present on any of the ancestors or not [5].

Following algorithm can be used to implement Read Lock on any Node:

WriteLock (node NODE)

- If and only if, Node has no lock at all and an xRLock is not acquired on parent of NODE
- Traverse from NODE to ROOT to check a READ/WRITE Lock on any node.
  1. If no READ/WRITE lock is present,
     a. WRITE lock is granted to NODE. [Assign TRUE to WLock of NODE]
     b. Grant xW lock to all nodes in the path from NODE to ROOT before finding any xW/xR lock on the NODE or ROOT.
        i. If an xR/xW lock or ROOT is found just do nothing and terminate.
     c. Increase the Writeref of every node on this path to the ROOT. [locking done successfully].
  2. If Write Lock is present on an ancestor,
     a. Check whether the thread that acquired WRITE lock on ANCESTOR is the same that is asking for WRITE lock on NODE.
        i. If same, grant WRITE lock on NODE [No need to grant xW lock] [Assign TRUE to WLock of NODE] and increase the Writeref of every node on this path to the ROOT.
        ii. If not, Write Lock cannot be granted.
  3. If READ Lock is present on an ancestor,
     a. Write Lock cannot be granted.

5 Unlocking

One may wonder that why are we working with all these reference counts. Simple answer to this is, these counts are helpful while unlocking. Whenever a Read lock has to be unlocked at any node, the ReadRef at each node in the path from 'this' node to root is decremented by one and if ReadRef is zero at a particular node then xRref is decremented to zero because if there is no read lock in sub-tree, there is no need of an xRLock on this node.

UnlockRead (node NODE)

1. Release Read lock from NODE. [Assign FALSE to RLock of NODE]
2. Decrease Readref of all the nodes from NODE to ROOT by 1.
3. If at any node, say ANCESTOR of Node, value of Readref is 0, i.e. no READ lock on any descendant node,
    a. If the value of Writeref is 0, i.e. no WRITE lock on any descendant node, release xR lock. [Assign FALSE to xRref of ANCESTOR]
    b. If the value of Writeref is greater than 0, i.e. WRITE lock is/are acquired on any descendant node, release xR lock [Assign FALSE to xRref of ANCESTOR] and acquire xW lock [Assign TRUE to xWref of ANCESTOR]

Similarly, when a write lock has to be unlocked at any node then writeref at each node in the path to root is decremented by 1 and if writeref at a node is zero, xWref is also decremented to zero because if there is no Write lock in sub-tree there cannot be xWLock.

.

UnlockWrite (node NODE)

1. Release WRITE lock from NODE. [Assign FALSE to WLock of NODE]
2. Decrease Writeref of all the nodes from NODE to ROOT by 1.
3. If at any ancestor of NODE, value of Writeref is 0, i.e. no WRITE lock on any descendant node,
    a. If the value of Readref is also 0, i.e. no READ lock on any descendant node, release xW lock. [Assign FALSE to xWref of ANCESTOR]
    b. If the value of Readref is greater than 0, i.e. READ lock is/are acquired on any descendant node, release xW lock [Assign FALSE to xWref of ANCESTOR] and acquire xR lock [Assign TRUE to xRref of ANCESTOR]

6 Conclusion and Future Work

There should be a mechanism for checking the state of any node one more time before giving the access on that node in case of hierarchical structures of larger height. One of the approaches can be DCLP (Double Check Locking Pattern) which is under consideration and review [6].

References

[1] Bächle, S., Härder, T., Haustein, M.P.: Implementing and Optimizing Fine-Granular Lock Management for XML Document Trees. In: Zhou, X., Yokota, H., Deng, K., Liu, Q. (eds.) DASFAA 2009. LNCS, vol. 5463, pp. 631–645. Springer, Heidelberg (2009), doi:10.1007/978-3-642-00887-0_56

[2] Johnson, R., Prandis, I., Ailamaki, A.: Improving OLTP Scalability using Speculative Lock Inheritance. Journal Proceedings of the VLDB Endowment 2(1) (August 2009)

[3] Pandis, I., Jhonson, R., Hardavellas, N., Ailamaki, A.: Data-Oriented Transaction Execution. Journal Proceedings of the VLDB Endowment 3(1-2) (September 2010)
[4] Byun, C., Yun, I., Park, S.: A New Optimistic Concurrency Control in Valid XML. Journal of Information Science and Engineering 25(1), 11–31 (2009)
[5] Concurrency Control Part 2, http://inst.eecs.berkeley.edu/~cs186/fa06/lecs/20cc2.pdf
[6] Double Checked Locking and the singleton pattern, http://www.ibm.com/developerworks/java/library/j-dcl/index.html
[7] Readers-writer Lock, http://en.wikipedia.org/wiki/Readers-writer_lock
[8] Suggestions for multiple-reader/single-writer lock?

ECG Arrhythmia Classification Using R-Peak Based Segmentation, Binary Particle Swarm Optimization and Absolute Euclidean Classifier

Milan S. Shet, Minal Patel, Aakarsh Rao, Chethana Kantharaj, and Suma K.V.

Department of Electronics and Communication,
M.S. Ramaiah Institute of Technology, Bangalore-560054, India

Abstract. This paper proposes a novel technique to classify arrhythmias from ECG signals using time domain and frequency domain approaches. The ECG signal is pre-processed using Fast Fourier Transform (FFT). It is then segmented into beats after detecting the R-peaks. The Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) used for Feature Extraction pack most information in fewest coefficients. The Binary Particle Swarm Optimization (BPSO) algorithm used for Feature selection reduces dimensionality by selecting subset of original variables. The proposed Absolute Euclidean Classifier (AEC), which uses the absolute values of the features instead of their actual values, is found to improve the Classification Rate significantly. Feature Extraction using DCT/DWT and Feature Selection using BPSO, together with pre-segmentation process results in an improved Classification Rate and a reduced number of selected features for the proposed Arrhythmia Classification system. Experiments conducted on MIT-BIH Database show an enhanced performance as compared to other systems.

Keywords: Absolute Euclidean Classifier, Binary Particle Swarm Optimization, Arrhythmia Classification, Discrete Cosine Transform, Discrete Wavelet Transform, Segmentation.

1 Introduction

Arrhythmia is any disorder of the heart rate or rhythm. It disturbs simultaneous cardiac contraction sequences and reduces the cardiac pumping efficiency. The electrocardiogram (ECG or EKG) is a diagnostic tool that measures and records the electrical activity of the heart. Interpretation of these activities allows the analysis and diagnosis of a wide range of arrhythmias, which can vary from minor to life threatening. Classification of the ECG signal is achieved by finding the characteristic shapes of the ECG that discriminate effectively between the required diagnostic categories. Conventionally, a typical heart beat is identified from the ECG and the component waves of the QRS, P and T waves are characterized using measurements such as magnitude, duration and area.

The proposed system involves pre-processing the ECG signals for removal of noise, detection of peaks and segmentation of the signal into beats. Feature extraction and feature selection significantly reduce the amount of information to

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 303–317.
springerlink.com

represent an ECG signal, thereby reducing computational time and cost. In our experiments, both Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) were used for feature extraction, yielding competitive results in both cases. Further, feature selection using Binary Particle Swarm Optimization (BPSO) improves accuracy of the classifier and results in selection of more interpretable features. The proposed Absolute Euclidean Classifier (AEC), which uses the absolute values of the DCT and DWT coefficients, instead of their actual values, is found to improve the classification rate significantly.

Several methods have been proposed for the classification of ECG signals. Khazaee et al. [1] extracted power spectral density (PSD) features of each heart beat with three timing interval features classifying cardiac abnormalities in MIT-BIH database. The combining of wavelet domain feature with RR- interval features can achieve high classification accuracy as reported in [2]. In [3] a PSO-SVM based approach has been proposed for feature selection and classification of cardiac arrhythmias. The MIT-BIH Arrhythmia Database[4], the MIT-BIH Atrial Fibrillation Database[5] and the MIT-BIH Malignant Ventricular Arrhythmia Database[6] are used in our experiments for performance evaluation. Classification has been done for the following four classes : normal sinus rhythm (N), paced beat (P), atrial fibrillation (AF) and ventricular fibrillation (VF).

The rest of the paper is organized as follows: Section 2 describes the various techniques of ECG signal pre-processing, namely, noise removal, R-peak detection, histogram equalization and segmentation. Section 3 and Section 4 give an overview of the dimensionality reduction techniques: feature extraction and feature selection. The proposed Absolute Euclidean Classifier is discussed in Section 5. Section 6 discusses the proposed Arrhythmia Classification Systems and Experimental Results. Concluding remarks are given in Section 7.

2 Pre-processing of ECG Signal

The basic task of ECG pre-processing involves R-peak detection. There are some difficulties one can encounter in processing ECG: irregular distance between peaks, irregular peak form, presence of low-frequency noise components in ECG due to patient breathing, etc. To address these problems, the pre-processing pipeline should contain specific stages so as to reduce influence of these factors. Fig. 1 demonstrates such a pipeline. Stage 1 involves removal of low frequency components, Stage 2 is R-peak detection and Stage 3 is segmentation. All these stages use the ECG signals sampled at a rate of f_s = 1000 samples/sec.

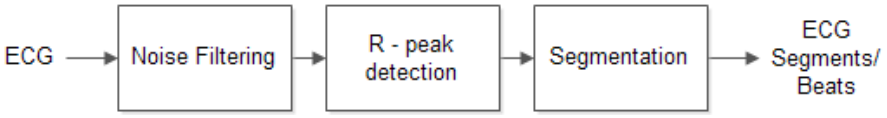


Fig. 1. Steps involved in pre-processing of the ECG Signal

A. Noise removal-FFT

This process involves the elimination of low-frequency noise components from the ECG signal. In other words, the uneven time-domain ECG signal is transformed to frequency domain using direct-FFT and the low frequency components are then removed. Next, a straightened ECG signal is restored with the help of inverse FFT. Fig. 2 (a) shows a sample ECG signal from the MIT-BIH database and Fig. 2 (b) is the straightened, reconstructed, noise free signal after removal of low frequency components.

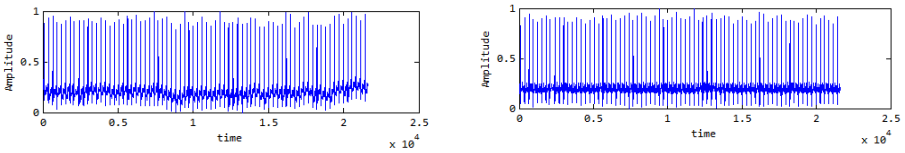


Fig. 2. An uneven sample ECG from the MIT-BIH Database

B. R-peak detection

R-peak detection involves finding the local maxima. Filtering is applied twice to locate the R-peaks with more accuracy. In the first filter, we use a windowed filter that "sees" only the maximum in the window and ignores all other values. A suitable window size is chosen and a filtered output is used to detect peaks. The second filter is a threshold filter. It uses the first filter output as well as the optimized window size calculated from the first filter to obtain an optimized set of peaks. Fig. 3 shows detected R-peaks in a sample ECG signal from the MIT-BIH database.

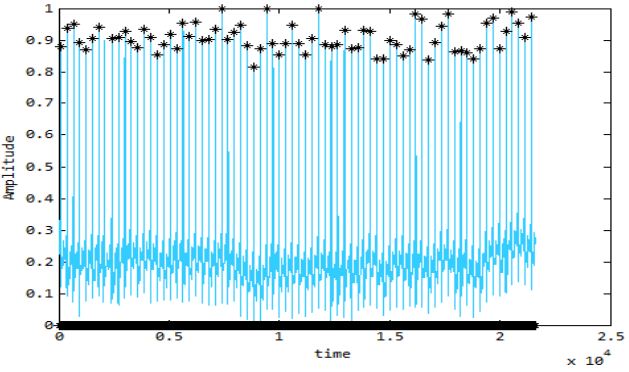


Fig. 3. Comparative ECG R-Peak Detection Plot

C. Segmentation

Segmentation of the ECG signal involves the extraction of single beats from the entire ECG. In order to form 1 segment/beat, the R-peaks detected in stage 2 are located. 50 samples before and 150 samples after the located peak form 1 segment/ beat. These segments are used for further processing. Fig. 4 shows a segment of a sample ECG signal with 50 samples before and 150 samples after the R-peak.

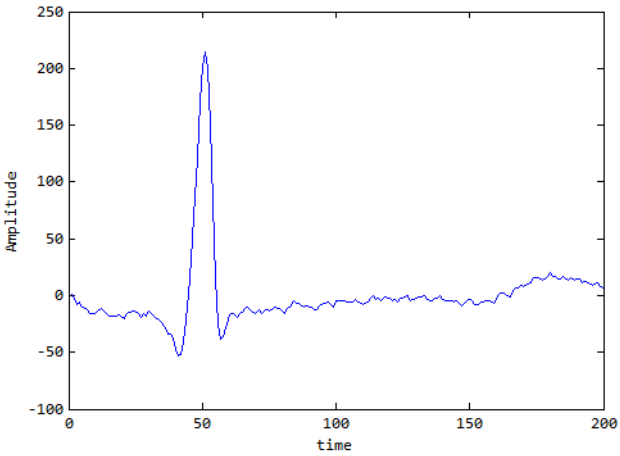


Fig. 4. A segment of a sample ECG signal of the MIT-BIH database

3 Feature Extraction

Feature extraction reduces dimensionality by projection of K-dimensional vector onto k-dimensional vector ($k < K$). In our experiments, the feature extraction process has been implemented using the 1D-Discrete Cosine Transform (DCT) and 1D-Discrete Wavelet Transform(DWT). The processes of feature extraction using DCT and DWT are elaborated in III-A and III-B respectively.

A. Feature Extraction using 1D-DCT

The Discrete Cosine Transform is given by Eqn. (1). In the 1D-DCT, most of the energy lies in the lower frequency components. It has the ability to pack most information in fewest coefficients.

$$y(k) = w(k) \sum_{n=1} x(n) \cos \frac{\pi(2n-1)(k-1)}{2N} \quad (1)$$

where,

$$k = 1, \ldots, N$$

$$w(k) = \begin{cases} \sqrt{\frac{1}{N}} & k = 1 \\ \sqrt{\frac{2}{N}} & 2 \leq k \leq N \end{cases}$$

B. Feature Extraction using 1D-DWT

One of the most commonly used technique for feature extraction is Discrete Wavelet Transform (DWT). Wavelets have many advantages compared to transforms like Discrete Fourier and Cosine Transforms. Functions with discontinuities and functions with sharp spikes usually take substantially fewer wavelet basis functions than sine-cosine functions to achieve a comparable approximation. DWT has the ability to provide spatial and frequency representations of the ECG signal simultaneously. The Haar Wavelet technique is a powerful technique for the multi-resolution decomposition of time series. It represents a signal by localizing it in both time and frequency domains. Dimensionality reduction is achieved when the 5-level-1D-Haar wavelet decomposition is performed and the approximation of the input ECG at each decomposition level is used as a feature vector. The dimensions of the feature vectors are 1x100, 1x50, 1x25, 1x13 and 1x7 corresponding to L1, L2, L3, L4 and L5 wavelet decomposition levels respectively.

4 Feature Selection-BPSO

Feature selection, implemented using Binary Particle Swarm Optimization (BPSO) [7], reduces dimensionality by selecting subset of original variables. Every particle in BPSO algorithm represents a possible candidate solution which is the feature subset. This algorithm operates on binary space where the particles are coded as binary strings and the velocities are constrained to the interval [0, 1] which is interpreted as change of probabilities.

BPSO for feature selection has the following advantages [8]: BPSO algorithm is derivative-free and is easy to implement. It has limited number of parameters and the impact of parameters to the solutions is small compared to other optimization techniques. It ensures convergence and the optimum value of the problem is calculated easily within a short time. The flowchart of BPSO used in feature selection process in shown in Fig. 5.

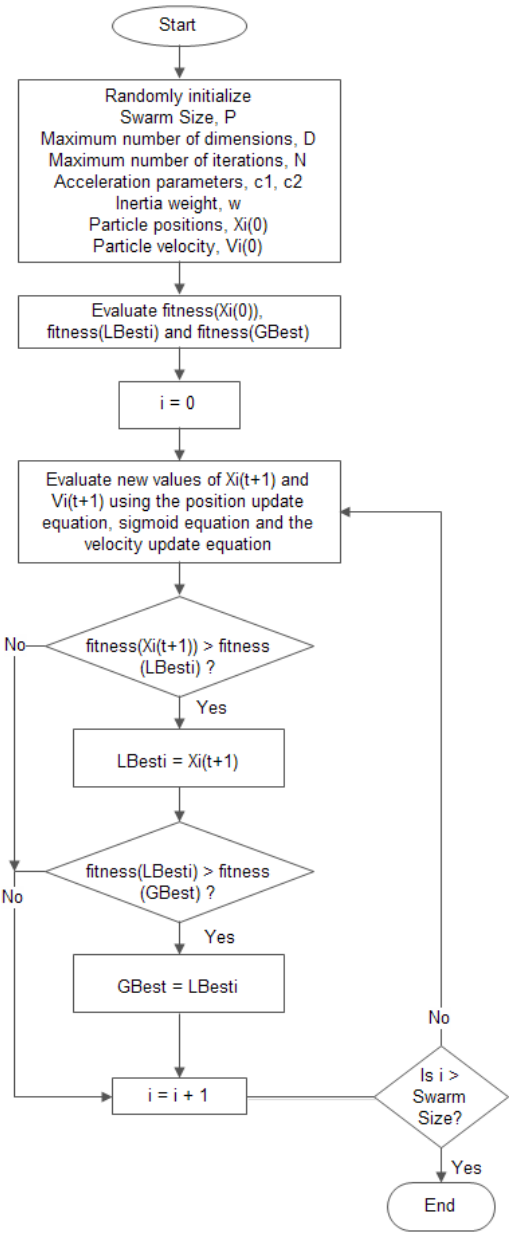


Fig. 5. Flowchart for BPSO

Let t (t = 1, 2, ..., $\text{iter}_{\max}$) denote the iteration number, i (i = 1, 2, ..., swarm size), the particle number, and **j** (**j** = 1, 2,..., N), the dimension number. Then the velocity update equation is given by:

$$v_i^j(t+1) = w \times v_i^j(t) + c_1 \times \text{rand}() \times (\text{lbest}_i^j - x_i^j(t)) + c_2 \times \text{rand}() \times (\text{gbest}^j - x_i^j(t)) \quad (2)$$

where c_1, c_2 are learning factors, w, the inertia weight, rand() is the uniform random number in the range [0, 1]. The sigmoid of velocity is calculated using:

$$\text{sigmoid}(v_i^j(t+1)) = \frac{1}{1+e^{(-v_i^j(t+1))}} \quad (3)$$

The position update equation is given by

$$x_i^j = \begin{cases} 1, & \text{rand}() < \text{sigmoid } v_i^j(t+1) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

The algorithm for BPSO [9] is given below:

1) Initialization.
2) Update velocity of particle i (Eqn. (2)).
3) Obtain sigmoid function of velocity (Eqn. (3)).
4) Update position of particle i (Eqn. (4)).
5) Repeat steps 2 through 4 for all particles of the swarm.
6) Calculate the fitness value according to Eqn. (7)
   - If fitness(x_i) < fitness(lbest_i), update the particle's best known position: $l_i \leftarrow x_i$
   - If fitness(lbest_i) < fitness(gbest), update the particle's best known position: gbest $\leftarrow \text{lbest}_i$
7) Repeat step 6 till a stopping condition is met.

Scatter Fitness Function: The fitness function in equation (7) evaluates the quality of evolved particles in terms of their ability to maximize the class separation term indicated by the scatter index among the different classes [10][11].

Let w_1, w_2, ..., w_L and N_1, N_2, ..., N_L denote the classes and number of ECG segments within each class respectively. Let M_1, M_2, ..., M_L and $M_\circ$ be the means of corresponding classes and the grand mean in the feature space. M_i is calculated as:

$$M_i = \frac{1}{N_i} \sum_{j=1}^{N} W_j^{(i)} \quad (5)$$

where i = 1, 2, ..., L and W_j^i, j = 1, 2, ..., N_i, represents the sample ECG segments from class w_i. The grand mean M_o is:

$$M_o = \frac{1}{n} \sum_{i=1}^{L} N_i M_i \quad (6)$$

where n is the total number of ECG segments in the chosen subset database. Thus, the scatter fitness function F is computed as follows:

$$F = \sqrt{\sum_{i=1}^{n} (M_i - M_o)^t (M_i - M_o)} \tag{7}$$

where, $(M_i - M_o)^t$ is the transpose of $(M_i - M_o)$

5 Absolute Euclidean Classifier

The classifier based on Euclidean distance, the Euclidean Classifier, is direct and simple and is commonly used for classification. The mean class values are used as class centers to calculate signal value distances for use by the Euclidean distance rule. For major level classification of a homogeneous area, this scheme is better. It's advantage comes from the minimum time it takes to classify [12]. In Arrhythmia Classification, it is employed to measure the similarity between the test vector and the reference vectors in the ECG signal gallery.

The proposed Absolute Euclidean Classifier (AEC) utilizes the straight-line absolute distance between two points for classification. For N-dimensional space, we propose the Absolute Euclidean distance between points p_i and q_i to be:

$$d = \sqrt{\sum_{i=1}^{n} (|p_i| - |q_i|)^2} \tag{8}$$

where $p_i(q_i)$ is the coordinate of p (q) in dimension i.

In Fig. 6, the distance d_2 from the proposed AEC block is found to be less than the distance d^1 from the EC block. Because of reduced Euclidean distance ($d_2 < d_1$), the Average Classification Rate (defined by Eqn. (9)) is significantly improved.

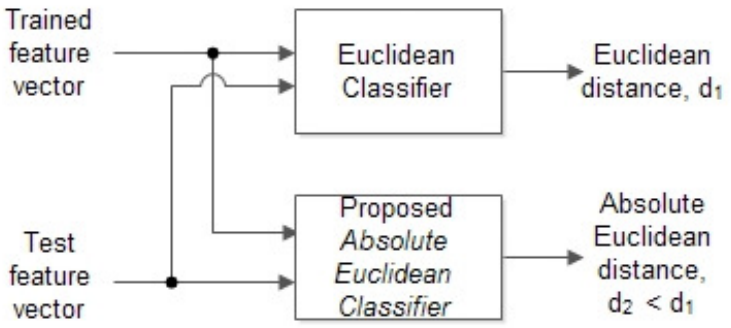


Fig. 6. Comparison between the EC and proposed AEC to calculate the feature vector distances d_1 and d_2

6 Discussion of Proposed Arrhythmia Classification Systems and Experimental Results

Experiments were carried out on different databases, namely, the MIT-BIH Arrhythmia Database [4], the MIT-BIH Atrial Fibrillation Database[5] and the MIT-BIH Malignant Ventricular Arrhythmia Database[6]. The beats obtained from all these databases are divided into two sets, a training set and a testing set. Classification Rate (CR) is defined as: ratio of the number of times a test ECG signals is correctly recognized to the total number of ECG signals in the testing set. It is expressed as a percentage.

$$CR = \frac{m}{t} \times 100\% \tag{9}$$

where m = No. of test ECG signals correctly recognized and
t = Total number of ECG signals in the testing set

A general block diagram of the proposed Arrhythmia Classification System is shown in Fig. 7. The blocks contain different pre-processing, feature extraction and feature selection steps for the ECG signals of the MIT-BIH Database.

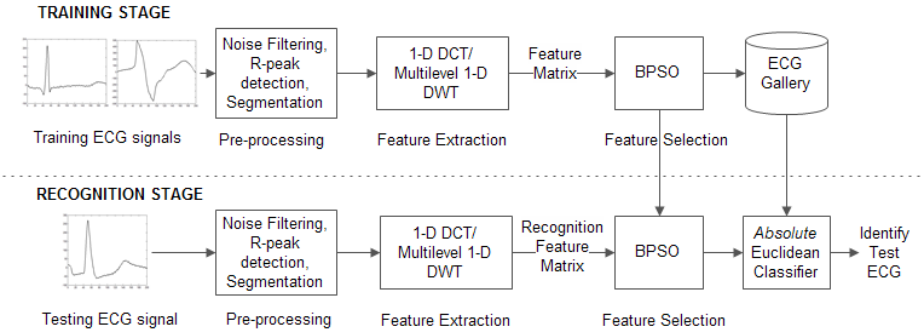


Fig. 7. Proposed Arrhythmia Classification System

Three sets of experiments are presented to evaluate the proposed techniques. The model is tested on Intel ® Core™ i5 CPU 2.67GHZ computer with 4.00GB RAM using 64-bit OS and 32-bit MATLAB ® signal processing tool for programming.

A. Experiment 1

In this experiment, we test the BPSO-based feature selection algorithm with feature vectors based on various sizes of DCT coefficients. The 1D-DCT is applied to the input ECG signal and only a subset of the DCT coefficients, i.e., the coefficients corresponding to higher energies, are retained. Subset sizes of 1x10, 1x20, 1x30, 1x40 and 1x50 of the original 1x200 DCT vector are used in this experiment as input to the subsequent feature selection phase. Table 1 and Fig. 9 show the Average

Classification Rate (ACR), Average number of Selected Features (ASF), Average Training Time (ATrT) and Average Testing Time (ATeT) for different feature vector dimensions using the BPSO based feature selection algorithms.

Table 1. Comparison of Average Classification Rate (ACR), Average number of Selected Features (ASF), Average Training Time (ATrT) and Average Testing Time (ATeT) for different feature vector dimensions

DCT_size	DCT (1x10)	DCT (1x20)	DCT (1x30)	DCT (1x40)	DCT (1x50)
ACR (%)	96.08	97.35	97.88	98.02	98.29
PCR (%)	97.08	98.54	99.17	98.75	98.75
ASF without BPSO	10	20	30	40	50
ASF with BPSO	10	20	30	37	44
ATrT (x 10^{-1} s)	25.65	26.64	27.02	26.58	27.65
ATeT (x 10^{-3} s)	17.73	19.29	18.92	18.50	18.02

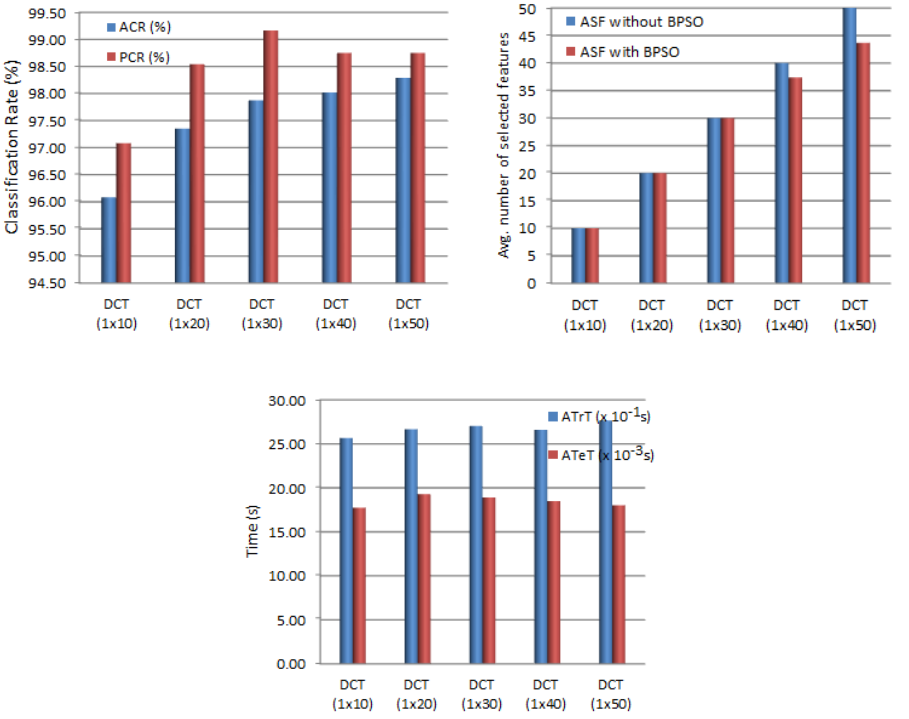


Fig. 8. Classification results for different DCT-feature based vectors. (a) Average Classification Rate (ACR) (b) Average number of Selected Features (ASF) (c) Average Training Time (ATrT) and Average Testing Time (ATeT).

The best ACR of 98.3% is achieved using the DCT (1x50) feature vector and the PSO-based feature selection algorithm. In this instance, the selection algorithm reduces the size of the original feature vector by nearly 12%. In general the BPSO has comparable performance in terms of classification rate but in all test cases the number of selected features is smaller when using the BPSO selection algorithm. On the other hand, in terms of computational time, BPSO selection algorithm takes less training time, which indicates that BPSO is computationally cheaper and effective.

B. Experiment 2

In this experiment, the DWT coefficient features have been extracted from each ECG signal. The 1-D Haar wavelet transform is applied to the input ECG signal reducing its size to half of its original size. 5-level wavelet decomposition is performed and the approximation of the input ECG signal at each decomposition level is used as a feature vector. The dimensions of the feature vectors are 1x100, 1x50, 1x25, 1x13 and 1x7 corresponding to levels L1, L2, L3, L4 and L5 wavelet decompositions respectively. Table 2 and Fig. 9 show Average Classification Rate (ACR), Average number of Selected Features (ASF), Average Training Time (ATrT) and Average Testing Time (ATeT) for different feature vector dimensions using the BPSO based feature selection algorithms. The best ACR of 97.2% is achieved using the DWT (1x50) feature vector and the BPSO-based feature selection algorithm using only 44 selected features.

Table 2. Comparison of Average Classification Rate (ACR), Average number of Selected Features (ASF), Average Training Time (ATrT) and Average Testing Time (ATeT) for different feature vector dimensions

DWT level	L1 (1x100)	L2 (1x50)	L3 (1x25)	L4 (1x13)	L5 (1x7)
ACR (%)	96.96	97.19	97.08	95.75	95.13
PCR (%)	97.92	98.13	97.71	97.71	96.88
ASF without BPSO	100	50	25	13	7
ASF with BPSO	79	46	25	13	7
ATrT (x 10^{-1}s)	39.53	34.54	36.53	29.70	41.86
ATeT (x 10^{-3}s)	27.61	26.87	28.35	22.75	32.79

C. Experiment 3 - Effect of Absolute Euclidean Classifier

In this experiment we prove the effectiveness of the proposed Absolute Euclidean Classifier (AEC). The distance between train and test vectors obtained using AEC is found to be less than the distance obtained using the Euclidean Classifier for signals of similar arrhythmia class and this distance is found to be more between the train and test vectors of different classes. Because of reduced Absolute Euclidean distance, the Average Classification Rate (ACR) (defined by Eqn. (9)) is significantly improved with the proposed Absolute Euclidean Classifier as shown in Fig. 10.

D. Comparison with other Arrhythmia Classification Systems

The proposed Arrhythmia Classification System is compared with exiting algorithms and methods like Multilevel Perceptron and Classification Tree (MLP-CT) [13], Self-organized ANN (S-O ANN) [14], Fuzzy Classifier (FC) [15], Genetic-ESVM with linear kernel (G-ESVM LK) [3] and Automated Patient-Specific Classification (APSC) [2]. The proposed algorithm has proven to be more efficient compared to these methods (Table 3).

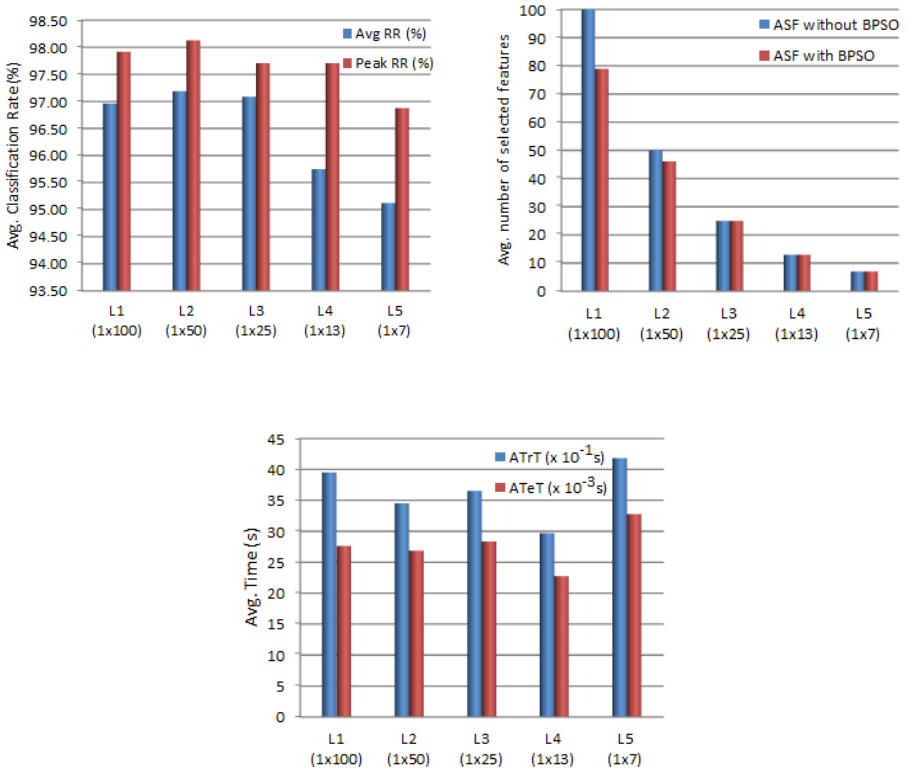


Fig. 9. Classification results for different DWT-feature based vectors. (a) Average Classification Rate (ACR) (b) Average number of Selected Features (ASF) (c) Average Training Time (ATrT) and Average Testing Time (ATeT)

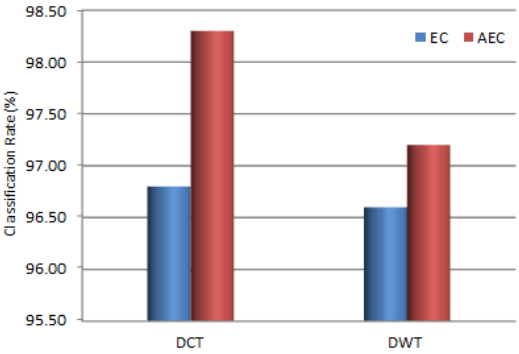


Fig. 10. Graph to demonstrate the increase in Classification Rate as a result of modification in the classifier equation

Table 3. Comparison with other Arrhythmia Classification Systems

Method	ACR (%)
MLP-CT [13]	87.90
S-O ANN [14]	92.88
FC [15]	93.13
G-ESVM LK [3]	96.83
APSC [2]	97.30
Proposed Method	98.30

In summary, the best results achieved using the proposed methods are shown in Table 4.

Table 4. Summary of Experimental Results

ECG signal length	21	
ECG segment length	200	
No. of segments per anomaly	200	
No. of training segments per anomaly	80	
No. of testing segments per anomaly	120	
	DCT+BPSO FS	DWT+BPSO FS
Average Classification Rate	98.3	97.2
Average No. of selected features	44	46
Average Training Time	2.	4.
Average Testing Time per ECG segment	18.02	32.79

7 Conclusion

This paper proposes a method for ECG arrhythmia classification, which is simple, yet very effective. We demonstrate that the use of the suggested preprocessing, feature extraction and feature selection techniques along with the proposed Absolute Euclidean Classifier enhances the classification rate. Experiments conducted on the MIT-BIH databases yielded an average classification rate of 98.30% using Discrete Cosine Transform and 97.20% using Discrete Wavelet Transform. The Binary Particle Swarm Optimization plays a significant role in decreasing the number of features selected, thereby reducing computational time and cost.

Future work includes a real-time implementation of the proposed system, improving the system so as to enable it to classify more types of arrhythmias, and exploring more techniques for feature extraction and selection.

Acknowledgement. The authors would like to thank Sergey Chernenko for his shared MATLAB demo of ECG processing.

References

[1] Khazaee, A., Ebrahimzadeh, A.: Classification of Electrocardiogram Signal with support vector machines and genetic algorithms using power spectral features. Biomedical Signal and Control 5, 252–263 (2010)
[2] Ince, T., Kiranyaz, S., Gabbouj, M.: A Generic and Robust System for Automated Patient-Specific Classification of ECG Signals. IEEE Transactions on Biomedical Engineering 56(5) (May 2009)
[3] Melgani, F., Bazi, Y.: Classification of Electrocardiogram Signals With Support Vector Machines and Particle Swarm Optimization. IEEE Transactions on Information Technology in Biomedicine 12(5) (September 2008)
[4] http://www.physionet.org/physiobank/database/mitdb/
[5] http://www.physionet.org/physiobank/database/afdb/
[6] http://www.physionet.org/physiobank/database/vfdb/
[7] Kennedy, J., Eberhart, R.C.: A discrete binary version of the particle swarm algorithm. In: Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics 1997, Piscataway, NJ, pp. 4104–4109 (1997)
[8] Peng, C., Xu, X.: A hybrid algorithm based on immune BPSO and N-1 principle for PMU multi-objective optimization placement. In: Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (April 2008)
[9] Hu, X., Shi, Y., Eberhart, R.: Recent advances in particle swarm, Evolutionary Computation Congress. In: CEC 2004, vol. 1, pp. 0-7803–8515-2 (June 2004)
[10] Liu, C., Wechsler, H.: Evolutionary Pursuit and Its Application to Face Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 22(6), 570–582 (2000)
[11] Ramadan, R.M., Abdel, R.F.: Face Recognition Using Particle Swarm Optimization-Based Selected Features. International Journal of Signal Processing, Image Processing and Pattern Recognition 2(2) (June 2009)
[12] Chelliah, S.: A Study of Euclidean classifier. Remote Sensing Division. Birla Institute of Scientific Research Jaipur, India

[13] Lin, Y.-J., Yang, J.-X.: ECG Patterns Recognition using Multilayer Perceptron and Classification Tree. In: Proceeding of International Medical Informatics Symposium, Taiwan (2007)
[14] Rogal Jr., A.S.R., Paraiso, E.C., Kaestner, C.A.A.: Automatic Detection of Arrhythmias Using Wavelets and Self-Organized Artificial Neural Networks. In: Ninth International Conference on Intelligent Systems Design and Applications (2009)
[15] Anuradha, B., Veera Reddy, V.C.: Cardiac Arrhythmia Classification Using Fuzzy Classifiers. Journal of Theoretical and Applied Information Technology (2008)

Design of Low Power High Speed 4-Bit TIQ Based CMOS Flash ADC

Parvaiz Ahmad Bhat[1] and Roohie Naaz Mir[2]

[1] Govt. Women's Polytechnique, Srinagar, India
b_parvaiz@yahoo.co.in
[2] National Institute of Technology Srinagar, 190006, J&K, India
naaz310@yahoo.co.in

Abstract. The analog-to-digital converter (ADC) is an essential part of system-on-chip (SoC) products because it bridges the gap between the analog physical world and the digital logical world. In the digital domain, low power and low voltage requirements are becoming more important issues as the channel length of MOSFET shrinks below 0.25 sub-micron values. SoC trends force ADCs to be integrated on the chip with other digital circuits. These trends present new challenges in ADC circuit design. This paper investigates high speed, low power, and low voltage CMOS flash ADCs for SoC applications.

The proposed ADC utilizes the Threshold Inverter Quantization (TIQ) technique that uses two cascaded CMOS inverters as a comparator. The TIQ technique proposed here has been developed for better implementation in SoC applications. The preliminary results show that the TIQ flash ADC achieves high speed, small size, low power consumption, and low voltage operation compared to other ADCs.

1 Introduction

Semiconductor technology is now approaching 100 nanometer feature size and will soon be below100 nanometer. This technology trend presents new challenges in analog-digital mixed signal circuit design. A mixed signal circuit must be integrated on a single chip along with logic and memory circuits to form a system-on-chip. The mixed signal circuit must operate at fast speeds along with digital logic and memory circuits; otherwise it becomes a bottleneck to the system.

1.1 Challenges in Designing ADC's for SOC

The major considerations in designing ADCs for the complete SoC are high speed, low power, and low voltage. In terms of high speed [1], presently 0.130μm CMOS technology allows processor speeds in excess of 2.4 GHz. However, the sampling speed of ADC's fabricated with an advanced BiCMOS process was around 200 mega samples per second (MSPS). High speed ADCs with a bipolar process operating up to 1.5 giga samples per second (GSPS) for digital oscilloscopes, digital RF/IF signal processing, direct RF down-conversion, and radar/ECM systems have also been produced recently.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 319–328.
springerlink.com

The next challenge is low power consumption. ADC's should be integrated with digital circuits on a single chip for the portable devices. All battery powered devices are now being designed to include low power techniques to prolong the battery life. Similarly, ADCs need low power architecture or a low power technique. Low voltage operation is one of the difficult challenges in the mixed signal ICs. The down-scaling of the minimum channel length to 0.065μm results in the reduction of the power supply voltage to 0.7 V [6]. A mixed- signal circuit designer faces a great challenge when designing an ADC that operates at low voltage because of the relatively high threshold voltage of the transistors. As a result, an ADC should be operated in a small voltage range.

1.2 Solid State Technology

The type of solid-state technology used to implement the converter also affects the speed of an ADC [2]. Three different types of solid-state technologies are currently used for high speed ADC implementations: CMOS technology, bipolar technology, and Gallium Arsenide (GaAs) technology. GaAs technology is the fastest of the three, and CMOS technology is the slowest. Bipolar technology allows faster operation and is compatible with the CMOS technology. However, BiCMOS technology requires more processing steps and higher cost compared to standard CMOS technology. Therefore, mixed-signal circuit implementation using only the standard CMOS technology is the preferred choice for SoC products [7].

The proposed ADC in this work utilizes the Threshold Inverter Quantization (TIQ) technique that uses two cascaded CMOS inverters as a comparator. The TIQ technique proposed has been developed for better implementation in SoC applications. The ADC is designed and simulated in 0.12μm CMOS and operates at 1GSamples/sec. Differential/integral nonlinearity (DNL/INL) errors are between –0.031 to 0.026 LSB and -0.024 to 0.011LSB, respectively.

The rest of the paper is organized as follows: section 2 describes the related work; section 3 introduces the proposed design and section 4 presents the ADC architecture; section 5 presents the simulation results. A conclusion is presented in section 6 and the references are listed in the end.

2 Related Work

Most of the researchers investigate techniques to enhance speed or reduce power consumption. [9] presents a 4-bit flash-type ADC suitable for ultra wide band applications. This design shows low power consumption due to use of transistors with reduced dimensions. [11] presents a simple and fast flash ADC using TIQ technique. It offers higher data conversion rates while maintaining comparable power consumption levels making t suitable for SoC integration using the standard digital CMOS process. [3] presents a 4-bit flash ADC for wide band applications. It uses clocked digital comparators that perform the track/hold function thus avoiding the harmonic and inter modulation distortion usually seen in high frequency signals. [4] presents an ADC that uses less number of analog components, small size and low power. It is suitable

for integration with DSP core for SoC applications. [10] presents a 4-bit flash ADC using preamplifiers and comparators to provide fast overdrive recovery. In order to enhance the speed, analog part of the ADC is fully pipelined.

Our work is aimed at developing the design techniques for Flash ADCs with emphasis on high-speed and low-power operation using TIQ as a comparator and comparing the performance of proposed ADC using three different types of encoders.

3 Proposed Work – TIQ Flash ADC

We propose high speed CMOS architecture with low power consumption, which is featuring the Threshold Inverter Quantization (TIQ) technique. Fig. 1 shows the TIQ schematic diagram. The main advantage of the TIQ based CMOS flash ADC design is a simpler comparator design. The idea is to use digital inverters as analog voltage comparators. This eliminates the need for high-gain differential input voltage comparators that are inherently more complex and slower than the digital inverters. The TIQ flash ADC also eliminates the need of reference voltages, which require a resistor ladder circuit. This simplicity in the comparator part provides both high speed and lower power consumption at the same time.

The analog quantization level of digital comparator is the switching threshold voltage of the quantization inverter. It is a reference voltage and is self-determined by the size ratio of NMOS and PMOS. The internal reference voltage, V_m, is defined as the input voltage V_{in} of the quantization inverter when the output voltage V_{O1} equals to V_i, where both PMOS and NMOS transistors are in saturation. Fig. 2 shows the static voltage transfer characteristic (VTC) of the inverter. The voltage V_{dd} is the supply voltage of the process. By changing the widths of the PMOS and NMOS devices with a fixed transistor length, we get different threshold voltage. The value of V_m is expressed in equation (1). All figures will be printed in black and white. All figures are to be numbered using Arabic numerals. Figures should always be cited in text in consecutive numerical order. Figure parts should be denoted by lowercase letters (a, b, c, etc.). Each figure should have a concise caption describing accurately what the figure depicts. Include the captions in the text file of the manuscript, not in the figure file.

$$V_m = \frac{\sqrt{\frac{\mu_p W_p}{\mu_n W_n}}\left(V_{dd} - \left|V_{Tp}\right|\right) + V_{Tn}}{1 + \sqrt{\frac{\mu_p W_p}{\mu_n W_n}}} \tag{1}$$

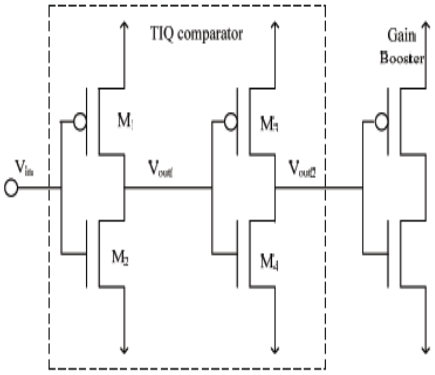


Fig. 1. TIQ Comparator Schematic diagram

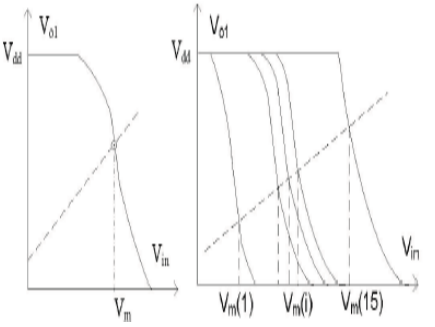


Fig. 2. Static VTC

4 ADC Architecture

The proposed flash ADC features the threshold inverter quantization (TIQ) technique for high speed and low power using standard CMOS technology that is compatible with microprocessor fabrication. Fig. 3 shows the block diagram of the TIQ flash ADC.

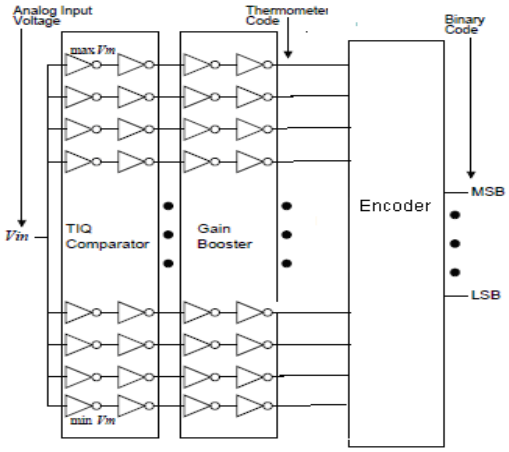


Fig. 3. Block Diagram of a TIQ Flash ADC

4.1 TIQ Comparator

TIQ role is to convert an input voltage (V_{in}) into logic '1' or '0' by comparing a reference voltage (V_{ref}) with the V_{in}. If V_{in} is greater than V_{ref}, the output of the comparator is '1', otherwise '0'. Two cascading CMOS inverters can be used as a comparator for high speed and low power consumption.

4.2 Gain Booster

Each gain booster consists of two cascading inverters with the same circuit as the comparator, but the transistor sizes of each gain booster are small and identical. The gain booster is used to increase voltage gain of the output of a comparator so that it provides a full digital output voltage swing. The propagation delay's trend is almost exponentially proportional [5] to the transistor length, but the voltage gain follows a logarithmic function. Therefore, both propagation delay and voltage gain should be considered together when we choose the size of the gain booster.

4.3 TC-to-BC Encoder

TIQ comparator array produce a thermometer code (TC), which needs to be converted into binary code (BC) using TC to BC encoder. Different types of encoder can be used to perform conversion. In our work we have used three types of encoders namely Fat tree encoder, ROM based encoder and a simple encoder. Simple encoder directly converts TC to BC unlike first two, which convert TC to BC in two steps.

5 Simulation Results

In this section, we present experimental results of the 4-bit TIQ flash ADC. The TIQ flash ADCs have been designed with standard CMOS technology [2] of 120nm with

ADS 2006A tool. The HSPICE models (BSIM3 level 49) have been used as the standard library. Table 1 lists the parameters of the 4-bit TIQ based Flash ADC. Figures 4 to 10 show the results of simulations carried out in this work.

5.1 Result Interpretation

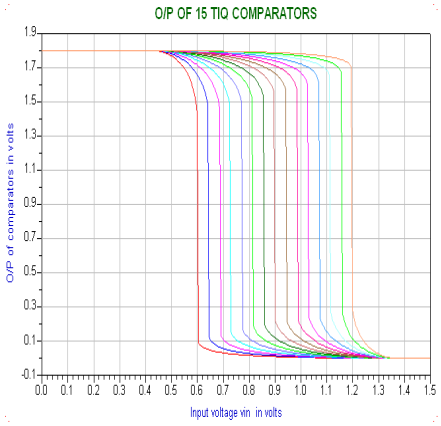


Fig. 4. Output for 4-bit TIQ comparator

Fig 4 shows DC simulation results of 15 TIQ comparators and shows the uniformity of 15 equally spaced invertors threshold voltages calculated from the equation (1) above.

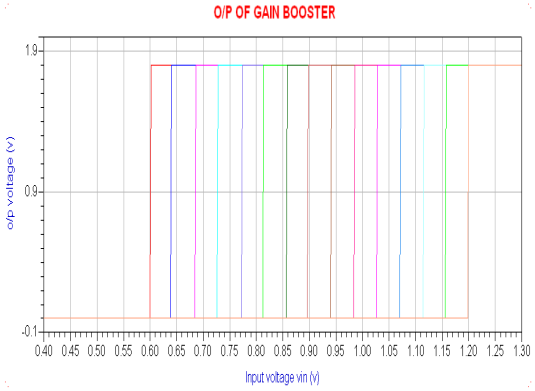


Fig. 5. Output of Gain Booster

Fig 5 shows the DC simulation results of the gain booster which consists of two invertors in cascade. Gain Booster is used to increase voltage gain of the output of comparator so that it provides a full digital output voltage swing with sharp transition.

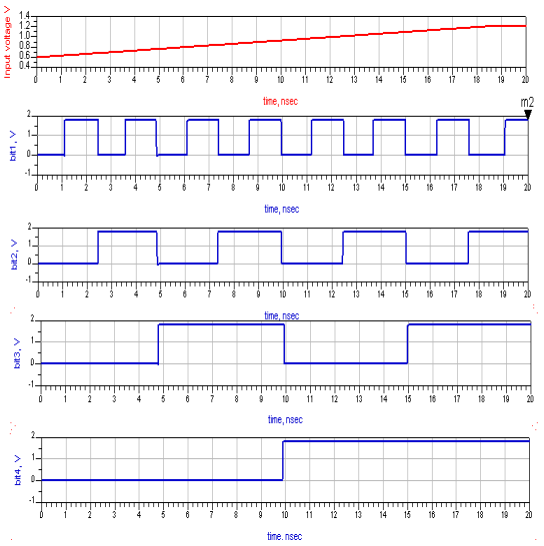


Fig. 6. Digital Output for 4 bit TIQ Flash ADC

Fig 6 shows the digital output of an ADC for the Ramp input which varies from 0.58V to 1.21V. In the figure, Bit 1 is the LSB bit and Bit 4 is the MSB bit.

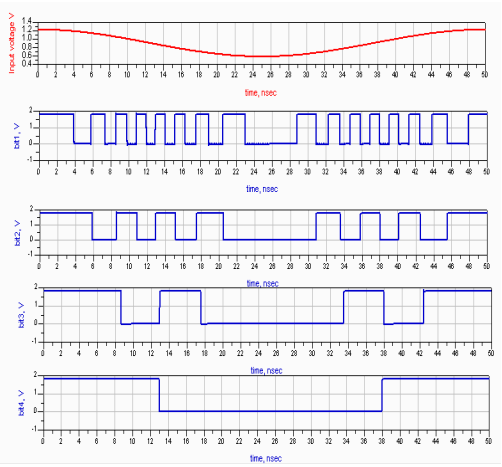


Fig. 7. Output of a 4-bit TIQ flash ADC with sinusoidal input 20MHz

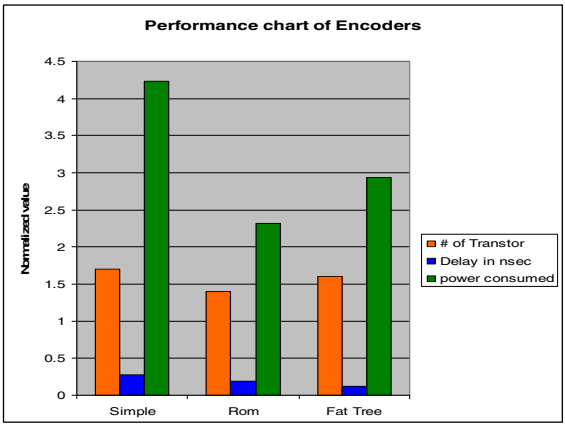


Fig. 8. Performance Chart of Different Encoders

Fig. 8 shows the performance chart of different types of encoders used with the three parameters (no. of transistors delay in nano-seconds) and power consumption in mWatts. From the figure, it is clear that the delay and the power consumption of fat-tree based encoder has least delay and power consumption.

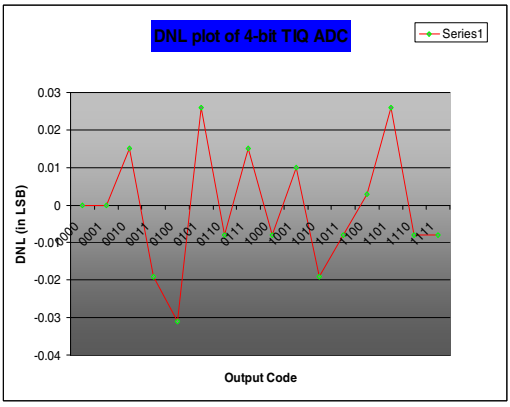


Fig. 9. DNL Curve of 4-bit TIQ based Flash ADC

Fig. 9 shows the DNL curve of 4-bit TIQ based ADC. The DNL range is from -0.031 LSB to + 0.026LSB.

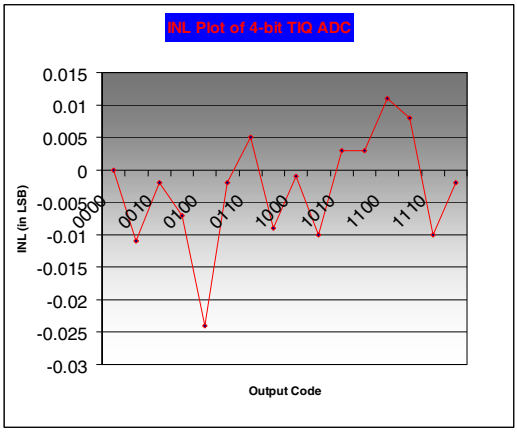


Fig. 10. INL Curve of 4-bit TIQ based Flash ADC

Fig. 10 shows the INL curve of 4-bit TIQ based ADC. The INL range is from -0.024 LSB to +0.011LSB.

6 Conclusion and Future Work

A simple and fast flash ADC architecture that uses two cascaded CMOS inverters as a comparator, called Threshold Inverter Quantization (TIQ) technique, has been developed. The TIQ flash ADC offers higher data conversion rates while maintaining comparable power consumption levels so that it is also highly suitable for the complete SoC integration using the standard digital CMOS process. The simulation test results showed that the fat tree encoder outperformed the commonly used ROM type encoder in terms of speed, power consumption, and area for the 4-bit TIQ flash ADC. As a future work we will improve the design in many ways. For low power design it is required to generate the MOSFET width automatically so that the power consumption of TIQ comparator block can be further reduced. To achieve high speed as well as high resolution it is possible to use 4-bit flash ADC in pipelined ADC structure. Moreover the time interleaved concept can be used to increase speed.

References

1. Demler, M.J.: High-speed Analog-to-Digital Conversion. Academic Press, Inc.
2. Van de Plassche, R.: CMOS Integrated Analog to Digital and Digital to Analog Converter. Kluwer Academic Publications (2004)
3. Wang, M., Chen, C.-I.H.: A High Spurious-Free Dynamic Range 4-bit ADC with Nyquist Signal Bandwidth for Wideband Communications. Appear in IEEE (2007)
4. Wang, M., Chen, C.-I.H.: Architecture and Design Synthesis of 2.5 G samples/s 4-b Pipelined Flash ADC in SoC Applications. Appear in IEEE (2005)
5. Waltari, M.E., Holonen, K.A.L.: Circuit techniques for low voltage high speed A/D converter. Kluwer Academic Publications (2004)

6. Donovan, C., Flynn, M.P.: A Digital 6-bit ADC in 0.25-μm CMOS. IEEE Journal of Solid-State Circuits 37(3) (March 2002)
7. Allen, P.E., Holberg, D.R.: CMOS Analog Circuit Design, 2nd edn. Oxford University Press, New York (2002)
8. Tangel, A., Choi, K.: The CMOS Inverter as a comparator in ADC designs. Pennyslvania State University, University Park, USA
9. Shehata, K.A., Ragai, H.F., Husien, H.: Design and implementation of a high speed, low power 4-bit flash ADC. IEEE (2009)
10. Wu, L., Huang, F., Gao, Y., Wang, Y., Cheng, J.: A 42 mW 2 GS/s 4-bit flash ADC in 0.18-μm CMOS. The Proceedings of IEEE (2009)
11. Iyappan, P., Jamuna, P., Vijayasamundiswary, S.: Design of Analog to Digital Converter Using CMOS Logic. The Proceedings of IEEE (2009)
12. Yoo, J.: A TIQ based CMOS flash A/D converter for system on chip applications. A PhD Thesis in Computer Science and Engineering. The Pennsylvania State University (May 2003)
13. Lee, D., Yoo, J., Choi, K., Ghaznavi, J.: Fat Tree Encoder Design for Ultra High Speed Flash A/D Converters. Pennsylvania State University (2002)
14. Yoo, J., Choi, K., Tangel, A.: A1-GSPS CMOS Flash Analog-to-Digital Converterfor System-on-Chip Applications. Pennyslvania State University, University Park, USA

A Reduced Complexity LDPC Decoding Algorithm Using Dynamic Bit Node Selection

Suvarna Hudgi and Siddram R. Patil

H. No 8-1304/10
Gandhi Nagar
Gulbarga 585104
shreesuvarna@gmail.com,
pdapatil@yahoo.com

Abstract. A simple and effective computational complexity reducing method for iterative message passing decoding algorithm of Low-Density Parity-Check (LDPC) codes is described. In each iteration, the algorithm selects the fraction of bit nodes with least reliability. At both check node processors as well as bit node processors, the extrinsic messages only for selected nodes are updated while for others the previous values are retained. The algorithm is based on a dynamic selection of bit nodes for updating the messages for each iteration. The complexity analysis and the simulation results shows that the method achieves up to 90 % saving in computations for high rate codes of code rate 0.9 compared to the standard Belief-Propagation (BP) algorithm while maintaining the same bit error rate performance.

1 Introduction

Gallager[1] introduced LDPC codes as a family of linear block codes with parity-check matrices containing mostly zeros and only a small number of ones. The "sparsity" of the parity-check matrices enables their efficient decoding by various message-passing decoding algorithms. MacKay and Neal [2][3] develop further interest on the LDPC codes and their decoding. Chung [4] has derived a theoretical threshold that is within 0.0045 dB away from Shannon capacity. The simulation results there in show that the LDPC code of length 10^7 achieves a bit error rate (BER) of 10^{-6} at a bit energy to noise density ratio (E_b / N_o) within 0.04 dB away from the one that is required to approach Shannon limit. This margin is the lowest amongst the codes discussed in literature.

An LDPC code is a linear block code specified in terms of a sparse $M \times N$ parity check matrix H whose elements are 0 and 1. A (N, w_c, w_r) LDPC code represents a code of length N where H has w_r number of 1's in each row and w_c number of 1's in each column. The parity check matrix of an LDPC code can be described by corresponding Tanner graph that displays the relationship between codeword bits and parity-checks. Each of the N code bits and M parity-checks in H are represented

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 329–337.
springerlink.com

by a node in the graph. The columns of H represent variable nodes or bit nodes, and the rows of H represent the check nodes in the corresponding Tanner graph. A graph edge joins a bit node to the nodes of the parity checks that include it. The degree of a bit node is the number of check equations that it participates in. Similarly, degree of a check node is the number of bit nodes which take part in that particular check. For the case where all bit nodes have the same degree and all the check nodes have same degree, then it is known as a regular LDPC code [3]. For such a case the code rate can be given by $1-(w_c / w_r)$. These degrees are different for the case of irregular LDPC codes where the irregularity is typically specified using two polynomials called bit node and check node degree profiles or degree distributions [5]. *Figure 1* shows an example of parity check matrix for which the Tanner graph can be shown by *Figure 2*. The parity check matrix represents a (10, 2, 4) regular LDPC code with rate ½.

$$H = \begin{bmatrix} 1\,1\,1\,1\,0\,0\,0\,0\,0\,0 \\ 1\,0\,0\,0\,1\,1\,1\,0\,0\,0 \\ 0\,1\,0\,0\,1\,0\,0\,1\,1\,0 \\ 0\,0\,1\,0\,0\,1\,0\,1\,0\,1 \\ 0\,0\,0\,1\,0\,0\,1\,0\,1\,1 \end{bmatrix}$$

Fig. 1. An example parity check matrix H

The decoding of LDPC codes is done using either a hard decision iterative decoding algorithm such as bit flipping (BF) or a soft decision iterative decoding algorithm such as belief propagation (BP) algorithm. The soft decision decoding of binary LDPC codes uses an iterative message-passing algorithm (MPA), which is an instance of Pearl's belief propagation (BP) algorithm operating on the Tanner graph of the code.

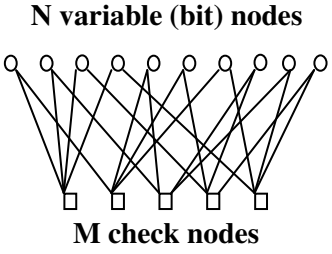


Fig. 2. Tanner graph representation

The message passing algorithm is also known as the sum-product algorithm, in which the messages are passed between the bit node and the check node. If there are no cycles in the graph and the graph is finite, then the sum-product algorithm after many iterations is equivalent to a maximum a posteriori (MAP) decoding algorithm.

In practice it is not possible to avoid cycles in an LDPC codes. Still the sum-product algorithm performs quite well. Operational diagram of the LDPC decoder is shown in Figure 3.

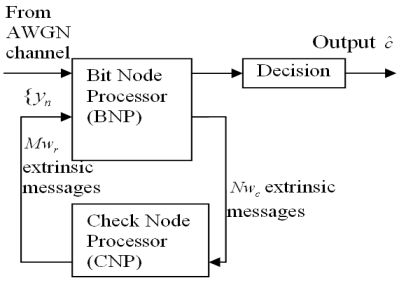


Fig. 3. The LDPC iterative decoder

Low-density parity-check (LDPC) code [1] can achieve good performance when decoded by the belief propagation (BP) decoding algorithm, but the computational complexity is rather high. It is practically important to simplify the iterative decoding algorithm to speed-up the computation, and reduce the complexity without performance degradation. Among the notable approaches for reducing the complexity of LDPC decoding are those of min-sum algorithm[6][7], scheduling based techniques[8][9][10] and forced convergence methods[11][12]. The forced convergence methods reduce the complexity cost of an iteration. The problem of forced convergence method is to select appropriate threshold for deciding the bits with high reliable value and low reliable value. However in such cases the threshold is code specific, sensitive g to Eb/No, and may lead to threshold wondering problems. Recently a reduced complexity decoding algorithm based on column layering is given in [13]. In our work, an algorithm is developed that offers a substantial reduction in computational complexity with the same bit error rate (BER) performance as that of standard BP.

A simple and effective computational complexity reducing method for iterative message passing decoding algorithm of Low-Density Parity-Check (LDPC) codes is described. In each iteration, the algorithm selects the fraction of bit nodes with least reliability. At both check node processors as well as bit node processors, the extrinsic messages only for selected nodes are updated while for others the previous values are retained. The algorithm is based on a dynamic selection of bit nodes for updating the messages for each iteration. The complexity analysis and the simulation results shows that the method achieves up to 90 % saving in computations for high rate codes of code rate 0.9 compared to the standard Belief-Propagation (BP) algorithm while maintaining the same bit error rate performance.

2 Standard BP Decoding of LDPC Codes

The basic decoding algorithm used for LDPC codes as reported in literature is described in this section. The (N, d_v, d_c) LDPC codes can be represented by a bipartite graph with N variable nodes on the left and $M = \left(Nd_v / d_c\right)$ check nodes on the right. A bipartite graph is specified by sequences $\left(\lambda_1, \lambda_2, \ldots, \lambda_{d_v}\right)$ and $\left(\rho_1, \rho_2, \ldots, \rho_{d_c}\right)$, here $\lambda_i(\rho_i)$ represents the fraction of edges with left (right) degree i, and d_v and d_c are the maximum variable degree and check degree, respectively.

Suppose the LDPC code C is used for error control over an AWGN channel with zero mean and power spectral density $N_o/2$. Assume BPSK signaling with unit energy, which maps a codeword $w = (w_1, w_2, \ldots, w_N)$ into a transmitted sequence $q = (q_1, q_2, \ldots, q_N)$, according to $q_n = (1 - 2w_n)$, for $n = 1, 2, \ldots, N$. If $w = [w_n]$ is a code word in C and $q = [q_n]$ is the corresponding transmitted sequence, then the received sequence is $q + g = y = [y_n]$, with $y_n = q_n + g_n$, where for $1 \le n \le N$, g_n's are statistically independent Gaussian random variables with zero mean and variance $\sigma^2 = N_o/2$. Let $H = [H_{mn}]$ be the parity check matrix which defines an LDPC code. We denote the set of bits that participate in check m by $N(m) = \{n : H_{mn} = 1\}$ and the set of checks in which bit n participates as $M(n) = \{m : H_{mn} = 1\}$. We also denote $N(m) \setminus n$ as the set $N(m)$ with bit n excluded, and $M(n) \setminus m$ as the set $M(n)$ with check m excluded. We define the following notations associated with the i th iteration:

$U_{ch,n}$: The log-likelihood ratio (LLR) of bit n which is derived from the channel output y_n. In BP decoding, we initially set $U_{ch,n} = 2y_n / \sigma^2$.

U_{mn}: The LLR of bit n which is sent from check node m to bit node n

V_{mn}: The LLR of bit n which is sent from bit node n to check node m

V_n: The a posteriori LLR of bit n

The standard BP algorithm is carried out as follows

Initialization: Set $i = 1$, and the maximum number of iterations to $I_{\max}$. For each m, n, set $V_{mn} = U_{ch,n}$

Step 1: Horizontal step, for $1 \le n \le N$ and each $m \in M(n)$, process

$$U_{mn} = 2 \tanh^{-1} \prod_{n' \in M(n) \setminus n} \tanh \frac{V_{mn'}}{2}$$

Step 2: Vertical Step, for $1 \le n \le N$ and each $m \in M(n)$, process

$$V_{mn} = U_{ch,n} + \sum_{m' \in M(n) \setminus m} U_{m'n}$$

$$V_n = U_{ch,n} + \sum_{m \in M(n)} U_{mn}$$

Step 3: Hard decision and stopping criterion test:

(i) Create $\hat{w} = [\hat{w}_n]$ such that $\hat{w}_n = 1$ if $V_n < 0$ and $\hat{w}_n = 0$ if $V_n \geq 0$

(ii) If $H\hat{w} = 0$ or the maximum iteration number $I_{\max}$ is reached, stop the decoding iteration and go to Step 4. Otherwise go to Step 1.

Step 4: Output $\hat{w}$ as the decoded codeword

3 Reliability Based Decoding

The proposed algorithm is based on the reliability of each received coded bit in LDPC decoder. In this section we discuss the meaning of reliability and basic concept of reliability decoding. Let x be in $GF(2)$ with elements {+1, -1}, where +1 is the 'null' element under the addition. Then the Log-Likelihood Ratio (LLR) value of the binary random variable is defined as

$$L(x) = \ln \frac{P(x=+1)}{P(x=-1)} \tag{1}$$

Inversely, given the LLR value we can calculate the probability of the bit as

$$P(x = \pm 1) = \frac{\exp(\pm L(x)/2)}{\exp(+L(x)/2) + \exp(-L(x)/2)} \tag{2}$$

If L_n is the LLR value for the n^{th} bit, it can be expressed as

$$L_n = \mathrm{sgn}(L_n)\left|L_n\right| \tag{3}$$

where sign and magnitude are separately written. Sign is the hard decision of the bit and magnitude represents the reliability of that decision. From equation (2) and (3), we can write an expression for the probability of bit being decoded correctly given the LLR value L_n for that bit as

$$P_n = \frac{\exp\left(+\left|L_n\right|/2\right)}{\exp\left(+\left|L_n\right|/2\right) + \exp\left(-\left|L_n\right|/2\right)} \tag{4}$$

and the bit is decoded with an error probability of

$$P_{en} = \frac{\exp\left(-\left|L_n\right|/2\right)}{\exp\left(+\left|L_n\right|/2\right) + \exp\left(-\left|L_n\right|/2\right)} \tag{5}$$

It can be seen in (5) that higher the reliability value $\left|L_n\right|$, lower is the probability of error P_{en} in decoding of that bit. As $\left|L_n\right| \to \infty$ then $P_n \to 1$ and $P_{en} \to 0$. It shows

that if value of $|L_n|$ for a bit is large then there is very less chance of a bit being in error. This motivated the research in this direction to select a fraction of bits for the next iteration.

During the process of iterative decoding, the log likelihood ratio (LLR) values will grow in magnitude with iteration and hence will be decoded with very less probability of error. For P_{en} of the order of 10^{-4} the L_n value is close to 10. This in turn indicates that enhancing the L_n value for the n^{th} bit beyond certain value, say 10, may lead to some redundant iteration. Also, the reliabilities of all the bits will not increase to a high value simultaneously. So at every iteration we can divide the bit nodes into two groups, one group with sufficiently large reliability value and another group with less reliable values. In [11] [12] the reliability value of the first group are not updated. This may be attributed to the fact that they have sufficiently large magnitude of LLR value. However the reliability values for the bits in the second group need to be updated. By doing so, we save the computations that are required to process the more reliable bit values. This can be continued in each iteration till the code is decoded successfully.

4 Dynamic Selection of Bit Nodes

In our approach setting threshold is not required as appropriate value of threshold setting is not straight forward. A fixed number of nodes to be processed in each iteration are selected and at the end of each iteration, the bit nodes are arranged in increasing order of magnitude of their LLR values. From the sorted list only the fraction of nodes which are having small LLR value are processed further in subsequent iteration. The number of nodes processed at each iteration are kept constant from the point of view of the hardware implementation of the algorithm. The fraction of nodes to be processed during decoding in each iteration is optimized in order to have the performance almost similar to that of standard BP algorithm. The fraction is observed to be $1-R$ for a rate R LDPC code for 1 to 5 dB Eb/No by extensive simulation. Let us indicate the fraction of bit nodes processed (active) while decoding by α and call it as saving factor. The range of α is between 0 and 1. The value of 1 for α indicates that all the bit nodes are active and there is no saving in computations and we refer this condition as standard BP. Operational diagram of the LDPC decoder with Dynamic Selection Of Bit Nodes is shown in Figure 3.

All other notations are same as standard BP algorithm. The proposed algorithm is given below.Initialization:

Set $i=1$, and the maximum number of iterations to $I_{\max}$.

For each m, n, set $V_{mn}=U_{ch,n}$ and $V_n=U_{ch,n}$

Set saving factor $\alpha=1-R$ and find $nf=\lfloor \alpha N \rfloor$

Step 1: $[a]=\text{sort}\left(|V_n|\right)$ for all n in increasing order where a is vector of indices of Sorted elements.

Step 2: Horizontal step, for $1 \le n \le N$ and each $m \in M(n)$,

If $n \in a(1:nf)$ process

$$U_{mn} = 2\tanh^{-1} \prod_{n' \in M(n) \backslash n} \tanh \frac{V_{mn'}}{2}$$

end

Step 3 Vertical Step, for $n \in a(1:nf)$ and each $m \in M(n)$, process

$$V_{mn} = U_{ch,n} + \sum_{m' \in M(n) \backslash m} U_{m'n}$$

$$V_n = U_{ch,n} + \sum_{m \in M(n)} U_{mn}$$

Step 4: Hard decision and stopping criterion test:

(i) Create $\hat{w} = [\hat{w}_n]$ such that $\hat{w}_n = 1$ if $V_n < 0$ and $\hat{w}_n = 0$ if $V_n \ge 0$

(ii) If $H\hat{w} = 0$ or the maximum iteration number $I_{\max}$ is reached, stop the decoding iteration and go to Step 4.

(iii) Otherwise go to Step 1.

Step 5: Output $\hat{w}$ as the decoded codeword.

With dynamic selection of bit nodes algorithm, by setting the saving factor to α, it is possible to save the computations at both message node and check node by the same factor. However, there is an over head of sorting the bit node messages at the start of every iteration. It is known that the sorting algorithm complexity is of the order of $n \log n$ for sorting n elements. With this over head, we can save $2\alpha d_v n$ additions per iteration at the variable node processor. Also we save $2\alpha d_c M$ multiplications, $\alpha d_c M$ hyperbolic tangent evaluations and $\alpha d_c M$ inverse hyperbolic tangent evaluations. Apart from this, we also save an equivalent proportion of memory access due to fraction of nodes are inactive and whose values are not updated in a given iteration. This in turn saves time as well as power as most of the power is consumed in memory access. The latency in decoding is also reduced due to the fact that only fractions of bit nodes are updated and fractions of messages at the check node processor are computed. During simulation of an regular (3,6), code rate ½, block length 1008, we observed that standard BP decoded 8634 coded blocks where as the decoder with $\alpha = 0.5$ decoded 12771 code blocks and with $\alpha = 0.6$ it decoded 13613 coded blocks for the same processing time. This shows that there is considerable increase in throughput of the decoder.

5 Simulation Results

The proposed algorithm is used to find the BER performance for three different codes and the same are shown in Figures 4-6. The first code is a randomly generated (1008,3,6) LDPC code with code rate ½ . The encoded bits are BPSK modulated and transmitted over the AWGN channel. At the decoder side, belief propagation

algorithm based on tanh rule is used and number of iterations are set to a maximum of 10.From Figure 4, at a saving factor of 50% the code slightly outperforms standard BP algorithm. In fact at medium or short code lengths, the BP algorithm becomes sub optimum, due to existing correlation among messages passed during the iterative decoding. As the new algorithm processes only fractions of nodes in a given iteration, this reduces the level of correlation and hence improves the performance. However at a BER of 1×10^{-4}, the performance for 60% saving is inferior by 0.04 dB which for 70% saving becomes inferior by 0.19 dB as compared to standard BP.

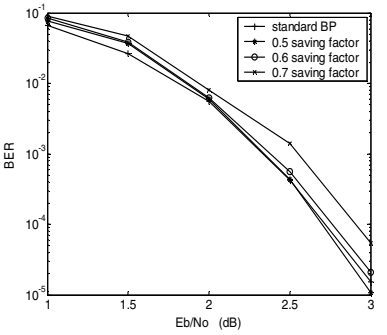


Fig. 4. BER performance of standard BP and DBS BP for $\alpha = 0.5, 0.6$ and 0.7 for $(1008, 3, 6)$

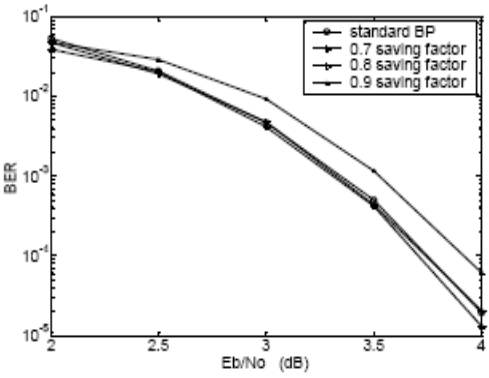


Fig. 5. LDPC code BER performance comparison between the standard BP decoding and decoding with various saving factor$\alpha = 0.7, 0.8$ and 0.9) for (724, 4, 16) structured BIBD based LDPC code

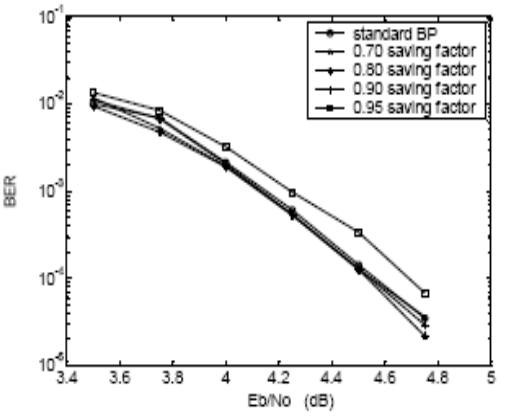


Fig. 6. BER performance comparison between the standard BP decoding and decoding with various saving factors (α=0.7,0.8,0.9 and 0.95) for (1810,3,10) structured BIBD based LDPC code

References

[1] Gallager, R.G.: Low-Density Parity-Check Codes. MIT Press, Cambridge (1963)
[2] MacKay, D.J.C., Neal, R.M.: Near Shannon limit performance of low density parity-check codes. Electron. Lett. 32(18), 1645–1646 (1996)
[3] MacKay, D.J.C.: Good error-correcting codes based on very sparse matrices. IEEE Trans. Inform. Theory 45, 399–431 (1999)
[4] Chung, S.Y., Forney Jr., G.D., Richardson, T.J., Urbanke, R.: On the design of low density parity check codes within 0.0045 dB of the Shannon limit. IEEE Commun. Lett. 5, 58–60 (2001)
[5] Luby, M.G., Mitzenmacher, M., Shokrollahi, M.A., Spielman, D.A.: Improved low-density parity ch eck codes using irregular graph. IEEE Trans. Inform. Theory 47, 585–598 (2001)
[6] Fossorier, M.P.C., Chen, J.: Near optimum universal belief propagation based decoding of low-density parity-check codes. IEEE Trans. Commun. 50(3), 406–414 (2002)
[7] Zhang, J., Fossorier, M., Gu, D., Zhang, J.: Two-dimensional correction for min-sum decoding of irregular LDPC codes. IEEE Commun. Lett. 10, 180–182 (2006)
[8] Zhang, J., Fossorier, M.: Shuffled belief propagation decoding. In: Proc. 36th Asilomar Conference on Signals, Systems and Computers 2002, vol. 1, pp. 8–15 (2002)
[9] Sharon, E., Litsyn, S., Goldberger, J.: Efficient serial message-passing schedules for LDPC decoding. In: Proc. Turbo-Coding Conference 2006 (2006)
[10] Wang, Y., Zhang, J., Fossorier, M., Yedidia, J.: Reduced latency iterative decoding of LDPC codes. In: Proc. 2005 IEEE Global Telecommunications Conference (2005)
[11] Fetweis, G., Zimmermann, E., Rave, W.: Forced convergence decoding of LDPC codes: EXIT chart analysis and combination with node complexity reduction techniques. In: Proc. 11th European Wireless Conference (2005)
[12] Bora, P.K., Zimmermann, E., Fettweis, G., Pattisapu, P.: Reduced complexity LDPC decoding using forced convergence. In: Proc. 7th International Symposium on Wireless Personal Multimedia Communications, WPMC 2004 (2004)
[13] Chi, Z., Wang, Z., Zhang, X.: Reduced Complexity Column Layered Decoding and Implementation for LDPC codes. IET Communications 5, 2177–2186 (2011)

Memory Optimized Design of Reciprocal Unit

Mahmad M. Nadaf, R.M. Banakar, and Saroja V. Siddamal

BVB College of Engg. & Tech.
Hubli
mahmadmn@gmail.com, banakar@bvb.edu, sarojavs@bvb.edu

Abstract. This paper presents the design of 32bit fixed point Q2.29 format reciprocal unit. The design is based on Newton-Raphson iteration method. The main contribution of the design is the initial values used in Newton-Raphson method is computed on the fly without storing them in a look up table which occupies block memory. All the values given to reciprocal unit are scaled from 1 to 2 ranges for reduced complexity in design and implementation. The design is tested on Xilinx Virtex-5 with target device XC5VTX240T which includes package FF1759 and speed -2. Using Xilinx Virtex-5 the design unit illustrates routing delay of 3.972ns and logic delay of 51.043ns.

1 Introduction

The division operation plays a vital role in many DSP applications. Designing a high-speed division is very much essential. The division operation (r) can be expressed as the product of the dividend (x) and reciprocal of the divisor (y),

$$r = \frac{x}{y} = x \times \left(\frac{1}{y}\right) \tag{1}$$

Most of the available reciprocals are based on floating point numbers using IEEE 754 format. This format is extensively used in manual calculations and representation but the fixed point data type is widely used in digital signal processing (DSP) and game applications, where performance is sometimes more important than precision. Hence an optimized design technique for reciprocal unit improves speed and accuracy.

The reciprocal unit can be computed using multiplicative iterative methods. We surveyed for different methods of multiplicative iterative methods such as Newton-Raphson [11], Taylor series [9] expansion, Secant method [1] and Brent method [2]. We found Newton-Raphson method solves non-linear equations based on linear model and needs only one initial assumption. Hence this process is fixed for iterative calculations.

There are two stages in reciprocal calculation, first the initial assumption and the second Newton-Raphson iteration. Look-up table solutions for the initial value approximation are common [8], but by nature they need quite large memories if high accuracy combined with low iteration counts. The initial approximation is obtained by various techniques [5] which includes linear approximation, direct look up table methods and table look up followed by multiplications. In order to reduce the look-up table memory-size, [2, 7, 6] introduces various technique. In these methods precision of the approximation increases (greater than 16 bits) and the size of the memory

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 339–345.
springerlink.com

required to implement the table look-up table also increases. Most frequent used table method is bipartite table method. This bipartite table method was first reported by Das Sarma and Matula [2], where the binary input is split into three separate k bit numbers, with k=n/3. This bipartite table method occupies 4352 Kbytes of memory to store the initial approximation. With little technical improvements a new initial approximation table method called symmetric bipartite table method (SBTM) was introduced. The SBTM method was enhanced and new table method called MBTM was used in [7] which occupy 384Kbytes of block memory for storing seed values. The MBTPA method used in [6] requires less memory compared to other table look up methods. This method uses 45.8Kbytes of memory to store the initial approximation values. In this paper we have proposed a reciprocal unit design which uses no block memory to store seed values and this seed values are calculated on-the fly. Q(2.29) fixed point format is used in our design approach.

The remainder of this paper is organized as follows. Section 2 gives initial assumption techniques. Section 3 describes Newton-Raphson iteration method. Section 4 looks at the Hardware Implementation. Section 5 presents Testing and results. Finally section 6 summarizes most important conclusions and the future work for this reciprocal unit.

2 Initial Assumption Technique in Heuristic Approach

The process of obtaining initial assumption value plays very important role in reciprocal design. There is no look-up-table generated for accessing initial guess. Rather we use heuristic method to find initial value on the-fly. To reduce complexity and design we consider the reciprocal unit input range from 1 to 2. We can have a single initial guess and in order to reduce Newton-Raphson iterations and to get accurate result we divided the 1 to 2 range into six sub divisions and each sub division has a distinct initial value. Assignment of initial value to sub divisions is based on bit position of fractional part of input numbers.

Since the range is 1 to 2 there is only possibility of integer part being 01_b and rest bits belongs to fraction part. Suppose if the integer bits be 10_b or higher than this value, it is considered to be out of range and that number is scaled by suitable scalar to bring the input number into the required range. The table 2.1 gives the input (x) and output range for respective subdivisions. Using this seed value we get exact reciprocal output in maximum 3 Newton-Raphson iterations.

2.1 On-the Fly Computation of Initial Assumption Values

When the design work started, we took a single value as the seed value and used it in the Q2.29 fixed point format input range of 1 to 2. We redesigned the heuristic approach and sectioned 1 to 2 input range into a total of 6 segments as shown in Table 1.

Table 1. Input and output range for designed reciprocal unit

Segment no.	Input range(x)	Output range(1/x)
1	1.750000-1.999999	0.502500-0.5714280
2	1.500000-1.740000	0.5714280-0.666666
3	1.250000-1.490000	0.6666665-0.800000
4	1.125000-1.240000	0.8000000-0.888888
5	1.062500-1.124000	0.8888889-0.941170
6	1.031250-1.062400	0.941180-0.9696969

From the above table it reveals that input range is geometrically divided into six parts for better performance.

In the next section we describe Newton-Raphson iterative method.

3 Newton-Raphson Iterative Method

The Newton-Raphson iteration method is most widely used [11] extrapolation method for solving non-linear equations. It uses a single (x_0) initial guess value close to the root (α) of the given non-linear equation f(x). Further iterations produce much better approximation to root value by converging towards root value for each iteration. We get x_1 after first iteration, then x_1 is much nearer to root α. Then we have,

Slope of tangent= $f(x) \div (x_0 - x_1) = f'(x_0)$

$$(x_0 - x_1) = f(x_0) \div f'(x_0) \tag{2}$$

$$x_1 = x_0 - (f(x_0) \div f'(x_0)) \tag{3}$$

And up to n terms,

$$x_{n+1} = x_n - (f(x_n) \div f'(x_n)) \tag{4}$$

As n $\rightarrow \infty$, $x_n \rightarrow 0$.

In order to compute reciprocal,

Consider $$f(x) = (1 + x) - c \tag{5}$$

Initial guess as x_0 and the priming factors are,

$$f(x_i) = 1 + x_i \tag{6}$$

And $$f'(x_i) = -(1 + x_i^2) \tag{7}$$

Simplifying we get, $$x_{i+1} = 2x_i - x_i^2 \times c \tag{8}$$

This can be implemented on hardware. The hardware processing of Newton-Raphson iteration doubles the accuracy in each iteration. This requires one squarer, one multiplier, one shifter and a subtraction unit.

3.1 Error Analysis for Newton-Raphson Method

Let $\varepsilon_i = (1 + c) - x_i$ be the error at the i^{th} iteration in the Newton-Raphson method.

$$\varepsilon_{i+1} = (1 + c) - x_{i+1} = (1 + c) - x_i(2 - x_i \times c) \tag{9}$$

The above equation can also be expressed as

$$\varepsilon_{i+1} = c(1 + c - x_i)^2 = c\gamma^2 \tag{10}$$

Where 'γ' is a constant and $\gamma = (1 \div c - x_i)$

From equation (10), we can conclude that the absolute error degrades quadratically in each Newton-Raphson iteration as the i^{th} error is proportional to the square of $(i\text{-}1)^{th}$ error.

4 Memory Optimized Reciprocal Unit

The following Fig. 1 shows the architecture of reciprocal unit. The architecture can be looked as two stages. In the first stage initial assumption i.e., seed value is found by above mentioned techniques.

The second stage includes Newton-Raphson extrapolation method. During the first iteration the input to the 2:1 MUX goes from initial seed value and for second and third iterations input to the MUX will be output of first and second iterated value. As mentioned above sections, the seed value is found on-fly without using any look-up-tables. Hence this architecture does not require any ROM memory and hence reduces complexity. Hence comparing with [4, 8] memory, power used and area also reduces drastically.

The seed value is copied to a register and shifted it left by one bit to get double of the input number. The squared output is truncated to 32 MSB bits (63:32) to retain precision. In next operation the input value is multiplied with squared output value and the result is truncated to 32 MSB bits. This result is subtracted from the one bit shifted seed value to get reciprocal. But for second and third iterations the input given to Newton-Raphson extrapolative unit is output of its previous iterated value. Finally after 3 Newton-Raphson iterations the design gives the 31 bit accurate reciprocal.

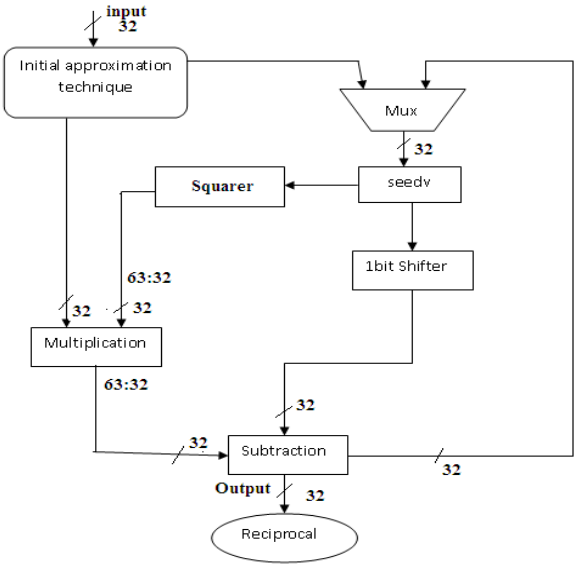


Fig. 1. Architecture of reciprocal unit

5 Functional Verification

In this section we will present the testing of our design for accuracy and results obtained from our design.

5.1 Testing the Reciprocal Unit

We tested the reciprocal unit for individually with number of random test vectors as well as some thoughtful ones. We even made available the facility of counter test by using a C-language program to provide standard output.

The Fig. 2 compares the system model results with designed hardware reciprocal unit results. Y axis shows the reciprocal values and X axis illustrates the input values in the range 1 to 2 for which the reciprocals are computed.

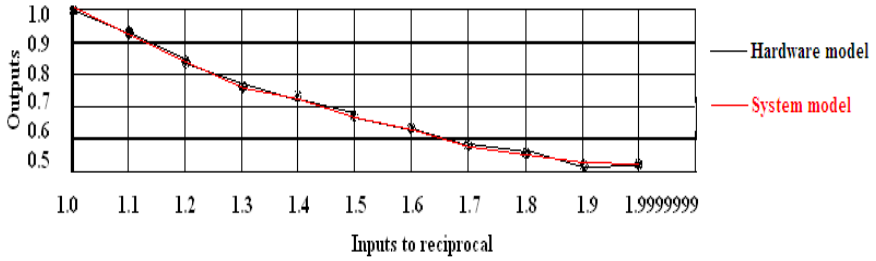


Fig. 2. Comparison of hardware reciprocal model with the system model

6 Result Analysis for Reciprocal Unit Design

The accuracy, latency and memory effectiveness of our method of hardware modeling of the reciprocal unit is demonstrated in this section.

The reciprocal unit design is implemented using Xilinx Virtex-5, XC5VTX240T device with package FF1759 and speed -2. The implementation occupies 116 Slice LUTs out of 149760 and all these LUTs are used as Logic, 65 IOs out of 680, 1 BUFG out of 32, 116 number of Flip Flop pairs used, 24 DSP48Es out 96. It gives reciprocal output in 55.015ns with 7.2% routing delay.

The Table 2 compares this design with different similar reciprocal designs which shows increase in performance. This table also has the details of the performance matrices used for comparing earlier design [2, 6, 7] which used look-up table method for initial assumption. We have used this reciprocal design in seismic migration lateral depth migration model.

Table 2. Comparison of hardware design with similar designs

Method	MBTPA [6]	MBTMA[7]	BTMA [2]	This work
Latency	<3	<2	<2	<1
Accuracy	2^{-28}	2^{-27}	NA	2^{-29}
Memory size	45.8 KB	384KB	4352KB	Null
Data format	Double precision IEEE 754 format			Q(2,29) fixed point format

7 Conclusion and Future Work

7.1 Conclusion

We have successfully designed the reciprocal unit in Xilinx Vertex5. The design uses no ROM memory to store initial values as they are calculated on-fly without any look-up-table. This design reduces area occupied by reciprocal unit design compared to [4] which using look-up-tables. And it is made to work on Q2.29 fixed point format which is extensively used in most of the DSPs.

7.2 Future Work

The division [3] operation can be implemented using the present reciprocal by including one multiplier.

References

[1] Brent, R.P.: Some Efficient Algorithms for solving Systems of non- linear Equations. SIAM J. Numer. Anal. 10, 327–344

[2] Das Sarma, D., Matula, D.W.: Faithful bipartite rom reciprocal tables. In: Knowles, S., McAllister, W.H. (eds.) Proceedings of the 12th IEEE Symposium on Computer Arithmetic, UK, pp. 17–28 (1995)
[3] Ferrari, D.: A Division Method Using a parallel Multiplier. IEEE Trans. Elctron. Comput. EC-16, 224–226 (1967)
[4] Chen, D., Zhou, B., Guo, Z., Nilsson, P.: Design and Implementation of reciprocal unit, pp. 1318–1321. IEEE Press (May 2005)
[5] Agrawal, G., Khandelwal, A.: A Newton Raphson Divider Based on Improved Reciprocal Approximation Algorithm. High Speed Computer Arithmetic EE328N (2006)
[6] Choo, I., Deshmukh, R.G.: An accurate linear approximation method utilizing a bipartite reciprocal table for a floating point divider. In: 2001 Canadian Conference on Electrical and Computer Engineering, vol. 2, pp. 1199–1204 (August 2002)
[7] Stine, J.E., Schulte, M.J.: Approximating Elementary Functions with Symmetric Bipartite Tables. IEEE Trans. Computers 48(8), 842–847 (1999)
[8] Kucukkak, U., Akkas, A.: Design and Implementation of reciprocal Unit using table look-up and Newton-Raphson iteration. In: Euromicro Symposium on Digital System Design, DSD 2004, August 31-September 3, pp. 249–253 (2004)
[9] Geers, M.G.D.: Enhanced solution control for physically and geometrically non-linear problems. International Journal for Numerical Methods in Engineering 46, 205–230 (1999)
[10] Schulte, M.J., Stine, J.E., Wires, K.E.: High-speed reciprocal approximations. In: Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers, vol. 2, pp. 1183–1187 (November 1997)
[11] Yoma, T.J.: Historically Development of the Newton-Raphson Method 37(4), 531–551 (1995)

Enhanced LZW Algorithm with Less Compression Ratio

Amit Setia and Priyanka Ahlawat

Department of Computer Engineering
National Institute of Technology, Kurukshetra, India
{amitsetia50,mannpammy}@gmail.com

Abstract. LZW is the one known compression algorithm appropriate for communication. LZW data compression algorithm is popular for data compression because it is an adaptive algorithm and achieves an excellent compromise between compression performance and speed of execution. LZW is a dictionary based data compression algorithm, which compress the data in a lossless manner so that no information is lost. LZW algorithm requires no extra communication from the encoder to the decoder. In this paper we present a scheme to eliminate the spaces from the data so that high compression factor achieves.

Keywords: compression, adaptive, encoding, decoding.

1 Introduction

LZW [Welch 1984] is a popular variant of LZ78 [Ziv and Lempel 1978], developed by Terry Welch in 1984. It is a dictionary based adaptive algorithm. The first 256 entries are occupied in the dictionary before any data is input. Most of the data file to which we want to compress contains many spaces. By eliminate these spaces from the data file, results in high compression factor or less compression ratio.

The rest of the paper is organized as follows. Section 2 reviews the basic concepts of compression and encryption. Section 3 presents the compression techniques. Section 4 describes the proposed system. Section 5 deals with results. Section 6 concludes the paper. Section 7 presents the future work.

2 Basic Concepts

There are various techniques to compression.

Compression Techniques:

a) LZW compression
b) Huffman coding
c) Arithmetic coding
d) Run length encoding
e) LZ 77
f) LZ 78 and so on.

In our system we are using LZW technique.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 347–351.
springerlink.com

3 Compression Technique

3.1 LZW Compression

It is a dictionary based algorithm. It is a variant of Lempel Ziv 78 compression algorithm. In this we initialize the dictionary to all symbols in alphabet. In common case of 8-bit symbols, first 256 entries of the dictionary 0 through 255 are occupied before any data is input. LZW token consist a pointer to the dictionary. When the dictionary is initialized, the next input character finds in dictionary.

The idea of LZW is that the encoder input symbols one by one and accumulates them in string I. After each symbol is input and is concatenated to I, dictionary is searched for string I. As long as string I is found in dictionary, the process continues. But at some point if adding text symbol suppose x causes the search to fail, string I is in dictionary but string Ix is not. At this point the encoder outputs the dictionary pointer that points to string I, and saves string Ix in the next available entry in the dictionary and then initialize string I to symbol x.

3.1.1 LZW Encoding Algorithm

Algorithm:

Step 1 Initialize dictionary to contain single character string.
Step 2 Read first input character prefix string ω from the data file.
Step 3 Read next input character k from the data file.
a) If no such k (input exhausted): output:=code(ω):Then EXIT
b) If ωk exists in dictionary: ω: = ωk; Repeat Step 3.
c) Else If ωk not in dictionary. Then output :=code (ω)
Dictionary: = ωk;
ω :=k;
Repeat step 3.
Step 4 End

3.1.2 LZW Decoding Algorithm

Algorithm:

Step 1 Read first input code and CODE=OLD code=input code with CODE=code (k), output=k, Fin char=k.
Step 2 Read next input code, CODE =INCODE=next input code.
If no new code: EXIT. Else:
Step 3 If CODE=code (ωk): stack = k
CODE: =code (ω)
Repeat Step 3
Else if CODE = code (k): output=k, Fin char=k.
Do while stack not empty:
Output = Stack top, Pop stack.
Dictionary=OLD code, k.
OLD code=IN code;
Repeat step 2.
Step 4 End

3.2 Advantages of LZW Algorithm

1. LZW data compression algorithm is an adaptive and very effective means to save storage space and network bandwidth.
2. LZW algorithm is easy to implement.
3. It is a lossless compression algorithm, so no loss of information is there.

4 Proposed Scheme

In LZW algorithm, dictionary size is determined by the general unary code.

Table 1.

n	a=(3+n*2)	nth codeword	no. of code words	Range of integers
0	3	0xxx	2^3 =8	0-7
1	5	10xxxxx	2^5 =32	8-39
2	7	110xxxxxxx	2^7 =128	40-167
3	9	111xxxxxxxxx	2^9 =512	168-679
		Total	680	

Now in LZW algorithm if we make dictionary of 680 entries. Each entry occupy 9 bits up to 680 entries, And if we want to increase dictionary size up to 2044 entries then each entry occupy 10 bits.

But in our proposed scheme we eliminate the spaces from the input data file. In our scheme if we make dictionary of 680 entries, Each entry occupy 10 bits, i.e. 1 bit higher than those in case of LZW encoding algorithm, 9bit data plus one parity bit used for checking the next character is space or not. If parity bit is set then next character is space otherwise not. But we save the 9 bits which is used for space. Now there are number of spaces in data file say n, Now we save n*9 bit space. But some extra parity bits are also sent, but many words in data file have small length. So spaces between these words occupy lot of space, so by eliminate these spaces we can achieve high compression.

4.1 Enhanced LZW Encoding Algorithm

Algorithm:

Step 1 Initialize dictionary to contain all 0 to 255 single character string.
Step 2 Read first input character prefix string ω from the input data file.

Step 3 Read next input character says k from the input data file.
a) If no such k (input exhausted): output:=code(ω):Then EXIT
b) If k=space, then set M.S.B bit of ω equals to 1, Repeat step 3.
c) If ωk exists in dictionary: ω: = ωk; Repeat Step 3.
d) Else If ωk not in dictionary. Then output :=code (ω)
Dictionary: = ωk;
ω :=k;
Repeat step 3.
Step 4 End

4.1.1 Enhanced LZW Encoding Algorithm Description

In step no. 3 (b) we set the parity bit or M.S.B bit of suffix ω. that helps in decoding site.

4.2 Enhanced LZW Decoding Algorithm

Algorithm:

Step 1 Read first input character ω, left shift the input character ω by 1. ie ω <<1. And result is stored in the carry bit. Then right shift the input character ω by 1. ie ω >>1. And CODE=OLD code=input code with CODE=code (k), output=k, Fin char=k.
Step 2 If carry bit = 1
Then next input character k = space.
Go to step 3.
Else Read next input character k from the input data file, left shift the input character k by 1. ie k <<1. And result is stored in the carry bit. Then right shift the input character k by 1. ie k >>1.
CODE = INCODE = next input code
If no new code: EXIT. Else:
Step 3 If CODE=code (ωk): stack = k
CODE: =code (ω)
Repeat Step 3
Else if CODE = code (k): output=k, Fin char=k.
Do while stack not empty:
Output = Stack top, Pop stack.
Dictionary = OLD code, k.
OLD code = IN code;
Repeat step 2.
Step 4 End

4.2.1 Enhanced LZW Decoding Algorithm Description

In step 1 we first left shift the character ω by 1 and result stored in carry bit. And now for retrieving actual data we right shift the character by 1.

5 Results

Table 2.

File size	Compression ratio with LZW compression	Compression ratio with enhanced LZW compression
5.7 kb	0.50	0.47
10 kb	0.61	0.53

We have taken two files one is of 5.7 kb and another is of 10 kb. When we compressed first file which is of 5.7 kb using LZW, the compression ratio of 0.50 achieved. But when we compressed same file using enhanced LZW then we achieved compression ratio of 0.47. In second case when we compressed another file of 10 kb size using LZW algorithm then we achieved compression ratio of 0.61. But when we compressed same file using enhanced LZW we achieved compression ratio of 0.50.

But compression ratios in case of enhanced LZW vary data to data.

6 Conclusion

The work present in this paper can be summarized as:

In this paper we accomplished a system which eliminate spaces from the data file so that we can achieve high compression.

7 Future Work

The proposed system can be combined with cryptography and steganography techniques for more security, which encrypt and hide the existence of the information.

References

[1] Jiang, J.: Pipeline algorithm of RSA data encryption and data compression. In: IEEE Proceeding of International Conference on Communication Technology, vol. 2, pp. 1088–1091 (1996)

[2] Kruse, H., Mukherjee, A.: Data compression using text encryption. In: IEEE Data Compression Conference, p. 447 (1997)

[3] Lin, M.B., Chang, Y.Y.: A new architecture of a two-stage lossless data compression and decompression algorithm. IEEE Transaction on VLSI Systems 17, 1297–1303 (2009)

[4] Welch, T.: A technique for high performance data compression. IEEE Computer 17, 8–19 (1984)

[5] Ziv, J., Lempel, A.: A universal algorithm for sequential data compression. IEEE Transaction on Information Theory 23, 337–343 (1977)

[6] Ziv, J., Lempel, A.: Compression of individual sequences via variable length coding. IEEE Transaction on Information Theory 24, 530–536 (1978)

Design of High Security and Performance System for Storage Devices Using AES

Vinodkumar I. Bellikatti[1], Chetan S.[2], Shivaputra[2], and Kushal K.S.[1]

[1] Dept. of M.Tech [VLSI Design & Embedded System],
Dr. Ambedkar Institute of Technology
Bangalore, India
vbellikatti6@gmail.com
[2] Dept. of Electronics & Communication,
Dr. Ambedkar Institute of Technology,
Bangalore, India
chetans31@gmail.com

Abstract. "All our dreams can't be translated into reality. But they can act as foundation stone for our glorious future"

As the technology of communication and storage improves, it needs high security and performance in both software & hardware. This paper describes the design of effective security system for implementing it to encrypt or decrypt the data in storage device. In this paper we are using O'Driocells matrix for mapping & inverse mapping that is used in S-Box calculation which reduces the total no of 1's to 51 which is less than previously published paper[5] and Mixcolumn is implemented using Combinational logic method instead of Xtime look up table[LUT]. Hence it is effective in terms of speed, low power and high performance. We proposed pipelined AES architecture that can offer low power and high throughput to increase the efficiency.

Keywords: AES, Encrypt/decrypt, FDE, ATM switch.

1 Introduction

The Advanced Encryption Standard (AES), standardized by NIST, National Institute of Standards and Technology, is a cryptographic algorithm replacement to DES (Data Encryption Standard) algorithm [1] ,[2] as the federal standard to protect sensitive information. AES has already received widespread use because of its high security, high performance in both hardware and software. Many implementations are done in software but it seems to be too slow for fast applications such as routers and some wireless communication systems. AES is a 128 bit symmetric data block cipher with 128, 192 or 256 bits key. The data block was described by arrays of bytes in 4 x 4 matrix (Called "State") and it has four basic steps operation as see in Fig .l: Sub Bytes, (or S-Box), ShiftRow, MixColumn, and AddRoundKey. These four steps are also known as layers. The four layer steps describe one round of the AES. Number of rounds is made vary according to the key size. The AES with 128-bit key size operates iteratively on those four basic steps for 10 rounds.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 353–359.
springerlink.com

2 Related Work

2.1 A FPGA Design of AES Core Architecture for Portable Hard Disk

This paper describes a high effective AES core hardware architecture for implementing it to encrypt/decrypt the data in portable hard disk drive system that apply to effectively in the terms of speed, scale size and power consumption to comply with minimum speed of 5 Gbps (USB3.0). We proposed the 128 bits data path of two different AES architectures design, Basic Iterative AES, which reuses the same hardware for all the ten iterations and, One Stage Sub Pipelined AES, with one stage of outer pipelining in the data blocks that both of them are purely 128 bits data path architecture that different from the previous public paper. The implementation result on the targeted FPGA, the basic iterative AES encryption can offer the throughput of 3.85 Gbps at 300 MHz and one stage sub pipelined AES can offer the throughput to increase the efficiency of 6.2 Gbps at 481 MHz clock speed.

3 System Architecture

The One Stage Sub-pipeline structure in Fig.1(Data Block#l and #2).1t is an improvement of the basic iterative architecture with respect to speed. It has just one stage of pipeline within the data block. The data block is replicated once. In One stage the same work load is shared by two data blocks.

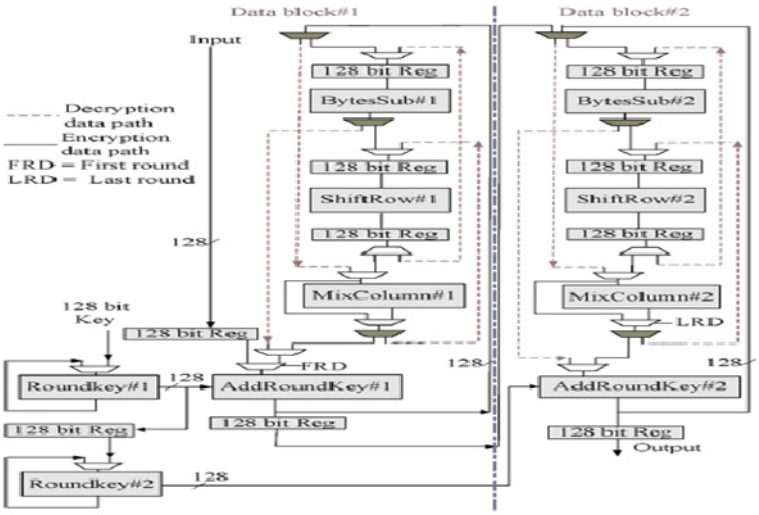


Fig. 1. A 128 bits data path of one stage sub pipelined AES

3.1 S-Byte Calculation

The S-Box operation is a non-linear byte substitution. It composes of above transformations as shown in figure 2.

1) Modular inversion in GF(2^4) - this stage is to compute B(x) = A^{-1}(X) for an 4-bit input word (in GF(2^4) where m(x) = x^4 +x+1 is taken as a field polynomial; {OO} is mapped to itself).

2) Affine Transformation: This sub-step is performed in GF(2) and defined by. D(x) = δ*B(x)$_{mod(x^8+1)}$ ϕ C(x) where b = {IF} = x^4 + x^3 + x^2 + x +1 for the encryption process and b = {4A} = x^6 + x^3 + x for the decryption.

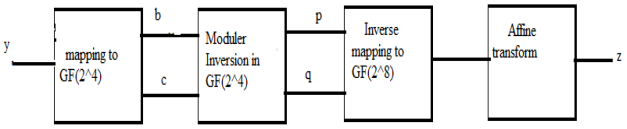


Fig. 2. Sub bytes design flow for composite field Inversion

Working in the composite field, multiplicative inverse is leisured. However, forth and back, we have to map elements in GF(2^k) into GF(2^n)^m where k = mn. Therefore both transform and inverse transform matrices are needed. Elements in GF(2^8) can be mapped to element in GF(2^4)^2 by using the polynomial r(x) = x^2 + x + β^{14} where β^{14} denotes the element in GF(2^4) of which I (x) = x^4 + x + I is the primitive irreducible polynomial. The resulted mapping and inverse mapping matrices are given in eqn. (2) and (3) respectively.

$$M=\begin{bmatrix} 1&0&1&0&0&0&1&0\\ 0&0&0&0&1&1&0&0\\ 1&1&0&1&0&0&1&0\\ 1&0&1&0&0&0&0&0\\ 1&1&0&1&1&0&0&1\\ 0&1&0&0&0&1&0&0\\ 1&0&1&0&0&1&1&0\\ 1&0&1&1&0&1&1&0 \end{bmatrix} \quad IM=\begin{bmatrix} 0&0&1&1&0&1&0&1\\ 1&0&0&0&0&1&1&0\\ 0&0&1&0&0&1&0&1\\ 0&0&0&0&0&0&1&1\\ 1&1&0&0&0&0&1&0\\ 1&0&0&0&0&0&1&0\\ 1&0&0&1&0&0&0&0\\ 0&1&1&1&1&0&1&0 \end{bmatrix} \qquad (2,3)$$

The upper-left element in the above matrices denotes the least significant bit. An advantage of mapping elements form GF(2^8) to GF((2^4)^2) is the simpler multiplicative inverse computation since inversion is performed in GF(2^4). For such a small field size, inversion using either the direct truth table mapping or table look up consumes small area. Moreover, in Rijndae1 system data are treated naturally in byte format. Let data (byte) be expressed as A = {bc} = bx + c , the inversion of A, say B = A^{-1} = {pq} = px + q. For the field polynomial r(x) = x^2 + Cx + D , one can have where

$$p = b\Delta^{-1} \tag{4}$$

$$q = (Cb \oplus c)\Delta^{-1} \tag{5}$$

$$\Delta = c(Cb \oplus c) \oplus b^2 D\,. \tag{6}$$

$$\Delta = bcC \oplus c^2 \oplus b^2 D \tag{7}$$

For GF($(2^n)^2$) , the polynomial in the form of $r(x) = x^2 + x + A$ always exists [12]. As such, C and D can be set to {1} and {9} (in GF(2^4)) respectively. Fixed-coefficient multiplication (i.e.,b^2 D) as well as squaring units are relatively simple according to their small field size. The multiplications required in computing eqn. (4), (5) and (6) can be done straight away in GF(2^4) or can be further simplified by making use of composite field GF($(2^2)^2$) [7].

3.3 ShiftRow Transformation

The ShiftRow process operates on individual row with individual offset byte of state. In the state arrangement, data are fed into a square matrix in row order. To operate the ShiftRow transformation, we need register#l to store the whole data before byte swapping. This can result in the unsmooth data flow. However, the implementation is not very difficult. Due to we have designed the ShiftRow transform throughput is 128 bits per clock cycle for support the high throughput of hard disk. We used register#2 to be a pipeline arrangement (see Fig. 3 below) such that the data are arranged in order and ready for the following operation, MixColumn transformation.

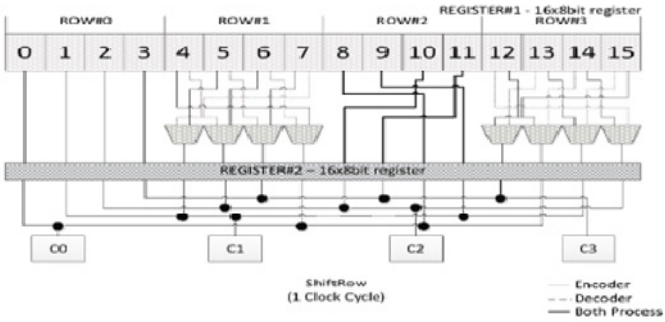


Fig. 3. ShiftRow and Inverse ShiftRow (dash line)

3.4 MixColumn Transformation

The mix column transformation operates on each column individually. Each byte of a column is mapped into a new Value that is a function of all four bytes in that column. The transformation can be expressed by the following matrix Multiplication on State.

$$\begin{bmatrix} 02 & 03 & 01 & 01 \\ 01 & 02 & 03 & 01 \\ 01 & 01 & 02 & 03 \\ 03 & 01 & 01 & 02 \end{bmatrix} \begin{bmatrix} c_{0,0} & c_{0,1} & c_{0,2} & c_{0,3} \\ c_{1,0} & c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,0} & c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,0} & c_{3,1} & c_{3,2} & c_{3,3} \end{bmatrix} = \begin{bmatrix} c'_{0,0} & c'_{0,1} & c'_{0,2} & c'_{0,3} \\ c'_{1,0} & c'_{1,1} & c'_{1,2} & c'_{1,3} \\ c'_{2,0} & c'_{2,1} & c'_{2,2} & c'_{2,3} \\ c'_{3,0} & c'_{3,1} & c'_{3,2} & c'_{3,3} \end{bmatrix}$$

Each element in the product matrix is the sum of products of elements of one row and one column. In this case, the individual additions and multiplications are performed in

GF (2^8),The MixColumn transform of an AES can be expressed as C'(x)= C(x)a(X)mod(x4+]) and each column is considered as a polynomial with coefficients Ci,c define in GF(2^8).

Multiplication is implemented using Combinational logic method by following steps.

1) If the low bit of B is set, XOR the product p by value of A
2) Keep track of weather the high bit of A is set to one
3) Rotate A one bit of left & discarding the high bit and making the low bit to have a value of zero
4) If A's high bit have a value of one prior to their rotation, XOR A with the hexadecimal number 0x1b
5) Rotate B one bit to the right, discarding the low bit and making the high bit have value of zero

3.5 AddRoundKey, KeyScheduling Transformation

The KeyScheduling expands the initial 128-bit cipher keys to generate the round keys. The two methods for the key expansion are commonly used, i.e., the round keys can be generated on-the-fly with the data transformation, or they are pre-calculated and stored for later use. In this paper, the round keys applied to the data transformation for encryption or decryptions are calculated on-the-fly. The agility of the key expansion deals with the situation that the cipher keys are changed frequently. The implementation of the key generation for encryption is illustrated in Figure 4. Every word (32 bits) of the next state is the XOR of the current word in the same position with its left neighboring word. For example, the word in C1 is calculated as w'[C_1] = w [C_1] + w [C_0]. For words in the position Ch its neighboring word is in position C3. A transformation Rot Word is applied

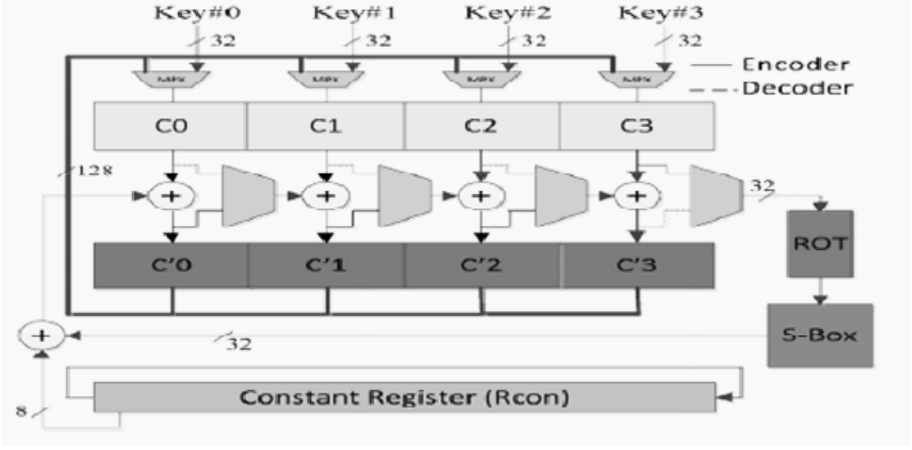


Fig. 4. KeyScheduling Structure

to the word in position C3 prior the XOR, followed by an XOR with the round constant RCon. So, we need two 128 bit register to store round keys. The Rot Word register is a circular word shift register. And The Rcon is a feedback word shift register. For the one stage sub pipe1ined AES structure, we have two different KeyScheduling modules which share the load of ten iterations. The KeyScheduling#1 generates the first five keys and the KeyScheduling#2 generates the last five keys.

4 Implementation Details

The core system itself as well as the other necessary circuits are designed and implemented using ***Verilog.*** The code is ***Verilog,*** which could be easily implemented on targeted FPGA devices, without changing the design. Synthesis and Place & Route are achieved on Xilinx ISE 13.1i with Xilinx's device family. This is used for writing, debugging and optimizing efforts, and also for fitting, simulating.

5 Conclusion

The proposed Security System can be chosen upon speed or throughput requirement for supporting storage device data encryption. In this system S-Box makes use of O'Driocells matrix for mapping and inverse mapping. The mix column can be implemented with combinational logic. It is more efficient than previous implementation, in terms of speed, power and throughput. Hence it can be effectively used in communication system to encrypt/decrypt the data. A synthesizable ***Verilog*** code is developed for the implementation of both encryption and decryption process. The design is verified via the FPGA implementation with Xilinx family.

References

[1] Daemen, J., Rijmen, V.: AES Proposal: Rijndael (Version 2). NIST AES, `http://csrc.nist.gov/publications/`, `http://csrc.nist.gov/CryptoToolkit/aes/rijndaellRijndaelammended.pdf`

[2] NIST Federal Information Processing Standards (FIPS PUB 197) Advanced Encryption Standard (November 2001), `http://www.nist.gov/aes`

[3] CAST, Advanced Encryption Standard Core, `http://www.cast-inc.com/cores/aes/index.shtml`

[4] Elbirt, A., Yip, W., Chetwynd, B., Paar, C.: An FPGAbased performance evaluation of the AES block cipher candidate algorithm finalists. IEEE Trans. of VLSI Systems 9(4), 545–557 (2001)

[5] ThongKhome, K., Thanavijitpun, C.: FPGA Design of AES C Architecture for Portable Hard Disk

[6] Kim, C.H.: Improved Differential Fault Analysis on AES Key Schedule

[7] Wong, M.M., Wong, M.L.D.: A High Throughput Low Power Compact AES S-box Implementation using Composite Field Arithmetic and Algebraic Normal Form Representation

[8] Rijmen, V.: Efficient implementation S-box, http://ftp.comms.scitech.susx.ac.uk/fft/crypto/rijndaelsbox.pdf
[9] Jutla, C., Kumar, V., Rudra, A.: On the Complexity of Isomorphic Galois Field Transforms. IBM Research Report, vol. RC22652, W0211–W0243 (November 2002)
[10] Chantarawong, S., Noo-intara, P., Choomchuay, S.: An Architecture for S-Box Computation in the AES. In: Proc. of Information and Computer Engineering Workshop 2004 (ICEP 2004), pp. 157–162 (2004)
[11] Hodjat, A., Verbauwhede, I.: Minimum Area Cost for a 30 to 70 Gbits/s AES Processor
[12] Jing, M.-H., Chen, J.-H., Chen, Z.-H.: Diversified MixColumn Transformation of AES

Implementation of Lifting Scheme Based DWT Architecture on FPGA

Naagesh S. Bhat

M.S. Ramaiah School of Advanced Studies, Bangalore
bnsnagesh@gmail.com

Abstract. The JPEG (Joint Photographic Experts Group) 2000 encoder is an entirely hardware implementation of a JPEG 2000 compression codec that is based on the ISO/IEC 15444-1 standard. The JPEG 2000 standard, finalized in 2001, defines a new image-coding scheme using state-of-the-art compression techniques based on wavelet technology. Its architecture is useful for many diverse applications, including Internet image distribution, security systems, digital photography, and medical imaging. In this paper, we propose an efficient VLSI architecture for the implementation of one-dimension, lifting scheme based discrete wavelet transform (DWT). Both of the folded and the pipelined schemes are applied in the proposed architecture. The architecture has been coded in Verilog HDL, and then verified successfully by the platform of Xilinx 10.1 on Virtex-4 device.

1 Introduction

There has been a long history in the development of wavelet transform [1]. Discrete wavelet transformation is now adopted to be transform coder in both JPEG 2000 [2] still image coding and MPEG-3 [3] still texture coding. In this paper, we mainly focus on the design of the DWT core for JPEG 2000.

JPEG 2000 is the emerging next generation still image compression standard. With the inherent features of wavelet transform, it provides multi-resolution functionality and better compression performance at very low-bit rate compared with the DCT based JPEG [4] standard. Discrete wavelet transform (DWT) has been widely used in many different fields of audio and video signal processing. Recently, DWT is being increasingly used as effective solutions to the problem of image compression. One well-known example is that DWT has been adopted by the JPEG2000, one of the several popular image compression standards defined by the Joint Picture Expert Group (JPEG), due to the efficient decomposition of a signal into several components (subbands) with DWT. In general, DWT can be implemented by direct convolution and several DWT architectures implemented by filter convolution have been proposed. The 5/3 reversible and 9/7 irreversible filters are chosen for lossless compression.

This paper organization is as follows. In Section 2, the lifting scheme algorithm is described and compared with classic implementation. The mathematical computation of Lifting based DWT architecture is depicted in Section 3. The proposed DWT architecture is depicted in Section 4. Development of test benches is described in Section 5. The result of DWT Architecture is depicted in Section 6. Finally, a conclusion is given is Section 7.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 361–369.
springerlink.com

2 Lifting Scheme

DWT can decompose the input samples in multi resolution as in Figure 1. The implementation of the discrete wavelet transform is based on the filter banks, where H and G denote a high-pass filter and a low-pass-filter, respectively. After each filtering, the number of the output samples is decimated by a factor of 2. The samples generated by the high pass filters are completely decomposed; meanwhile, the other samples generated by the low pass filters are applied to the next-level computation for further decomposition.

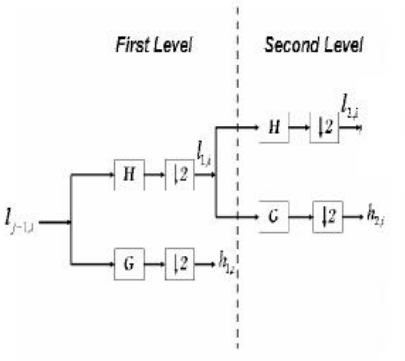


Fig. 1. 2 Level DWT

The lifting scheme is a new algorithm proposed for the implementation of the wavelet transforms. It can reduce the computational complexity of DWT involved with the convolution implementation. Furthermore, the extra memory required to store the results of the convolution can also be reduced by in place computation of the wavelet coefficient with the lifting scheme. Each decomposition stage has predict and update blocks, which extracts the relevant information and passes on to the next stage. Both of them are linear phase (symmetrical) filters. The lifting scheme architecture is shown in Figure 2.

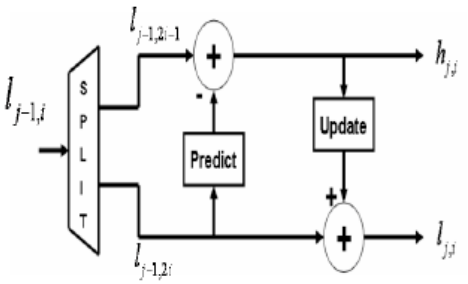


Fig. 2. Lifting Scheme

The lifting scheme consists of the following three steps to decompose the samples, namely, splitting, predicting, and updating. Figure 3 illustrates the three steps associated with the lifting scheme based DWT for the one-dimensional signal.

- **Split step:** The input samples *x* is split into even samples and odd samples.
- **Predict step:** The even samples are multiplied by the predict factor and then the results are added to the odd samples to generate the detailed coefficients.
- **Update step:** The detailed coefficients computed by the predict step are multiplied by the update factors and then the results are added to the even samples to get the coarse coefficients.

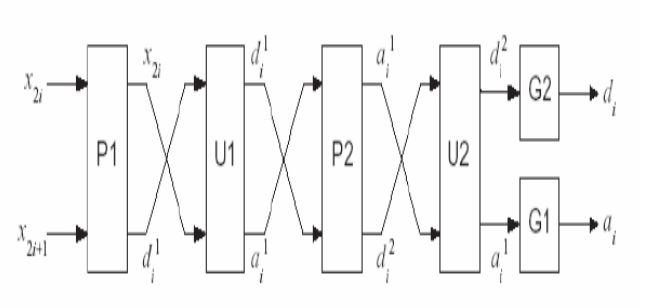


Fig. 3. Lifting Scheme Architecture

$$\text{Predict P1: } d_i^1 = \alpha\,(x_{2i} + x_{2i+2}) + x_{2i+1} \tag{1}$$

$$\text{Update U1: } a_i^1 = \beta\,(d_i^1 + d_{i-1}^1) + x_{2i} \tag{2}$$

$$\text{Predict P2: } d_i^2 = \gamma\,(a_i^1 + a_{i+1}^1) + d_i^1 \tag{3}$$

$$\text{Update U2: } a_i^2 = \delta\,(d_i^2 + d_{i-1}^2) + a_i^1 \tag{4}$$

$$\text{Scale G1: } a_i = \zeta * a_i^2 \tag{5}$$

$$\text{Scale G2: } d_i = d_i^2 * \zeta \tag{6}$$

For the given set of Predict and Update stages, assuming the value of **i = 0,** the equation can be finalized. By re-arranging all the values and the constant co-efficient, the final equation can be derived.

$$ai = (3 * \gamma.\beta.\delta.\zeta + \delta.\zeta + \beta.\zeta)\,[\,\alpha\,(x_0 + x_2) + x_1 + \alpha\,(x_0 + x_{-2}) + x_{-1}] + \zeta.\delta.\beta.\gamma\,[\alpha(x_2 + x_4) + x_3 + \alpha(x_{-2} + x_{-4}) + x_{-3}] + \zeta.\delta.\gamma\,(x_0 + x_2 + x_0 + x_{-2}) + \zeta * x_0 \tag{7}$$

$$di = 1/\zeta\,[(2 * \gamma.\beta + 1)\{\alpha\,(x_0 + x_2) + x_1\} + \gamma.\beta\,\{\alpha\,(x_0 + x_{-2}) + x_{-1} + \alpha(x_2 + x_4) + x_3\} + \alpha\,(x_0 + x_2)] \tag{8}$$

3 Mathematical Calculations

The equation which was obtained from the architecture has some of the constant terms namely Alpha, Beta, Gamma, Delta and Zeta and some of the input terms from

x-4 to x+4 which is of 8-bit data. The values of the constant terms are given in Table 1. The values of the inputs x_{-4} to x_{+4} is given is Table 2.

Table 1. Constant Co-efficient Values

Alpha	1.58613434200
Beta	0.05298011854
Gamma	0.88291107620
Delta	0.44350685220
Zeta	1.14960439800

Table 2. Values of X-4 till X+4

x-4	x-3	x-2	x-1	x	x+1	x+2	x+3	x+4
122	110	102	91	72	54	48	27	6

The architecture states that

- The input should be 8-bit and have signed representation
- The output should be 16-bit and signed representation.
- Lifting co-efficient should be 8-bit signed representation.
- All the intermediate outputs have to be stored.

The lifting co-efficient can be multiplied with a common term as 32 or 64 so that representing that number might be much easier. The lifting co-efficient has to multiplied in such a manner that the final values should be able to compute as 8-bit signed number and it does not have any number which is beyond decimal like 83.000 [Ref 3]. This can be achieved by taking only the positive values and discarding all other decimal point values. For example: XY.~~xy~~. By doing these the values of all the lifting co-efficient will be having a decimal without any point values so that the calculations can be much easier. The constant terms final value is given in Figure 4.

By substituting the final values in Fig 4 in the equations 7, 8 and discarding the LSB for the adders and multipliers, the values of ai and di will be

$ai= (3 * \gamma.\beta.\delta.\zeta + \delta.\zeta + \beta.\zeta) [\alpha (x_0 + x_2) + x_1 + \alpha (x_0 + x_{-2}) + x_{-1}] + \zeta.\delta.\beta.\gamma [\alpha(x_2 + x_4) + x_3 + \alpha(x_{-2} + x_{-4}) + x_{-3}] + \zeta.\delta.\gamma (x_0 + x_2 + x_0 + x_{-2}) + \zeta * x_0$

$ai= (57) [50 (72 + 48) + 54 + 50 (72 + 102) + 91] + 6 [50(48 + 6) + 27 + 50(102 + 122) + 110] + 30 (72 + 48 + 72 + 102) + 1 * 72$

The summation of 72 and 48 will be 60 since 72+48=120 will be a resultant of 9-bit number. Discarding the LSB makes the value of 120 as 60.

$ai = (57) [50 (60) + 54 + 50 (87) + 91] + 6 [50(27) + 27 + 50(122) + 110] + 30 (60 + 87) +72$

$ai = (57) [46 + 54 + 67 + 91] + 6 [50(27) + 27 + 50(122) + 110] + 30 (60 + 87) +72$

$ai = (57) [50 + 79] + 6 [50(27) + 27 + 50(122) + 110] + 30 (73) +72$

ai = 65

The same calculation process is taken for **di** as well.

di= 1/ζ [(2 * γ.β +1){ α (x_0 + x_2) + x_1} + γ.β { α (x_0 + x_{-2}) + x_{-1} + α(x_2 + x_4) + x_3} + α (x_0 + x_2)]

di= 6 [(35){ 50 (72 + 48) + 54} + 12 { 50 (72 + 102) + 91 + 50(48 + 6) + 27} + 50 (72 + 48)]

di = 39

Expression	Value of Expression	Multiplication Term	Product	Final Value
3 * γ.β.δ.ζ + δ.ζ + β.ζ	0.89765625	64	57.45	57
ζ.δ.β.γ	0.023359375	256	5.98	6
ζ.δ.γ	0.47046875	64	30.11	30
2 * γ.β +1	1.1031875	32	35.302	35
γ.β	0.046878906	256	12.001	12
1/ζ	0.8121875	32	25.99	26
α	1.57375	32	50.36	50

Fig. 4. Constant Terms for the eqution

4 DWT Architecture

The DWT architecture consists of a top module which is interconnected with a lot of sub blocks as shown in Figure 5. Each sub-block is having dedicated functions, operations to be performed and different Delay, Power and Area.

SIPO: SIPO shift registers are used at the inputs for DWT architecture which fetches the inputs serially with the help of enable. Whenever enable signal is high, the serial data passes through the registers so that the data can be taken for further calculations. When enable signal is low, irrespective of the serial input data then registers will be not fed with the serial data.

ADDER: Adder is one of the major building blocks for the architecture. 8-Bit signed adder is used in the circuit. The operation is performed as per the value of the two 8-bit signed integers. The output is a 9-bit signed number. As per the architecture, all the intermediate outputs should be in an 8-bit register. Since the adder's output is of 9-bit, the LSB of the adder is omitted and the remaining 8-bits are considered as the sum of the adder and stored in the register.

Multiplier: Multiplier is another major building block of the architecture. It is an 8-Bit signed constant coefficient multiplier. Again the output of multiplier is more than 8-bits so the LSB's will be truncated and the first 8-bit MSB data's will be considered as the product of the multiplier. There are 7 constant co-efficient multipliers for this architecture.

PISO: Parallel in serial out is the final stage of the architecture where the entire computation completes and the final values will be stored in 8-bit registers. These PISO

are controlled by a piso_load signal. Whenever the signal goes high, the 8-bit registers will invoke the PISO program and then the final values will displayed on to the screen one bit by another.

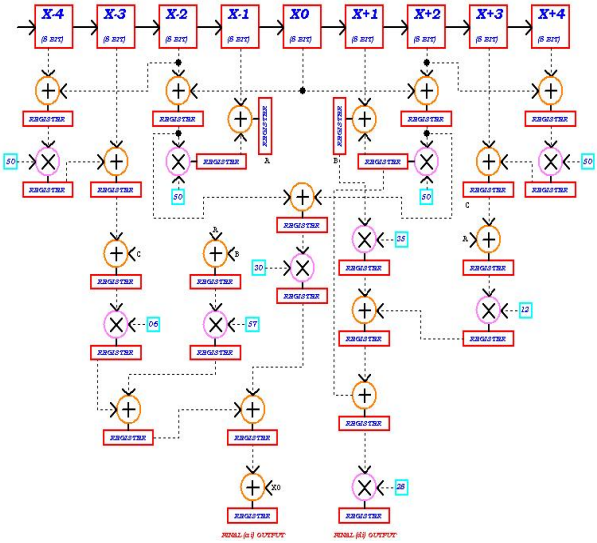


Fig. 5. DWT Architecture

Apart from all these blocks there are Parallel in Parallel out shift registers which are supposed to be added after each and every operation of ADDER and MULTIPLIER. At every positive edge of clock the operations are performed and then at the next clock cycle the data is passed on to the shift registers. The entire architecture consists of Seventeen 8-bit signed adders and Ten 8-bit signed Constant Coefficient multiplier. So that it requires Twenty-Seven 8-bit PIPO registers.

The sub-blocks which are integrated with the 2-level DWT architecture are Signed Adders, Signed Constant Co-efficient multipliers, Serial In Parallel Out shift register, Parallel in Serial out Shift registers and registers at every intermediate outputs.

In the 2-level DWT architecture of modified lifting scheme as per the equation stated in Figure 5, the number of sub-blocks is

- Nine 8-Bit Serial in Parallel Out shift Registers
- Seventeen 8-Bit signed Adders
- Ten 8-Bit signed multipliers
- Two 8-Bit Parallel in Serial out Registers

By integrating all the sub-blocks in the main module, the top module will be created. These modules are supposed to be instantiated in the top module. By instantiating these modules the control will be switching over those modules as per the link provided.

5 Developments of TestBenches

Figure 6 represents the block diagram of the test bench module used for testing the 2-level DWT architecture and the stimulus test vectors. It includes a clock generator to provide the clock for the entire system, one Parallel in Serial Out shift register, one Serial in Parallel Out shift register which is used for inputs and output data conversion and adders and multipliers circuit which is used to perform the internal operations in the entire system.

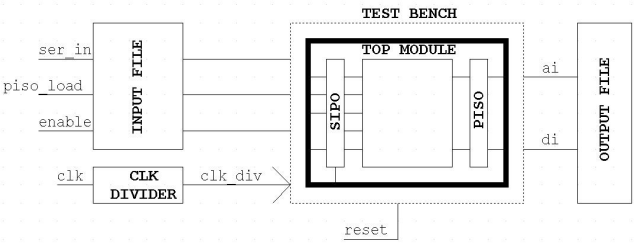


Fig. 6. Testbench with file I/O Ports

The test bench module reads the input data from the file and stores it in an internal frame buffer. Three signals are sent out from TBM to the DUT. The first signal SER_IN will be sending the serial data continuously to the DUT. The second signal PISO_LOAD will be logic high whenever the data is loaded at the output which is sent through Parallel in serial out. The third signal ENABLE controls the input signal whenever enable goes to logic low, the DUT will not process the INPUT signal. This can be used as a control signal for the input. The TBM drives the CLK_DIV by means of a signal CLK and the output of the Clock generator will be the divided value of the clock as the user requirement. The final output ai and di which is 8-bit data is passed on to a parallel in serial out and it is displayed one by one and will be written in the file as per the time intervals.

6 Results and Discussions

We can perform a behavioral simulation on your design before synthesis with the Simulate Behavioral Model process. This first pass simulation is typically performed to verify RTL (behavioral) code and to confirm that the design is functioning as intended. Behavioral simulation can be performed on a source file available in the Behavioral Simulation view which is shown in Figure 7. The post place and route simulation is also called as Post Synthesis simulation. A post synthesis simulation (Figure 8) models interconnect delay, as well as gate delay. This type of simulation will most accurately match the behavior of the actual hardware. However, for large designs, it can take a significant amount of time to extract the interconnect delay values from the place-and-route information, and a significant amount of time to run the actual simulation.

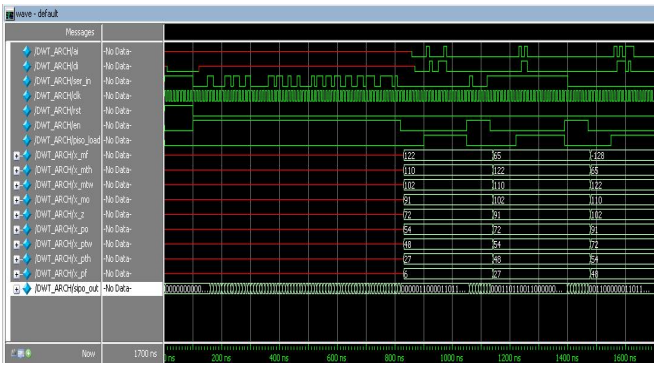


Fig. 7. Pre-Synthesis Output Waveform

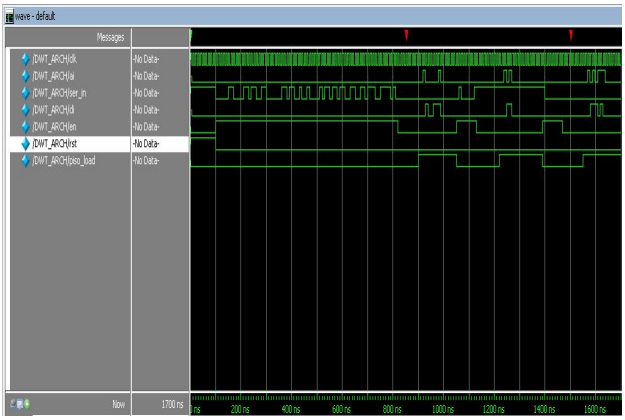


Fig. 8. Post-Synthesis Output Waveform

Chipscope is a set of tools made by Xilinx that allows you to easily probe the internal signals of your design inside an FPGA, much as you would do with a logic analyzer. In a design that uses much of the FPGA's memory, there may not be much memory left over for the Chipscope cores. Also, Chipscope cannot sample as quickly as an external logic analyzer. Calling ICON, ICA and VIO IP cores (Figure 9). By adding **ICON**, we can block unused boundary port, disable JTAG clock and disable boundary scan component with the help of control signals. By adding **ILA**, we can trigger the inputs and outputs. By adding **VIO,** we can control the input and output ports. Adding Chip Scope Definition and Configurations file (Figure 9). This will generate .cdc file which has the information of input and output ports, trigger width, sample data width which will be mapped to the main module. This file requires another top module which will be instantiated with all the control signals for **ICON**, **ILA** and **VIO** and the **DUT** for the architecture.

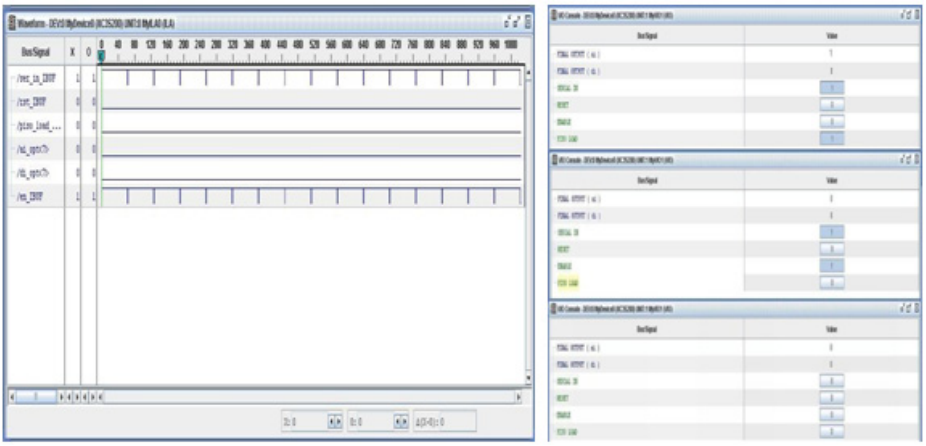


Fig. 9. Chip Scope IP Core and Configuration File

7 Conclusions

A lifting based DWT core is proposed in this paper. Folding architecture is adopted to reduce the hardware cost and to achieve the higher hardware utilization. Multiplication is realized with co-efficient represented in constant form. It is a compact and efficient DWT core for the hardware implementation of JPEG 2000 encoder. The future work will be the optimization of the memory organization of the overall JPEG2000 system.

References

1. Chen, P.Y.: VLSI implementation for one dimensional multilevel lifting-based wavelet transform. IEEE Trans. on Computers 53(4), 386–398 (2004)
2. Andra, K., Chakrabati, C., Acharya, T.: A VLSI Architecture for Lifting-based Forward and Inverse Wavelet Transform. IEEE Trans. on Signal Processing 50(4), 966–977 (2002)
3. Huang, C.-T., Tseng, P.-C., Chen, L.-G.: Efficient VLSI Architectures of Lifting-Based Discrete Wavelet Transform by Systematic Design Method. In: Proc. IEEE Int'l Symp. Circuits and System, vol. 5, pp. 565–568 (May 2002)
4. Sweldens, G.: The Lifting Scheme: A Custom- Design Construction of Biorthogonal Wavelet. Applied and Computational Harmonic Analysis 3, 186–200 (1996)
5. Alam, M., Rahmana, C.A., Badawy, W., Jullien, G.: Efficient distributed arithmetic based DWT architecture for multimedia application. In: Proceedings of The 3rd IEEE International Workshop on System on Chip for Real-Time Application (2003)
6. Verma, A.K., Ienne, P.: Improved use of the carry-save representation for the synthesis of complex arithmetic circuits. In: Int. Conf. Computer-Aided Design (ICCAD 2004), San Jose, CA, USA, November 7-11, pp. 791–798 (2004)

Mobile Based E-Mail Reading System

Azath M.[1] and Channamallikarjuna Mattihalli[2], Member IEEE

[1] Department of Computer Science
[2] Department of Electrical and Computer Engineering
Debre Berhan University, Ethiopia
azathhusssain@gmail.com, ckmattihalli@ieee.org

Abstract. This paper is to read out all the messages in one's e-mail account through mobile. When ones the user calls up, depending on the password entered by him/her, the user is auto authenticated into the respective e-mail account. Later, we are going to download all the read messages in the users account and convert the same from text to speech (TTS) and read out the same through mobile. Mobile based e-mail reading system comes into picture when blind people need to have their mails read to them. The aim here is to read any message present in the in-box. When the user authenticates himself through his User-ID and Pin, he gets a list of messages present in his in-box. The user has the option of listening to un-read, read messages with or without content. When the required message is selected, it is converted from text to voice and is read out.

Keywords: DTMF, API, TTL, MAX 232, RS 232.

1 Introduction

The scope of this paper under given time constraint encompasses design and development of mobile interfacing circuits which connects mobile signal to computer to read emails. Designing aspects include detailed study of approved to DTMF decoder microcontroller, MAX'232, serial port Development of software incorporates the usage of dynamic link library concepts, speech technology and mailing concepts. This paper designing can be further made used to avail other services like weather forecasts, news update, stock market updates etc. This paper that aims in reading emails through mobile. The very first task of this project is to design interface card that is to be connected to serial port of the system to handle ample number of inputs and outputs. The serial port is to be programmed to configure the input and output ports. The second task is designing mobile interfacing circuits which should work out to be very economical, efficient and accommodates approved circuits of mobile so that it can be implemented easily avoiding dangerous voltages to pass the circuits. Most of the R&D people investing more on speech technology but it is not so much convenient to user, because of unique accent and situation of the user.

1.1 Literature Survey

There are several related works; some of them are as follows: Most of these applications are oriented to blind users and provide access to common programs and operating

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 371–378.
springerlink.com

systems. This is just a partial list of applications using speech synthesis. Emacspeak Driver for the Doubletalk and LiteTalk Synthesizers

- Emacspeak Driver for Braille 'n Speak, Braille Lite, and Type 'n Speak
- Linux Device Driver for the DoubleTalk PC
- Screader: a Text-to-Speech Application for Linux
- Slackware96 Rootdisk for the Blind

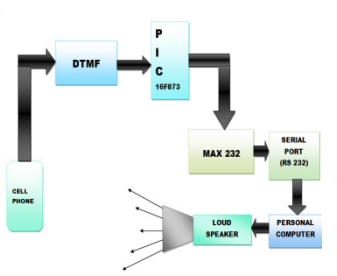
There is also a company Net phonic communications involved in working in the same related area that provides two products with similar features. The present development in this project includes design, development and implementation mobile based email reading system which provides core functionality provides an open architecture in which it is easy to enrich the system with new features. Next was to go through books in depth on SPC Digital Exchanges by Redmill and Valder and mobile Switching systems and networks by Thiagrajan Vishwanathan which gave a broad, yet fairly in depth and up to date coverage of telecommunication switching systems and networks.

Went through related several website to get clear understanding of hardware components and about its interfacing details for the project and requirements in designing the hardware.

Next step was to handle mailing concepts; went through javaMail API studied in detail for handling mailing concepts. Final step was to handle speech technology went through website www.cloudgarden.com and downloaded trial version of cloud garden and studied the details about cloud garden in website in order to get familiar with it. Went through java Speech API, studied in detail for use its usage in this project. After collecting this information's, conducted an exhaustive analysis on how to design, develop to implement final product.

2 Proposed System

The end user who wants his mails to be read out dials his Secret identification number on a mobile handset. Mobile operates on frequency whereas micro-controller is digital in nature. Hence a DTMF decoder is required for the conversion of frequencies into digits. DTMF Decoder is followed by Microcontroller. Here we use PIC 16F873 Microcontroller. It works on TTL logic whose voltage range is from 0 to +5V.



Here it is programmed such that it provides a connection after two rings from the mobile.RS232 is used for the serial communication between the microcontroller and the PC. Therefore to connect RS232 to a microcontroller, a voltage converter like MAX232 is used to convert TTL logic levels to RS232 voltage levels and vice versa.

MAX 232 is used for converting RS232 signals to TTL voltage levels which is acceptable to the PIC. PIC has build-in serial communication capability; hence the RS232 is connected to the PC through serial port. The text to speech (TTS) conversion is implemented on the PC (uses Java programming).The speech signal which is the output of the input text is dragged out of PC using speakers. This speech signal is nothing but the speech version of the e-mails which is heard by the user on the mobile or telephone line.

- DTMF:

DTMF consist of two sine wave of given frequencies. Using frequency filters individual frequencies are selecting, so that they can be passed very easily through lines. DTMF was designed only control signals.DTMF is an important for the voice communication control. DTMF also used to dial numbers; sometimes floating codes are transmitted using DTMF usually through a CB transceiver.

- PIC 16F873:

A microcontroller is a highly integrated chip which performs controlling functions. It is called as one-chip microcomputer .A µ-controller is distinguished from a µ-processor in that it has many capabilities useful for real-world interfacing built into the chip. The PIC is a microcontroller from Microchip Technology Inc. The PIC has RISC Architecture with individual memory for program and data.

The one we will be using is the PIC16F873. PIC is having on-board many peripheral devices such as RAM, EPROM, timers, input output ports, oscillator, analog to digital converter etc. The PIC16F873 has 4096 words of memory for program, 192 bytes of RAM, and 128 bytes of EEPROM and can operate with clocks up to 20 MHz on 8 bits of data.

- Decoding DTMF:

It is difficult to detect and recognize DTMF. Many ASIC are used, many microprocessors are used to transmit and receive DTMF most of the time MT8870 is used for compatibility. So, DTMF generated rectangular pulses and RC filters works reliably. The mentioned MT 8870 uses two 6th order band pass filters with switched capacitor

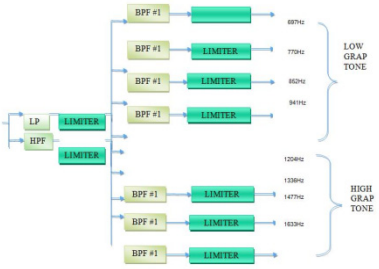


Fig. 1. Dual Tone Multi Frequency Decoder

These produce pure sine waves even from distorted inputs, with any harmonics suppressed. The scheme used to identify the two frequencies associated with the buttons that has been pressed is shown above. Here, the two tones are first separated by a LPF and a HPF. The P/B cut off frequencies of LPF is slightly above 1000Hz, where as the HPF is slightly below 1200Hz.

The output of each filter is next converted into a square wave by a limiter & then processed by a bank of BPF with narrow P/B's. The four BPFs in the low frequency channel have center frequencies at 697Hz, 770Hz, 852Hz & 941Hz. The 4-BPFs in higher frequency channel have center frequencies at 1209Hz, 1336Hz, 1477Hz & 1633Hz. The detector following each BPF develops the necessary de-switching signal if its input voltage is above a certain threshold.

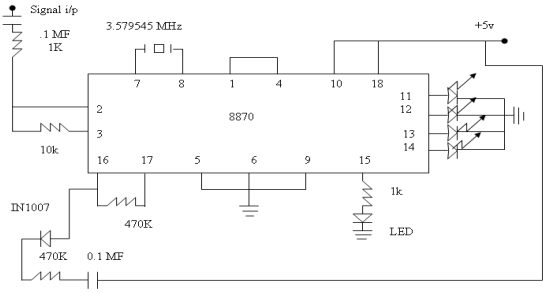


Fig. 2. IC MT8870 is used to detect and recognize DTMF

The digital implementation of a DTMF signal involves adding two finite length digital sinusoidal sequences with latter simply generated by using Look-up tables or by computing a polynomial expansion. The digital tone detection can be easily performed by computing DFT of DTMF signal & then measuring the energy product in 8-DTMF frequencies. Minimum duration of a DTMF signal is 40ms. Thus with a sampling rate of 8 KHz, these are utmost 0.04 x 8000 = 320 samples available for decoding each DTMF digit. The actual number of samples used for the DFT computation is less than this number is chosen so as to minimize the difference between the actual location of the sinusoid and the nearest integer value DFT index K.

The DTMF decoder computes the DFT samples closest in frequency to the 8-DTMF fundamental tones & their respective 2^{nd} harmonics. In addition, a practical DTMF decoder also computes the DFT samples closest or the frequency to the second harmonics corresponding to each of the fundamental tone frequencies. The DTMF signal generates by the handset has negligible second harmonics. The DFT computation scheme employed is a slightly modified version of Goertzel algorithm.

The DFT length N determines the frequency spacing between the locations of the DFT samples and the time it takes to compute the DFT sample. A large N makes the spacing smaller, providing higher resolution in frequency domain but increases the computation time.

The frequency fk in Hertz corresponding to DFT index (Binary number) K is given by

$$fk = k * FT / N \quad (3.1)$$

FT= sampling frequency. To minimize leakage, it is desirable to choose N appropriately.

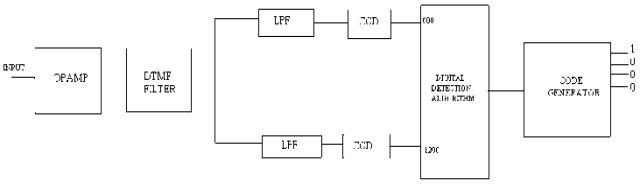


Fig . 3. Functional Block Diagram of DTMF decoder

Referring given circuit, it displays the number dialled from the mobile set using the DTMF mode. This circuit can also show the number dialled from the mobile of the called party. This is particularly helpful for receiving any number over the mobile. The DTMF signal—generated by the mobile on dialling a number—is decoded by DTMF decoder MT8870 which converts the received DTMF signal into its equivalent BCD number that corresponds to the dialled number. Pin number 11 to 15 are connected to Port B0-B5 as shown in schematic circuit diagram.

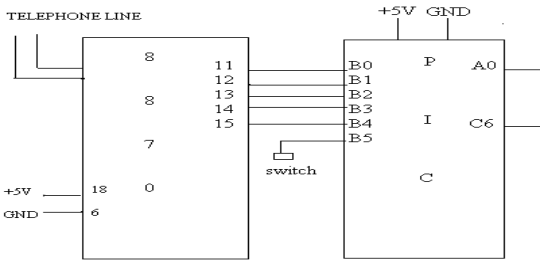


Fig. 4. PIN Connections of DTMF Decoder and PIC16F873

3 Result

The system developed had a lot of short comings initially, but gradually it was cleared. Things like not able to differentiate between read and unread messages were a major concern, as any individual is not allowed to meddle with the server database nor the client server.

When the user accesses his in-box with our setup, and listens to his mails, we download the date of particular messages and store it in a file, under subsequent accessing of that particular in-box, the dates of the downloaded messages and those present in the file (on the server) are compared, if there is mismatch then new messages have arrived, giving the constraint that as the number of users increase the server disk space exponentially increases.

The use of 2 different sets of digits as authentication-

- USERID
- PIN

Initially it was sufficient that every user had one pin for authentication, but then came the question of secrecy; and so we adopted the idea of having 2 different sets of numbers for verification. The proposed mobile based email reading system is developed to access your mailbox remotely using mobile. This project can be made use of in remote places where there is no internet connection.

Initially, the user of this setup must register him in our website. Once registered, the user is given a 4 digit secret code. This secret code is used by the user to access his/her in-box remotely by calling up our service.

The features that are available are:

- Users can remotely login to our homepage and get registered. On registration, the user is given a 4 digit unique password.
- User can call up this service and enter the secret code in order to access his in-box remotely.
- On accessing his in-box, the user is given the following three options
  - Press 1 for reading all read messages
  - Press 2 for reading all unread messages
  - Press 3 for quitting from the system
- On selecting the option, the subject and the from addresses of all the mails are read out and asks the user to press 2 to read the contents.
- The contents of that mail are downloaded and read out to the remote user over the phone.
- User can press 3 to quit the system and finally '#' to disconnect from the system.

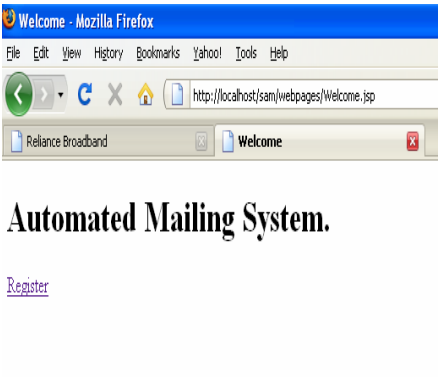


Fig. 5. Welcome Page

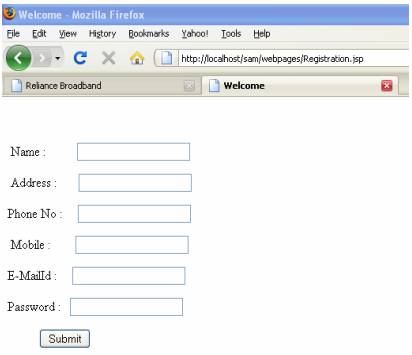


Fig. 6. Registration Page

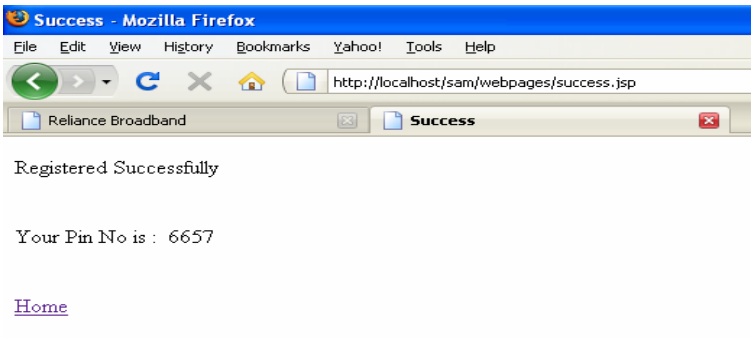


Fig. 7. Successful Registration

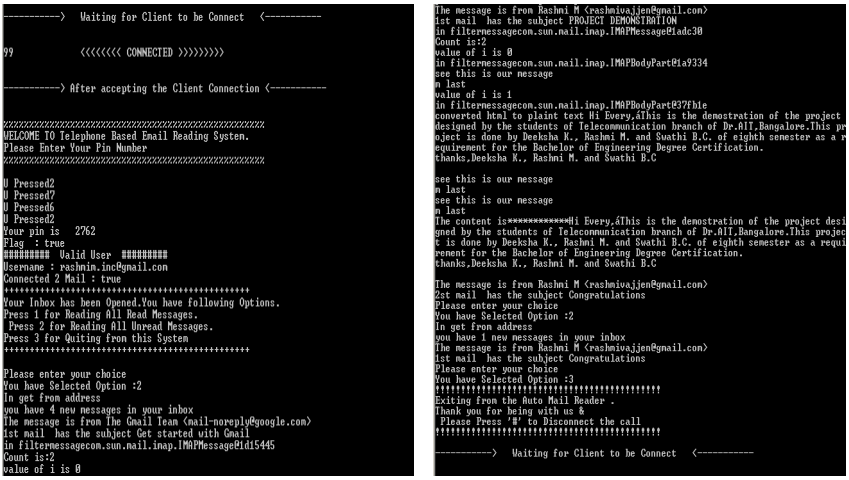


Fig. 8. Execution process

4 Conclusion

The new configuration we ventured on and proposed to make the task of accessing Electronic-Mails as easily as listening to it over a more common and easily available medium such as a mobile is successfully attained.

The preliminary proposal of making the task easier for the user to have his in-box at his disposal was satisfied by making use of the telephone at the client side. Further the options of dispensing the mails and modifying the user's in-box was achieved with the help of Flags and the files that are used for comparison. Any user on the move can avail our service if he has a mobile device, without the need of having a modem at his disposal. On top of everything, most of all, every system must be foolproof in accessing only user's particular in-box, and preventing any unauthorized access to others' in- box. This has been achieved by querying the user to enter a pin that is unique to every individual account.

5 Future Enhancements

There is a huge scope for future inclusions in our project "Automated E-mail Reading System". Presently, only one user can be hooked into this system. Instead the project can be upgraded so that many users can access the service simultaneously. This can be achieved by programming the microcontroller to receive many calls from different users.

It can be replaced by antenna and the whole system can be wirelessly controlled. If the e-mails have images, provisions for downloading them and displaying them on the mobile screen can be added.

Here the text is converted to English language which is globally accepted. This can be extended to other languages also. The clarity of the speech can be enhanced by using advanced speech SDK software's.

References

[1] Mark Palmer Microchip Technology Inc. AN858 Host UART.-pin PIC16F873
[2] Rob Hamerling Device include file pic16f873, containing Declaration of ports
[3] Bagchi, S., Mitra, S.K.: An Efficient Algorithm for DTMF Decoding Using the Subband NDFT. In: Proceedings of the IEEE Symposium on Circuits and Systems, Seattle, WA, pp. 1936–1939 (1995)
[4] `http://www.cloudgarden.com` for text to speech conversion

Design of ANFIS Controller Based on Fusion Function for Linear Inverted Pendulum

Abhishek Kumar and R. Mitra

Indian Institute of Technology, Roorkee, Haridwar
abhishekdri@gmail.com, rmtrafec@iitr.ernet.in

Abstract. "Rule number explosion" and "adaptive weights tuning" are two main issues in the design of fuzzy control systems. To overcome the problems, a method is implemented for control of the inverted pendulum (IP) using linear fusion function based on LQR mapping, and combines it with adaptive control scheme to tune controller parameters using ANFIS. By using fusion, the output variables of the system with four dimensions are synthesized as two variables: error and variation of error. The method is applied to the approximate linear model, and the experimental results show that this method has better tracking performance, disturbance resisting performance, and robustness against model parameter perturbance.

Keywords: LQR control, fusion function, ANFIS, fuzzy control, inverted pendulum.

1 Introduction

From mid 1990's, fuzzy neural network has been applied to the control of under- actuating systems. However, two key problems in fuzzy neural network control still remains, namely, the optimal tuning of controller parameters and the effective suppression of rule number explosion. For the first issue, ANFIS [1] provides an efficient solution and has been utilized to control the inverted pendulum system. For the second issue, due to its importance to control scheme implementation efficiency, it has attracted many researchers attention.

In this paper, we designed a new ANFIS based control scheme where four state variables are fused into two variables: error and variation of error to solve the problem of rule explosion. The coefficients used in linear fusion function are derived from the LQR feedback controller parameters of inverted pendulum. Experiment results of simulation show that this method has better control quality than LQR.

2 Modeling and Characteristics Analyzing of Inverted Pendulum

2.1 Inverted Pendulum Structure

Inverted pendulum, an underactuated (number of joints greater than number of actuators) mechanical system, has two degree of freedom (DOF) with ϕ the unactuated variable. The inverted pendulum system is composed of a cart moving on guideway[4] and

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 379–386.
springerlink.com

a pendulum which is fixed on the cart. The displacement of the cart can be measured by a sensor installed on one side of the guideway, and the angle signal can be measured by a coaxial angle sensor install in the bearing which articulates the pendulum to the cart.

On the other side of the guideway is mounted a DC permanent magnetic direct torque motor, driving the cart to move on the guideway. When the cart moves from left to right and vice-versa, torque acts on the pendulum to keep the whole system in stable mode. The cart that is shown as Fig. 1(a) is controlled by the function $F=u(t)$ moving in the x axis direction to keep the pendulum stable in the perpendicular plane. The cart mass is M, the pendulum mass is m, its length is $2L$.The cart is restricted to move within a fixed range. The reference position for x is zero *meter*, when cart is in the center of the chosen basic universe of discourse; and for $\phi(\theta)$ is π(zero) *rad*, when the pendulum is at a natural stable downward position. The motor input voltage range is $-5V$ to $+5V$.

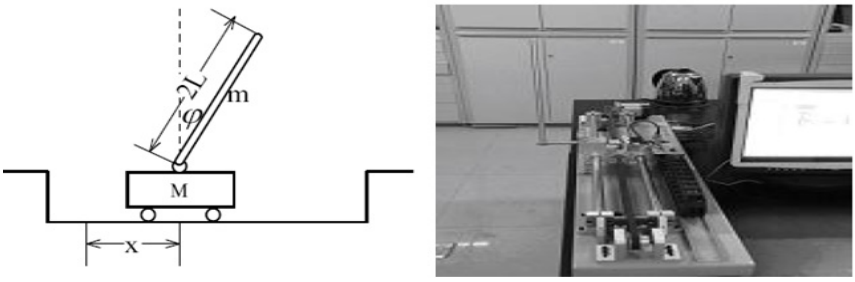


Fig. 1. (a)Inverted pendulum system sketch (b) Real time control effect picture of IP[2]

2.2 Mathematical Model of Inverted Pendulum

In this paper, the IP system, by Googol's GLIP2001 [3] model; can be viewed as a rigid-body system of cart-pendulum when neglecting air resistance and various frictions. Build the one stage linear IP mathematical model near its vertical upright balanced state, the dynamic equations of the system can be found with help of the Euler-Lagrange equation as:

$$\begin{aligned}(M+m\ddot{x}+b\dot{x}+mL\ddot{\theta}\cos\theta-mL\dot{\theta}^2\sin\theta)&=F\\(I+mL^2)\ddot{\theta}+mgL\sin\theta&=-mL\ddot{x}\cos\theta\end{aligned}\tag{1}$$

Table 1 shows the parameters of the inverted pendulum used in the model. The linearized form of nonlinear system is derived, taking the state variables:

$$x_1=x, x_2=\dot{x}_1, x_3=\phi, x_4=\dot{\phi}$$

Gate the state space equations (with cart acceleration as input) as follows:

$$\begin{aligned}\dot{X}&=AX+Bu\\Y&=CX+Du\end{aligned}\tag{2}$$

$$A=\begin{bmatrix}0&1&0&0\\0&0&0&0\\0&0&0&1\\0&0&29.4&0\end{bmatrix},B=\begin{bmatrix}0\\1\\0\\3\end{bmatrix},C=\begin{bmatrix}1&0&0&0\\0&0&1&0\end{bmatrix},D=\begin{bmatrix}0\\0\end{bmatrix}$$

It is easy to obtain the state controllability and output controllability expression of IP system and it is as (3) and (4) respectively.

$$rank\begin{bmatrix}B & AB & A^2B & A^3B\end{bmatrix}=4 \quad (3);\quad rank\begin{bmatrix}CB & CAB & CA^2B & CA^3B & D\end{bmatrix}=2 \quad (4)$$

Table 1. Parameters of IP equations

Parameter	Definition	Value	Unit	Parameter	Definition	Value	Unit
g	Gravity constant	9.81	*N/Kg*	$\dot{x}$	Cart velocity	..	*m/s*
M	Mass of cart	1.096	*Kg*	$\ddot{x}$	Cart acceleration	..	m/s^2
m	Mass of rod	0.125	*Kg*	ϕ *or* $(\pi-\theta)$	Deflection of pendulum	..	*Rad*
b	friction coefficient of the cart	0.1	*N/m/s*	$\dot{\phi}$	Velocity of pendulum	..	*rad/s*
L	distance from rod rotation axis center to rod mass center	0.25	*m*	$\ddot{\phi}$	Acceleration of pendulum	..	rad/s^2
I	Inertia of rod	0.0034	$Kg\ m^2$	*u*	Cart acceleration as input	..	m/s^2
x	Cart displacement	..	*m*				

3 Implementation of Information Fusion

Inverted pendulum system is a multi-sensor system; multi-sensor is the basis of information fusion, and multi-source information is its object. The information fusion is that the multiple sensor or multi-source information is treated comprehensively, in order to obtain more accurate and reliable conclusion [6]. The approximate linear state equation near the natural upward equilibrium position of the inverted pendulum system can be obtained after some assumptions and approximations. It is well known that the linear system has characteristics of direct integration, so the inverted system state variables can be changed into integrated error *E* and error change *EC* by constructing a linear fusion[7] function.

In this paper, construct a linear fusion function on the basis of LQR gain mapping.

Define the inverted pendulum system's state space as in (2).

Choose the quadratic objective function as:

$$J=\frac{1}{2}\int_0^\infty (X^TQX+u^TRu)dt \quad (5)$$

For the inverted pendulum system, weighting matrix *Q* and *R* are used to balance the weight of the system's state vector *X* and *u*. Because of *Q* being a semi-definite matrix

and *R* being a definite matrix, the objective function is non-negative. On output, disturbances affecting the system, give an appropriate *u* that is called optimal control to make the system return to equilibrium position as soon as possible and at the same time make the objective function minimum.

Fusion function design steps [8] combining with optimal control are given as follow:

- Calculate the state feedback matrix *K* that can make the inverted pendulum system basically stable through LQR theory. For, *R*=1 and *Q*=[1000 0 200 0] ; *K*= [-31.623 -20.151 72.718 13.155].
- Construct fusion function $F_1(X)$ using state- feedback matrix *K* as (6):

$$F_1(X) = \frac{1}{\|K\|}\begin{bmatrix} K_x & K_\phi & 0 & 0 \\ 0 & 0 & K_{\dot{x}} & K_{\dot{\phi}} \end{bmatrix} \quad (6); \quad \|K\| = \sqrt{[(K_x)^2 + (K_\phi)^2 + (K_{\dot{x}})^2 + (K_{\dot{\phi}})^2]} \quad (7)$$

- Reduce the dimensions of input variable $X = [x, \phi, \dot{x}, \dot{\phi}]$

by $F_1(X)$, and obtain the comprehensive error *E,* error- change rate *EC* expressed as (8):

$$\begin{bmatrix} E \\ EC \end{bmatrix} = F_1(X)X^T \quad (8)$$

4 Controller Design

4.1 Anfis Based Control

ANFIS, which is proposed by Jang [1], is a connectional simulation of the fuzzy system concept and T-S inference model. For the considered inverted pendulum system of two inputs and one output, first order T-S model are adopted to express the fuzzy rules, that is

Rule1: IF x is A_1 *and y is* B_1 *THEN* $f_1 = p_1x_1 + q_1x_2 + r_1$

Rule2: IF x is A_2 *and y is* B_2 *THEN* $f_2 = p_2x_1 + q_2x_2 + r_2$

Where, A_i and B_i are fuzzy sets corresponding to input variables, and their membership functions are $\mu_{Ai}(x)$ and $\mu_{Bi}(y)$. The output of the ANFIS system is computed by:

$$f = \sum_{i=1}^{2} \overline{W}_i f_i = \sum_{i=1}^{2} \overline{W}_i (p_i x + q_i y + r_i) \quad \text{Where,} \overline{W}_i = \frac{\mu_{A_i}(x)\mu_{B_i}(y)}{\sum_{j=1}^{M} \mu_{A_i}(x)\mu_{B_j}(y)} \quad (9)$$

is normalized weight meaning how much it matches the corresponding rule. p_i, q_i and r_i are consequent parameters which can be updated using linear least square estimation (LSE) algorithm, and the parameters of membership functions $\mu_{Ai}(x)$ and $\mu_{Bi}(y)$ have to be updated using nonlinear gradient descent algorithm.

Choosing step-size of *0.01* with '***ode3***' solver and simulation time *10s*.The *3003* data sets are obtained with the initial conditions listed in table 2 for ANFIS training and checking.

Table 2. ANFIS Training Data Generation

Sl.no.	Initial conditions				Data-matrix
	x	θ	$\dot{x}$	$\dot{\theta}$	
(i)	0.01	0	0	0	1001×3
(ii)	0.1	π/4	0	0	1001×3
(iii)	0.1	-π/4	0	0	1001×3

To make the universe of discourse for *E, EC* and *u* in the range [-1 1], a scale factor (K_u) of 200 is chosen here for output dataset. Input variable *e* and *ec* are divided into three fuzzy subsets with Gaussian membership functions. At last, the ANFIS controller is designed with the following steps in MATLAB ANFIS GUI [9] editor:

1) Load data: Load training and test data to ANFIS editor;

2) Generate fuzzy inference system: Load initial FIS system selecting grid partition with Gaussian MF and rules connect inputs-outputs linearly.

3)Train fuzzy-neural inference system: Training parameters are hybrid learning algorithm, zero minimum allowable error, and 30 training epochs. The previously mentioned datasets are used to train the initial fuzzy inference system of step (2);*trnRMSE is 0.000006.*

4) Test ANFIS: After ANFIS training is finished, the loaded test data are used to test the trained system.

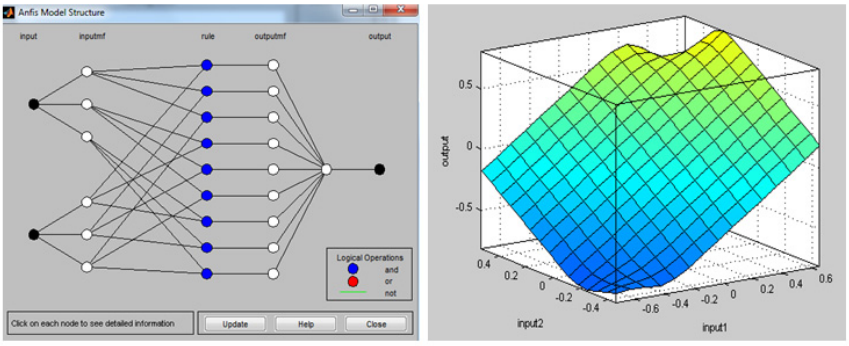


Fig. 3. (a) Topological structure (b) Control rule surface of designed ANFIS controller

5 Simulation Results and Analysis

The design is tested in the MATLAB with SIMULINK [9] environment. During debugging process, the given initial states of inverted pendulum including the cart

displacement and pendulum angle will especially affect the system's stability. Assume pendulum angle deviating from the vertical upright direction as ϕ. We have done several experiments to show the performance of our control system.

The first experiment is to check the performance of LQR controller with different sets of *Q* and *R* matrix when control input is a step input. The best of them is chosen to compare the results with fuzzy and ANFIS control. Fig. 4 shows the step response of the system with LQR without any disturbances. LQR control is able to stabilize the pendulum within 3*s* with zero steady-state error; hence the design criterion is satisfied.

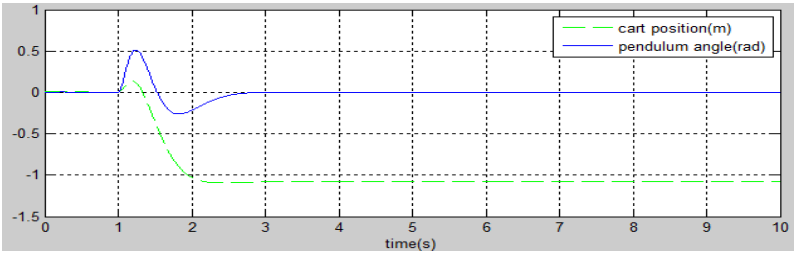


Fig. 4. Stabilization of IP using LQR control

In the 2nd experiment, linear fusion using LQR mapping without fuzzy controller is tested which results exactly same performance as LQR. This reveals that, the role of fusion function is just to reduce the dimension of inputs to the controller, whereas the controller part is nothing but LQR gain in modified form.

In the 3rd experiment, a pulse of amplitude *4V* for time-interval 4-6*s* used as disturbance which is superposed with control input and the result is shown in Fig.5.

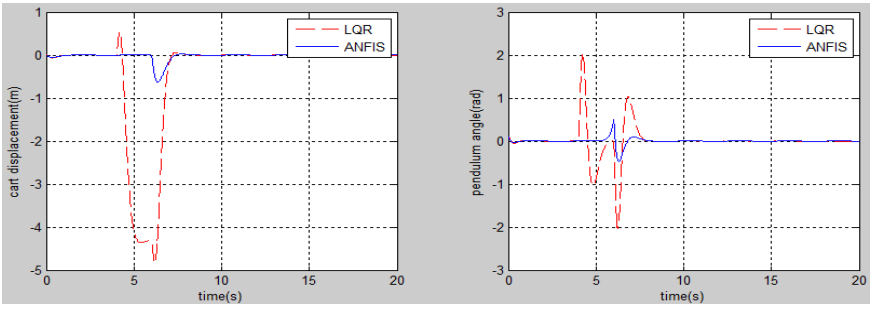


Fig. 5. Simulation results comparison of ANFIS and LQR control for (a)cart displacement, (b) pendulum angle, when a small disturbance is superposed with control input

In the last experiment, we can see that in the LQR control method, the cart displacement crosses its real universe of discourse which will cause instability in the system.

6 Conclusion

As a result the design process of fuzzy controller has been much simplified and the control quality has also been improved significantly. Experimental evidence suggests that the LQR mapped ANFIS controller employed exhibits greater robustness to complex dynamics and uncertain parameters compared with the standard LQR controller. Such kinds of fuzzy controllers have clear design ideas, satisfactory reliability and practicability. It provides a better foundation for the real time control to the other complex underactuated system.

References

[1] Jang, J.-S.R., Sun, C.-T.: Neuro Fuzzy Modelling and Control. IEEE Proc. 83, 378–406 (1995)
[2] Han, Y., Liu, Y.: One Rod Inverted Pendulum Controller Design Based on Self-Adaptive Fuzzy PID with Fuzzy. In: Proceedings of the IEEE, 8th World Congress on Intelligent Control and Automation, China, pp. 4891–4894 (July 2010)
[3] Googol Technology, GLIP series User's Manual (2006)
[4] Liu, H., Duan, F., Gao, Y.: Study on Fuzzy Control of an Inverted Pendulum System in the Simulink Environment. In: Proceedings of 2007 IEEE on Mechatronics and Automation, pp. 937–942 (August 2007)
[5] Wang, L., Zheng, S., Wang, X., Fan, L.: Fuzzy Control of Double Inverted Pendulum Based on Information Fusion. In: IEEE International Conference on Intelligent Control and Information, pp. 327–331 (August 2010)
[6] Han, Y., Liu, Y.: Reduced –Dimension Fuzzy Controller Design Based on Fusion Function and Application in Double Inverted Pendulum. In: IEEE International Conference on Industrial Mechatronics and Automation, vol. 2, pp. 337–340 (2010)
[7] Ladeneva, Y.N.: Automatic estimation of parameters to reduce rule base of fuzzy control complex systems, a master thesis, Puebla, Mexico (August 2006)
[8] Qu, Z., Xie, W., Zhou, Q.: Variable Composition Based Adaptive Fuzzy Control of Double Inverted Pendulum. In: IEEE Conference on Fuzzy System and Knowledge Discovery, vol. 2, pp. 768–772 (2010)
[9] The Mathworks, Using Matlab version 7.10.0, The Mathworks, R2010a

Appendix I

Table 3. The Parameters of Anfis Control Rules

	Premise parameter				**Conclusion parameter**		
	C1	***σ1***	***C2***	***σ2***	p_i	q_i	r_i
R1	-0.7695	0.1924	-0.4827	0.155	0.324	0.003685	-0.4184
R2	-0.7695	0.1924	0.04043	0.2513	0.921	0.921	0.0001988
R3	-0.7695	0.1924	0.5698	0.1251	0.9215	0.9209	0.000382
R4	-0.03591	0.3069	-0.4827	0.155	0.9208	0.9216	0.000311
R5	-0.03591	0.3069	0.04043	0.2513	0.9206	0.9206	0.0000006
R6	-0.03591	0.3069	0.5698	0.1251	0.9216	0.9217	-0.0002321
R7	0.6896	0.2152	-0.4827	0.155	0.9202	0.9209	0.0003893
R8	0.6896	0.2152	0.04043	0.2513	0.9214	0.921	-0.0004457
R9	0.6896	0.2152	0.5698	0.1251	0.001398	0.05935	0.1879

Proposing Modified NSGA-II to Solve a Job Sequencing Problem

Susmita Bandyopadhyay[1] and Arnab Das[2]

[1] Department of Production Engineering, Jadavpur University, Kolkata, India
bandyopadhyaysusmita2009@gmail.com
[2] NIMTT, Kolkata, India
joy1981das@gmail.com

Abstract. In this paper, a bi-objective job sequencing problem is proposed. The objectives are – 1) total weighted tardiness of jobs and 2) total deterioration cost. The proposed problem has been solved by a modified version of the original Nondominated Sorting Genetic Algorithm-II (NSGA-II) which is one of the commonly applied Multi-Objective Evolutionary Algorithm in the existing literature. NSGA-II has been modified by introducing a novel mutation algorithm that has been embedded in the original NSGA-II. The experimental results show the Pareto optimal solutions and conclusions are drawn based on the results.

Keywords: Mutation, Multi-Objective Evolutionary Algorithm (MOEA), Non-dominated Sorting Genetic Algorithm (NSGA-II), Pareto optimal solution, Job sequencing.

1 Introduction

Job sequencing is a problem of deciding over the correct sequencing of jobs in a manufacturing system. In Multi-Objective Problems, a vector of decision variables optimizes a vector of objective functions. Because of the presence of a number of objectives, we get a set of optimal solutions, called Pareto Optimal solutions, instead of a single optimal solution.

In this paper, we have also modified one the most widely used MOEA techniques known as Non-domination Sorting Genetic Algorithm – II (NSGA-II) (Deb et al. 2002). The existing literature shows a variety of improvements to NSGA-II, such as controlled elitism, scalarizing fitness function, and so on. In this paper we have modified NSGA-II, by introducing a mutation algorithm embedded in NSGA-II.

2 Literature Review

A significant level of research studies on multi-objective scheduling and sequencing are observed in the existing literature. Tuong and Soukhal (2010) investigated a bi-objective job sequencing problem with minimization of total weighted tardiness and due date cost and solved the problem by decomposing the problem in to an assignment problem. Tavakkoli-Moghaddam et al. (2010) developed a fuzzy multi-objective

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 387–392.
springerlink.com

linear programming to solve a job sequencing problem with two objectives – total weighted tardiness and makespan. Some other remarkable research studies are the studies by Karimi et al. 2010, Venditti et al. 201). Various techniques adapted to solve the multi-objective problems in the literature are Particle Swarm Optimization (Sha and Lin 2010), Ant Colony optimization (Yagmahan and Yenisey 2010), Differential Evolution (Wanga et al. 2010), Tabu Search (Erenay et al. 2010), NSGA-II (Zandieh and Karimi 2010), Agent Based Modelling (Sabouni and Yazdani 2010).

3 Problem Formulation

At first, we provide the assumptions and notations in subsections 3.1.

3.1 Assumptions and Notations

The assumptions and notations are provided in Figure 1 and Figure 2 respectively.

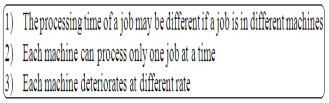


Fig. 1. Assumptions

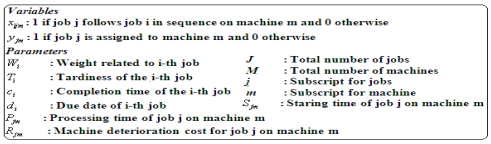


Fig. 2. Notations

$$Min\ Z_1 = \sum_{i=1}^{J} W_i T_i \qquad (1)$$

$$Min\ Z_3 = \sum_{m=1}^{M} \sum_{j=1}^{J} R_{jm} y_{jm} \qquad (2)$$

Subject to the constraints:

$$\sum_{\substack{i=1, i\neq j \\ j\neq 1}}^{J} x_{ijm} = 1 \qquad (3)$$

$$\sum_{\substack{j=1 \\ j\neq j}}^{J} x_{ijm} \leq y_{im} \qquad (4)$$

$$c_j \geq S_{jm} + P_{jm} \qquad (6)$$

$$\sum_{m=1}^{M} y_{im} = 1 \qquad (5)$$

$$T_i \geq c_i - d_i \qquad (7)$$

$$x_{ijm} + x_{jim} = 1 \qquad (8)$$

$$c_i, S_{jm}, T_i \geq 0 \qquad (9)$$

Fig. 3. Formulated Problem

3.2 Problem Formulation

The formulated problem is given below in Figure 3.

Objectives (1) and (2) minimize the total weighted tardiness of all jobs and the total deterioration cost of all the jobs respectively. Constraint (3) ensures that only one job precedes each job. Constraint (4) states that if job j follows job i then both job I and job j belong to machine m, assuming that only one job follows a job and only one job precedes a job. Constraint (5) states that each job is assigned to exactly one machine. Constraint (6) states that the completion time of job j is greater than or equal to the sum of the starting time of job j on machine m and the processing time of job j on machine m. Constraint (7) defines the tardiness. This constraint states that the

tardiness of a job j must be greater than or equal to the completion of job j minus the due date of job j. Constraint (8) ensures that either job i will follow job j or job j will follow job i. Constraint (9) ensures that c_i , S_{jm} and T_i must be positive quantities.

4 Modified NSGA-II Algorithm

For experimentation, we have taken 6 jobs and 4 machines. Thus we have taken 12 variables – I) the first 6 variables are for sequencing 6 jobs, II) the last 6 variables are for assigning machines to 6 jobs. The modified NSGA-II algorithm as applied in this paper is given in Figure 4.

```
MODIFIED NSGA-II

Initialize variables
Generate random population of size N
Evaluate values of each objective
Perform Nondomination Sort on the population,
    to assign rank and crowding distance value to each chromosome
For each generation
  If probability <= Crossover_probability
        Perform Tournament Selection to generate mating pool
        For each chromosome in mating pool
           Perform order crossover
           Add the offspring to the offspring population
           Evaluate objectives of each offspring
        End For
        Combine offspring population with the parent population
              to get intermediate population
        Perform Nondomination Sort on Intermediate chromosome population
        Select the best N chromosome based on rank and crowding distance
  Else
        Perform mutation over the entire population
        Perform Nondomination Sort
  End If
End For
```

Fig. 4. Modified NSGA-II

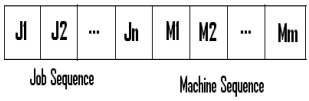


Fig. 5. Chromosome Representation

The main components of the algorithm are depicted below.

4.1 Chromosome Representation

The genotype of the chromosome is presented in Figure 5. The total size of the chromosome has been 12 – for 6 jobs' sequence and assignment of machines for 6 jobs. The algorithm for initialization is given in Figure 6.

```
INITIALIZATION

Initialize flag=0, j=1 and initialize elements of array flg[] to 0
While flag != 0
   Generate random number r
   If r <= 0.16 Then
         Assign temp = 1
   Elseif r > 0.16 & r <= 0.32 Then
         Assign temp = 2
   Elseif r > 0.32 & r <= 0.48 Then
         Assign temp = 3
   Elseif r > 0.48 & r <= 0.64 Then
         Assign temp = 4
   Elseif r > 0.64 & r <= 0.8 Then
         Assign temp = 5
   Else
         Assign temp = 6
   End If
   If flg [temp] != 1 Then
         Assign chromosome[j]=temp
         Calculate j=j+1
         Assign flg [temp] = 1
   End If
   If j > 6 Then
         Assign flag=1
   End If
End While
```

Fig. 6. Initialization of Chromosomes

```
ORDER CROSSOVER

Choose 2 chromosomes
Randomly generate two crossover sites r1 and r2
Copy genes from r1 to r2 from parent 1 in to child
Delete these particular gene values from parent 2
Copy the remaining gene values from parent 2
  to the empty positions of child 1 in the same order as in parent 2
```

Fig. 7. Order Crossover

Before selection is performed, the individuals in the population are ranked on the basis of non-domination: all non-dominated individuals are classified into one category (with a dummy fitness value, which is proportional to the population size, to provide an equal productive potential for these individuals). The crowding distance is also calculated (see equation (10) below) to keep a diverse front.

$$d_I = \sum_{i=1}^{B} \frac{f_i(x) - F}{f_i^{\max} - f_i^{\min}} \tag{10}$$

Where d_I : crowding distance for individual (chromosome) I; $f_i^{\max}$ and $f_i^{\min}$ are the maximum and minimum objective values of the i-th objective respectively.

4.2 Crossover and Mutation

In this paper, order crossover (Figure 7) has been performed based the structure of the chromosomes. The mutation algorithm is given in Figure 8.

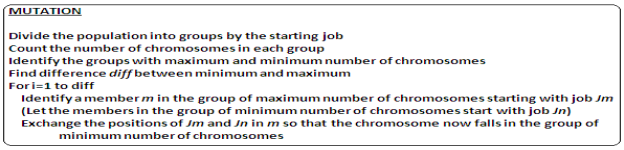


Fig. 8. Mutation Algorithm

The population is first divided into a number of groups starting with different job numbers. Next the number of chromosomes in each group is counted and the groups containing the maximum and minimum number of chromosomes and their difference are calculated. The resulting number represents the number of chromosomes which will be mutated. The resulting population is again subjected to Nondomination Sort in order to find the Nondominated chromosomes.

5 Experimentation and Results

The experimentation with Matlab has been conducted for crossover probabilities of 0.1 to 0.9 and generations 10 to 100. Figure 9 shows the Pareto optimal solutions for probability values of 0.5 (Figure a), 0.6 (Figure b), 0.7 (Figure c), whose results have been observed to be improving from probability 0.5 to 0.7. Each graph shows the Pareto optimal solutions for generations 50, to 100, since these generations have resulted better values.

The horizontal axis represents the weighted tardiness and the vertical axis represents the cost of deterioration. For generations 70, 80, 90 the Pareto optimal solutions show best performance for probability 0.7. A set of Pareto optimal solutions is shown in Figure 10 for probability 0.7 and for generations 50 to 100 (topmost row). Figure 11 shows the execution time data for various generation number and probabilities. From Figure 13 it is clearly observed that the execution time increases with the in the crossover probability and generation numbers.

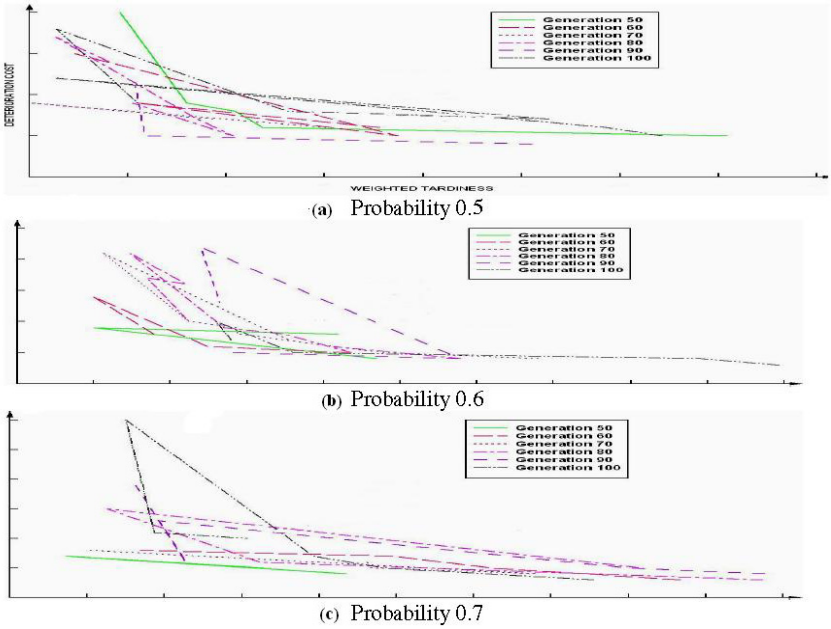


Fig. 9. Pareto optimal solutions

50		60		70		80		90		100	
1.2827	15	2.9031	13	1.6875	16	3.2205	13	3.2272	14	2.5655	13
0.51675	17	2.1653	15	1.6993	15	1.2712	16	2.7009	15	1.7542	15
1.603	14	1.79	17	2.3444	14	0.67866	25	0.87835	23	1.4779	17
		0.80625	18	0.60909	18	2.7577	15	0.98449	16	0.75173	40
								0.78945	29	0.86007	21
										1.2207	20

Fig. 10. Pareto Optimal Solutions

	Generations									
Prob.	10	20	30	40	50	60	70	80	90	100
0.1	1.156	2.219	3.203	4.359	5.266	6.406	7.406	8.25	9.078	10.203
0.2	1.297	2.625	3.344	4.593	5.422	6.703	8.172	8.735	10.297	11.531
0.3	1.265	2.828	3.843	4.719	6.266	6.922	8.172	9.891	10.172	12.094
0.4	1.547	2.328	4.063	5.496	5.875	8.141	9.172	9.859	12.281	13.813
0.5	1.594	3.031	4.079	5.781	6.656	7.672	9.563	11.11	12.765	13.281
0.6	1.547	2.812	4.531	6.172	7.641	8.672	10.422	12.063	13.453	15.296
0.7	1.859	3.375	4.719	6.672	7.75	9.406	10.938	11.609	14.157	16.218
0.8	1.734	3.75	5.468	6.515	8.468	9.782	11.609	12.844	15.203	17.046
0.9	1.922	3.547	5.547	7.296	8.64	10.593	12.031	13.266	16.14	17.719

Fig. 11. Execution Times

6 Conclusion

A multi-objective job sequencing problem with two objectives is formulated in this paper and has been solved by modified NSGA-II evolutionary multi-objective algorithm. The modification has been made by proposing a mutation algorithm which shows effective results.

References

1. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multi-objective Genetic Algorithm. NSGA–II. IEEE Transactions on Evolutionary Computation 6, 182–197 (2002)
2. Safa, E.F., Ihsan, S., Ays_egül, T., Manoj Kumar, T.: New solution methods for single machine bicriteria scheduling problem: Minimization of average flowtime and number of tardy jobs. European Journal of Operational Research 201, 89–98 (2010)

3. Karimi, N., Zandieh, M., Karamooz, H.R.: Bi-objective group scheduling in hybrid flexible flowshop: A multi-phase approach. Expert Systems with Applications 37, 4024–4032 (2010)
4. Yazdani, S.M.T., Jolai, F.: Optimal methods for batch processing problem with makespan and maximum lateness objectives. Applied Mathematical Modelling 34, 314–324 (2010)
5. Sha, D.Y., Lin, H.-H.: A multi-objective PSO for job-shop scheduling problems. Expert Systems with Applications 37, 1065–1070 (2010)
6. Reza, T.-M., Babak, J., Fariborz, J., Ali, G.: The use of a fuzzy multi-objective linear programming for solving a multi-objective single-machine scheduling problem. Applied Soft Computing 10, 919–925 (2010)
7. Huynh, T.N., Ameur, S.: Due dates assignment and JIT scheduling with equalsize jobs. European Journal of Operational Research 205, 280–289 (2010)
8. Luca, V., Dario, P., Carlo, M.: A tabu search algorithm for scheduling pharmaceutical packaging operations. European Journal of Operational Research 202, 538–546 (2010)
9. Wanga, L., Pan, Q.-K., Suganthan, P.N., Wang, W.-H., Wang, Y.-M.: A novel hybrid discrete differential evolution algorithm for blocking flow shop scheduling problems. Computers & Operations Research 37, 509–520 (2010)
10. Betul, Y., Mutlu, Y.M.: A multi-objective ant colony system algorithm for flow shop scheduling problem. Expert Systems with Applications 37, 1361–1368 (2010)
11. Zandieh, M., Karimi, N.: An adaptive multi-population genetic algorithm to solve the multi-objective group scheduling problem in hybrid flexible flowshop with sequence-dependent setup times. Journal of Intelligent Manufacturing (2010), doi:10.1007/s10845-009-03747

Verification Platform for FPGA Based Architecture

Adesh Panwar

BEL-Bangalore, India
adeshpanwar@bel.co.in

Abstract. System on Chip (SoC) refers to a system designed by integrating Intellectual Property (IP) cores such as CPUs, DSPs and various other high end functions on a single chip. Traditional simulation and emulation techniques for verification of such chips has become unaffordable due to increasing complexity, high cost, more time to market and large number of scenarios required to cover absolute verification. This paper describes new verification platform for complex embedded systems based on FPGA. The robustness and re-use capability of proposed approach makes it applicable at different integration levels and different phases of the project life cycle. Experiment and performance analysis on Radar Signal Processing module proves the effectiveness of the proposed platform.

1 Introduction

System design has become complex, where functionalities of Processors, Memories, and Communication networks, all are integrated into a single chip. An important aspect of such design is its Verification strategy. Verification is a means to ensure that the design meets the functional requirements as defined in the specification.

Verification of such chips become extremely challenging, as all the design blocks when verified at their levels, are reliable but when put together may lead to new problems. Dedicated interfaces between system modules, intrinsic system complexity and non detailed IP cores also create integration problems.

To verify any design, functional verification is used and it can be classified into simulation-based method and emulation-based method [1].Rapid Prototyping Systems (RPS) can also be used to accurately model the prototype of the intended SOC. It provides significantly higher simulation speed than software simulator [2].

Various FPGA based Emulation techniques have been proposed [3] [4] [5].Most of them have tried to reduce time and effort for verification. Proposed methodology is similar to them but much simpler and far more cost effective. Our methodology does not require any expensive tools, emulation boards or instruments. Also, it does not require learning a new language or modeling technique.

This paper is organized as follows: Section 2 gives a system overview. Section 3 describes the proposed environment and details about the verification flows. Section 4 report's conclusions and future works.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 393–398.
springerlink.com

2 System Overview

Radar signal processing like any other FPGA based design are complex in nature. As shown in fig.1, it involves various functional blocks which may be custom design or any IP core.

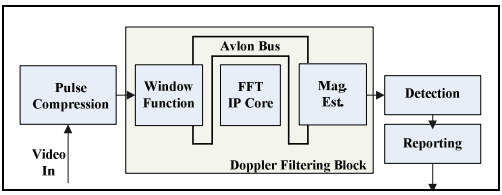


Fig. 1. Radar signal processing block diagram

This design is built around FPGA from Altera Corp. [6], which has 40K logic elements and more than 400 user I/Os. FPGA is configured using Small Outline Integrated Circuit package (SOIC) with 64 Mb capacities. A 16 bit, 10(MSPS) differential ADC is used to sample Video. A configurable SRAM from Cypress Semiconductor Corp. [7] is used for sample storage.

Doppler filtering block consist of Altera's FFT vs9.0 [6] core based on Avalon® Streaming (ST) compliant input and output interfaces. Quartus II Version 9.0 [6] provides Integrated Development Environment for this design development.

For the above design, it is required to verify all the components at different integration levels during all the development phases of design. Verification based on simulation is increasingly error prone, since simulation environments are often based on approximate computation [8] depending on level of the adopted model. So, test benches have to be rewritten for complete tests when transition from simulation to Emulation is done.

Just as the HDL based design blocks are reusable; it is desirable to reuse the verification platform to enhance overall productivity. If platform developed for one design can be used for other similar designs, a significant amount of the verification time is saved for subsequent design.

3 Proposed Verification Platform

In this paper, as shown in Fig. 2 we propose a low cost platform for the verification of complex system. This platform supports execution at different verification flow using the FPGA device performance. It is flexible, robust and quickly adapts to different abstraction levels of verification flows and offers significant reduction in time required to meet a new development phase.

An emulator is provided on the FPGA consisting of a HDL based controller and input-output wrappers. The controller reads data from an available DPRAM module on the board hosting the FPGA chip.

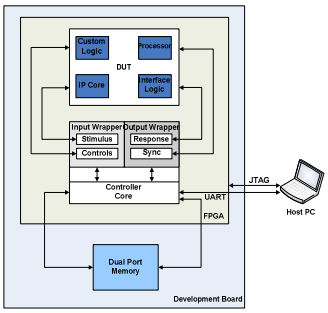


Fig. 2. Proposed Verification Platform Structure

The controller is interfaced with host PC over serial line. Through this channel commands and data are sent and emulation results are read out. On reception of initiation command form host PC, controller reads the stimulus from the DPRAM and provides it to the input wrapper at the DUT clock. The input wrapper provides the necessary controls along with the test pattern. In this way, modules with high pin counts and standard interfacing protocols can be verified without saturating number of pins available on FPGA. In addition, variety of DUVs can be interfaced with minor modification on Input and Output wrappers. The emulated response is captured back by the output wrapper and written into the DPRAM. It is sent to the host application for further analysis.

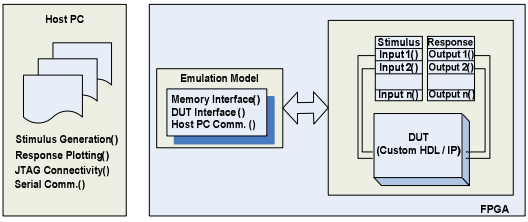


Fig. 3. Block diagram of the proposed wrapper

The proposed emulation platform is interfaced with a supporting software environment running on the host PC. As shown in fig.3, it is used to generate various test patterns, memory initialization, and response analysis.GUI provides selection among various patterns available and initiates the verification process. The I/O signals generated for the targeted module are elaborated to provide a binary format that is eventually sent to controller via the UART. Received responses from DUV are plotted, and are available for further analysis in the form of user readable files.

The following incremental verification methodology has been used-

3.1 Logic Simulation

For the verification of single processing module or preliminary integrated subsystems, the supporting environment includes RTL simulation and mathematical bit true models providing the DUV stimuli and expected response.

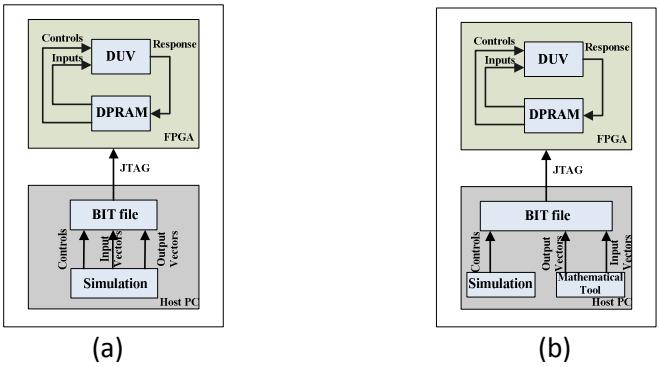


Fig. 4. Verification platform structure using (a) Logic Simulation (b) Equivalence check

Fig.4 (a) shows the platform structure using Logic Simulation. This approach has been used for standalone verification of Doppler filtering block consisting of IP core, windowing function and Magnitude estimation.

Altera-ModelSim [9] has been used to carry RTL Level Simulation. Test benches are written to provide stimulus to DUV (here, Doppler Filter Block). Both the stimulus and responses are recorded in a file and then written in a DPRAM. The emulated response is compared against the previously recorded values to verify the module's functionality.

3.2 Accelerated Simulation

In this method, a mix of simulation tools and mathematical tools (like MATLAB [10]) are used. For IP core verification, vendors generally provide mathematical bit true models along with the core. Input vectors and responses are calculated using these models in mathematical tools. Controls are generated thorough test bench. They are applied to the emulated DUV in combination with the simulation templates (Fig. 4(b)).

Here the main gain achieved is reduction in verification time as it does not require RTL simulation and performance evaluation is comparable to application like situations. This methodology is useful in case of IP cores verification and their integration with other blocks.

3.3 Emulation using HDL Controller

This verification mode is suitable for complex blocks, combination of blocks or block communicating over specified protocol. Here the stimulus and expected response are computed in the application program running on Host PC with help of mathematical models and tools. Stimulus is written in DPRAM using UART. Control signal generation, stimulus application, DPRAM interfacing and Host interfacing are performed using HDL controller. The emulator response is recorded on DPRAM and sends to host pc via UART.

The programmability of HDL Controller gives the flexibility of various test vector application without changing the setup. The software at host pc initiates this process and finally verifies the emulated output.

Slow communication with host PC is only required at set up. Here simulation is not required, thus greatly reducing the verification time. It facilitates verification at different abstraction levels.

3.4 Rapid Prototyping

This final verification strategy uses same previous environment with small modification for system level verification. The input and output wrappers are modified to mimic the proper system level protocols. This shows the robustness of attaching the wrappers to any module of DUV.

Distinct advantages of our proposed strategy are:

1) It is based on HDL controller that allows managing the whole emulation process.
2) Proposed system does not suffer any pin-out constraints, since wrappers are sizable according to the DUV requirements.
3) This guarantees a complete scalable approach, with possibility to interconnect new custom logics.
4) Proposed platform doesn't require any expensive equipment.
5) Standard HDL based design which eliminates need to learn a new language.

4 Conclusion

The proposed platform environment has been implemented and applied for the verification of a complex integrated system under development for Radar signal processing.

Table 1. Occupation figures regarding the implementation of this methodology for Doppler Filtering Block

Family	Stratix III
Devices	EP3SE
Logic Utilization	14%
DSP Block 18 bit	4%
Block Memory Bits	14%

Acknowledgments. Authors are very much thankful to General Manager of MR SBU in Bharat Electronics Ltd., Bangalore to give us an opportunity for publishing this technical paper. Thanks to our dear colleagues whose kind supports make this study and design possible.

References

1. Pixley, C.: Functional verification 2003: technology, tools and methodology. In: International Conference on ASIC, vol. 1, pp. 1–5 (October 2003)
2. Rashinkar, P., Paterson, P., Singh, L.: System on A Chip Verification- Methodology and Techniques (2002)

3. Costa, D.: ESFFI-a novel technique for the emulation of software faults in COST components. In: IEEE Int'l Conference and Workshop on Engineering of Computer Based Systems, pp. 197–204 (2001)
4. Kafka, L., Danek, M., Novak, O.: A Novel Emulation Technique that Preserves Circuit Structures and Timing. In: IEEE International Symposium on System-on-Chip, pp. 1–4 (2007)
5. Black, B., Shen, J.: Calibration of Microprocessor Performance Models. IEEE Computer 31(35), 59–65 (1998)
6. http://www.Altera.com
7. http://cypress.com
8. Wolf, W.: A decade of hardware/software co design. IEEE Computer 36, 38–43 (2003)
9. ModelSim® Altera User's Manual Version 6.1 (2006)
10. Matlab R2009b

Implementation and Analysis of Downlink Scheduling for IEEE 802.16 Using Controlled Priority Queuing

Z.M. Patel and U.D. Dalal

S.V. National Institute of Technology,
Suart, India
{zmp,udd}@eced.svnit.ac.in

Abstract. Scheduling algorithms play vital role in MAC layer for wireless broadband networks such as WiMAX. Though WiMAX standard defines scheduling service classes, it does not specify the actual packet scheduling mechanism to achieve QoS guarantee. This paper discusses scheduling objectives, scheduling algorithms and resource allocation strategies in WiMAX network considering PMP mode and time division duplex (TDD) operation. A novel scheduling scheme called Controlled Priority Queuing (CPQ) is proposed which takes into account available bandwidth and priority of service queues in making scheduling decisions. CPQ is implemented in VHDL along with well-known RR and DRR algorithms. The performance is compared in terms of throughput and queuing delay considering real hardware implementation scenario. It was observed that CPQ throughput is found better in case of rtPS and nrtPS service class.

1 Introduction

The success of next generation wireless technologies depends on the performance of their schedulers to deliver high data throughput and meet Quality-of-Service (QoS) commitments. Keeping this in mind, the MAC layer of IEEE 802.16 standard [1,2] is designed to support variety of applications and services through its QoS support [3]. With fast air link, asymmetric downlink/uplink capability, fine resource granularity and a flexible resource allocation mechanism, WiMAX can meet QoS requirements for a wide range of data services and applications.

QoS mechanism of WiMAX [4] classifies traffic in service flows and it is possible to assign QoS requirement per flow. The IEEE 802.16 MAC defines five types of service flow or QoS classes: Unsolicited Grant Service (UGS), real-time Polling Service (rtPS), extended real-time Polling Service (ertPS), non real-time Polling Service (nrtPS) and Best Effort (BE). UGS is designed to support constant bit rate (CBR) traffic such as T1/E1 and VoIP without silence suppression. The rtPS supports real-time service flows that generate variable bit rate (VBR) traffic on a periodic basis e.g. video or audio streaming. An ertPS gets allocations periodically but its size is dynamic. Thus it combines features of UGS and rtPS and supports application such as VoIP with silence suppression. nrtPS is designed for non real-time service flows that are delay-tolerant but may need high bandwidth such as File Transfer Protocol (FTP).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 399–406.
springerlink.com

2 QoS and Scheduling

2.1 QoS Provisioning and Scheduling Objectives

The basic approach in provisioning QoS in IEEE 802.16 PMP networks is that Base Station (BS) does the scheduling for both uplink (UL) and downlink (DL) transmission. The data packets are associated to service flows with well defined QoS parameters so that the BS scheduler can correctly determine the packet transmission ordering over the air interface.

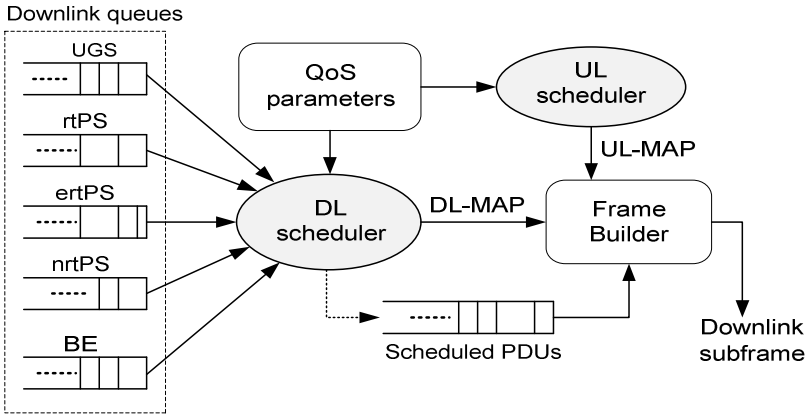


Fig. 1. Base Station scheduler operation with serveice queues

The MAC layer is connection oriented and supports both continuous and bursty traffic. So it can handle constant bit rate (CBR), real time variable bit rate (VBR) and non real time VBR and BE traffic. By ensuring proper resource allocation, scheduler tries to meet QoS requirements of various traffic flows. The QoS requirements are supplied by subscriber stations (SSs) in dynamic service addition (DSA) and dynamic service change (DSC) MAC management messages [1,2] at the time of connection set up. These requirements are specified in terms of minimum reserved transmission rate (MRTR), maximum sustained rate (MSTR), tolerated jitter, latency, traffic priority and loss rate. In addition to this, subscriber station (SS) can explicitly requests bandwidth during connection. As shown in Fig. 1, BS scheduler analyzes QoS parameters, downlink queue state information and bandwidth requests to allocate slots to each SS in uplink and downlink directions.

Fig. 2 shows how scheduling decisions are carried in the downlink sub-frame of 802.16 PMP frame. DL-MAP contains information about downlink grants whereas UL-MAP contains information about time given to each SS to access channel in the immediately following uplink sub-frame.

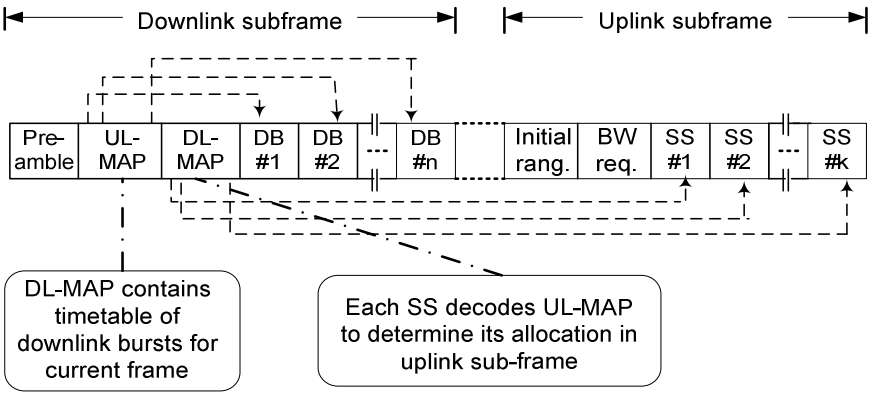


Fig. 2. WiMAX frame structure emphasizing scheduling aspect

The scheduling scheme must be able to

(1) allocate minimum number of slots to each connection to fulfill basic QoS need e.g. MRTR
(2) allocate unused slots to some connections to maximize the resource usage
(3) order the slots to improve delay and jitter

2.2 Slot Allocation

The slot allocation [5] is based on the type of service flow, QoS requirements, bandwidth (BW) request size B_i and state of queues. Suppose that S_i stands for slot size i.e. the number of bytes i^{th} connection can send in one slot. Though slot duration is fixed, the number of bytes in a slot may vary depending upon the current modulation and coding scheme (MCS) associated with that connection. If *FPS* denotes number of frames per second then the number of slots N_i within each frame for i^{th} connection can be calculated by using following expression

$$N_i = \frac{B_i}{FPS * S_i} = N(B_i, S_i) \tag{1}$$

The result of expression (1) can be floating point number, while the number of slots is always an integer value. So one could round the result to the largest integer value. Besides, for flexible resource allocation, one would consider bandwidth (BW) request size R_i for uplink allocations. Thus no. of slots can be calculated from following expressions.

$$N_i = \left\lceil \frac{R_i}{S_i} \right\rceil = R(R_i, S_i) \tag{2}$$

The equation (2) can be also applied to *downlink connections* with only difference that *queue size* Q_i instead of request size will be taken into account.

3 VHDL Implementation

In this section, implementation of two well known scheduling algorithms RR and DRR and new proposed scheme named as CPQ will be discussed. The complexity of these scheduling schemes is lower than EDF[6], WFQ[7] and hierarchical[8] schedulers. These three schedulers are implemented in VHDL and their output analyzed to estimate *average throughput* and *queuing delay*. Following are variables/functions used in the pseudo-code for the scheduling algorithms:

- *ActiveList*: list of indices of nonempty service queues
- *ExtractFlow*(*p*): extracts flow type from packet *p*
- *ExistInActiveList(i)*: checks whether service flow *i* exist in ActiveList
- *UpdateActiveList(i)*: updates *ActiveList* when service flow *i* is added or removed
- *Enqueue(i,p)* : insert packet *p* in service queue *i*
- *Dequeue* ($queue_i$): removes a packet from head of service queue *i*
- *NextActiveList(i)*: selects next lower priority flow after *i* from ActiveList
- *Complete MAP_ message()*: forms DL-MAP message for current frame
- *Send TDD_frame()*: sends TDD downlink subframe to PHY layer

The upper layer packets are first classified and inserted into appropriate QoS class queue (buffer). This operation is common for all type of scheduler.

3.1 RR Scheduling

In RR, packets from each queue are selected in cyclic manner serving one packet at a time in round-robin order. Empty queues are skipped. The primary benefit of RR is that an extremely bursty or misbehaving flow does not degrade the quality of service delivered to other flows, because each flow is isolated into its own queue.

3.2 DRR Scheduling

Deficit-Round-Robin (DRR) is a well-known scheduling algorithm [9] originally developed for IP networks. DRR scheduling combines the ability of providing fair queuing in the presence of variable length packets with the simplicity of implementation. DRR defines the state variable *deficit Counter (DCi)* and allocation *quantum(Qi)* for each queue. The *DCi* of each active queue increased by *quantum(Qi)* when the queue has its turn. If the packet at the head of the queue is less than or equal to the variable *DCi*, the variable *DCi* is reduced by the size of packet and the packet is removed and sent to output port. The process will be repeated until either the *DCi* is less than or equal to zero or the queue becomes empty. When these conditions occur, the scheduler moves on to serve the next non-empty queue. If the queue is empty, the value of *DCi* is reset to zero.

3.3 Controlled Priority Queue (CPQ) Scheduling

Priority Queuing is the basis for a class of queue scheduling that are designed to provide simple method of supporting differentiated service [4] classes. But strict priority scheduling causes starvation and hence excessive delay and drop rate of low priority queues. Controlled Priority Queuing allows packets in a high-priority queue to be scheduled before packets in lower-priority queues only if the amount of traffic in the high-priority queue stays below a user-configured threshold. The pseudo code for CPQ (Fig.3) uses following variables/functions.

- λ_i : fraction by which queue *i* is allowed to use available bandwidth
- UB_i : bandwidth utilized by service queue *i* in present round
- *ABW* : available bandwidth
- *Update_ABW()*: function that updates available bandwidth at the end of each round

```
While (TRUE) do
        If ( ActiveList is not empty )
           i = an index of selected non-empty service queue
        While (TRUE) do
          If  (UB[i] < (ABW * λ[i]) and  queue_i not empty) {
                PacketSize = Size(Head (queue_i) );
                Dequeue (queue_i);
                UB[i] = UB[i]  + PacketSize;    }
           Else  {
                UB[i]=0;        // reset utilized bandwidth
                NextActiveList(i);       // select next  queue
             If (round is over )  update_ABW( ) ;     }
           If (Empty (queue_i) )  {
                   UB[i]=0;
                   UpdateActiveList(i);       // remove flow i
                   NextActiveList(i);
                If (round  is over ) update_ABW( );
                      break;  }
            If (ABW  ≤ 0)  {        // Is BW exhausted ?
                   Complete MAP_message( );
                   Send TDD_frame( );     }
          End while
End while
```

Fig. 3. Pseudo code for CPQ scheduling algorithm

Initially highest priority queue is selected to send packets. However it is only allowed to use *fraction* of Available Bandwidth (ABW) and when the usage exceeds the limit next lower priority queue will be served. Each queue has its own fraction by which it can utilize available bandwidth. ABW is updated at the end of *every round*

and utilized bandwidth (UB) is updated when packet is served. Unlike CPQ, DRR do not take in to account available bandwidth, instead it relies on Quantum and Deficit counter of each queue.

4 Simulation Results

The simulation assumes different input data rate for different class of queues. The input data rate is 1.99 Mbps for UGS, 1.136 Mbps for ertPS and rtPS and 0.568 Mbps for nrtPS and BE class of queue. The output of scheduler is analyzed for each service class. From this, it is possible to evaluate average throughput and average queuing delay of each service class. Then, the network traffic (i.e. no. of SS) can be increased gradually from low to heavy and its effect on average throughput and average queuing delay can be determined.

Plots from Fig. 4 to 7 show average throughput with RR, DRR and CPQ scheduling schemes. It is observed that CPQ offers better average throughput performance than DRR in case of rtPS and nrtPS class of applications. For UGS and BE classes, CPQ throughput is slightly lower than DRR under heavy traffic conditions. RR scheduler maintain average throughput constant for nrtPS and BE flows because total input traffic is low in these flows as compared to UGS and rtPS.

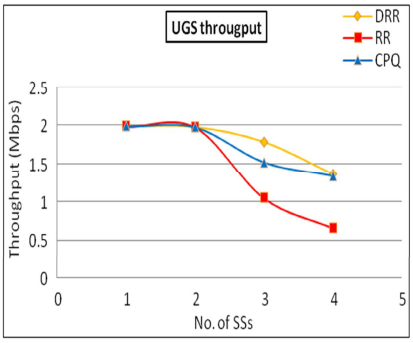


Fig. 4. Average Throughput - UGS

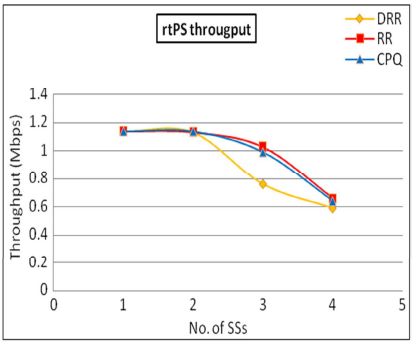


Fig. 5. Average Throughput – rtPS

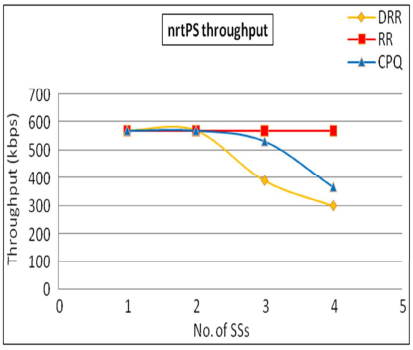


Fig. 6. Average Throughput – nrtPS

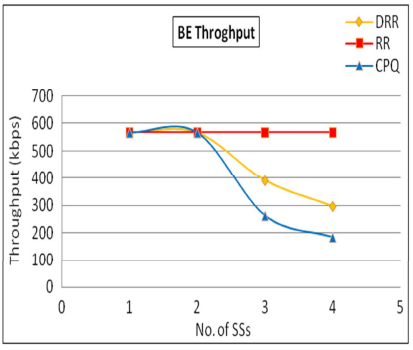


Fig. 7. Average Throughput - BE

The queuing delay of each service class is plotted from Fig. 8 to 11 considering RR, DRR and CPQ schemes. The results reveal that delay performance of RR is worst of all and that the controlled priority queuing (CPQ) scheme has less queuing delay than DRR for rtPS service class. For other classes, queuing delay of CPQ is almost equal to DRR.

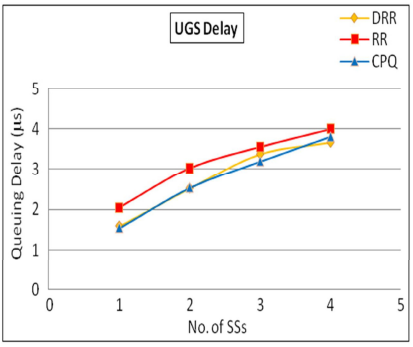


Fig. 8. Queuing Delay - UGS

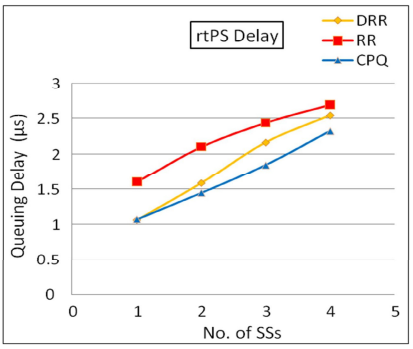


Fig. 9. Queuing Delay – rtPS

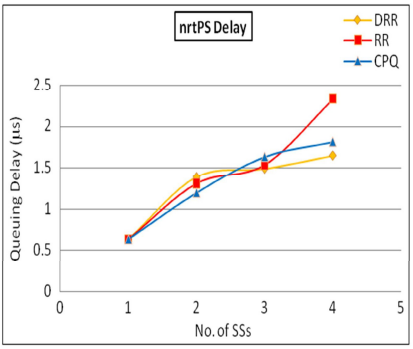


Fig. 10. Queuing Delay - nrtPS

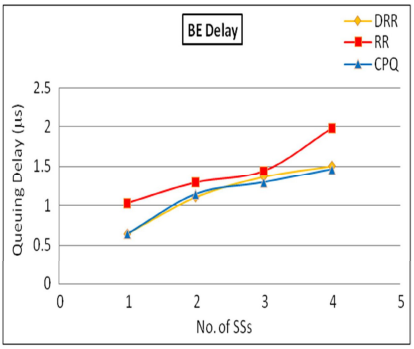


Fig. 11. Queuing Delay - BE

5 Conclusions

In this paper, scheduling solution for 802.16 base station is presented. The proposed CPQ algorithm for scheduling is based on simple and fast round robin concept. It is also not computationally expensive so scheduler will not burden BS with extensive calculations. It is implemented in VHDL and its performance in terms of average throughput and average queuing delay is evaluated using Modelsim simulator. The evaluation takes into account realistic HDL implementation scenario and packet analysis.

Two other classical schemes, RR and DRR are also implemented in VHDL and their performance is compared with CPQ. CPQ improves both average throughput and queuing delay for *rtPS* flow where as it improves average throughput for *nrtPS* flow.

References

1. IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Inter-face for Fixed Broadband Wireless Access Systems. ANSI/IEEE Std. 802.16-2004 (revision of IEEE Std. 802.16-2001)
2. IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Inter-face for Fixed Broadband Wireless Access Systems. Amendment 2:Physical and Medium Access Control Layers for Combined Fixed and Mobile Opera-tion in Licensed Bands. ANSI/IEEE Std 802.16e-2005
3. Ahmet, Y., Ivanovich, M., Yegin, A.: Survey of MAC based QoS implementations for WiMAX networks. Computer Networks: The International Journal of Computer and Telecommunications Networking 53(14), 2517–2536 (2009)
4. Cicconetti, C., Erta, A., Lenzini, L., Mingozzi, E.: Performance Evaluation of the IEEE 802.16 MAC for QoS support. IEEE Transactions on Mobile Computing 6(1) (January 2007)
5. Sayenko, A., Alanen, O., Hamalanen, T.: Scheduling solution for IEEE 802.16 Base station. Computer Networks: The International Journal of Computer and Telecommunication 52(1), 96–115 (2008)
6. Sivaraman, V., Chiussi, F.M.: Statistical Analysis of Delay Bound Violations at an Earliest Deadline First (EDF) Scheduler. Performance Evaluation: An International Journal 36-37, 457–470 (1999)
7. Demers, A., Keshav, S., Shenkar, S.: Analysis and Simulation of Fair Queuing Algorithms. Journal of Internetworking and Research 1, 3–26 (1990)
8. Wang, Dittman: Adaptive Radio Resource Allocation in Hierarchical QoS Scheduling for IEEE 802.16 Systems. In: 50th IEEE Global Telecommunications Conference (IEEE GLOBECOM 2007), pp. 4769–4774 (November 2007)
9. Shreedher, M., Varghese, G.: Efficient Fair Queuing using Deficit Round Robin. IEEE/ACM Transaction on Networking 4(3), 375–385 (1996)

Mechanism for Secure Content Publishing for Reporting Platform Hosted on Public Could Infrastructure

Bhanu Prakash Gopularam and Nalini N.

CSE Dept, NMIT,
Bangalore, India, 560 064
bhanuprakash.gopularam@gmail.com
nalinaniranjan@hotmail.com

Abstract. Cloud computing works on various service models like SaaS, PaaS, IaaS. The enterprises can outsource data and computation to cloud and benefit from cloud computing unique attributes. This paradigm also brings forth many challenges for data security and access control. A reporting platform is software which allows users to access content within it. The content hosted on reporting platform is developed by content publishers who are worried about intellectual property rights and content protection. The content contains data configuration information as well as database access query (sql-query) that needs to be run against a database. Upon request from user, the reporting platform connects to a database and executes the content and returns the transformed output. Later the outcome is formatted to user understandable format and delivered to user. When the reporting platform is deployed on public cloud environment one needs to provide stringent security for data in rest and in motion. The different entities accessing the content may reside in an untrusted domain and some of the parties (viz. database provider) may reside in a different enterprise cloud and needs to be accessed while serving the user request. In this work, we propose a generic scheme to enable content protection and fine-grained access control of the published data and protecting the data even from cloud providers. One unique problem for which we provide a solution is that the data confidentiality is ensured even when some computation is required on the content in cloud environment.

1 Introduction

Cloud computing provides computation, software applications, data access, data management and storage resources without requiring cloud users to know the location and other details of the computing infrastructure. Cloud computing infrastructures enable companies to cut costs by outsourcing computations on-demand. Cloud services are offered in different service models viz. Infrastructure as a Service (IaaS), Platform as a Service(PaaS) and Software as a Service(SaaS) The typical cloud deployment models are public cloud, community cloud, hybrid cloud and private. The main cloud information security objectives are dependability, trustworthiness, survivability (resilience). [U.S. DoD Software Assurance Initiatives]. Also the data confidentiality,

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 407–414.
springerlink.com

integrity and availability (CIA traid) are 3 important concepts of cloud software assurance for information system security.

Encryption transforms data using cryptographic algorithms. The Attribute Based Encryption is classified as CP-ABE (Ciphertext Policy-Attribute Based Encryption) [8] and KP-ABE (Key Policy-Attribute Based Encryption)[4]. In this paper, we use KP-ABE[4] for data transformations. Data is encrypted by set of attributes and user secret key is associated with a access structure. The internal nodes are threshold gates and leaf nodes are associated with attributes which is used to encrypt data. If encrypted data's attribute satisfy user secret key's access structure, user is able to decrypt a ciphertext.

The model also uses Proxy Re-Encryption scheme which aids in resolving issues with User revocation[3]. Here the proxy is able to convert ciphertext encrypted under Alice's public key into ciphertext that can be decrypted by Bob's secret key, this is called as Proxy Re-Encryption(PRE) technique.

Bilinear Pairings and Computation Assumptions: Let G1 and G2 be two cyclic groups of the same prime order q. A bilinear pairing is a map e : G1×G1 → G2 which satisfies the following properties:

- Bilinear: e(ga1, gb2) = e(g1, g2)ab for all g1, g2 ϵ G1 and a, b ϵ Z*q .
- Non-degenerate: there exists g1, g2 ϵ G1 such that e(g1, g2) ≠ 1.
- Computable: there is an efficient algorithm to compute e(g1, g2) for g1, g2 ϵ G1.

Homomorphic encryption is a form of encryption where a specific algebraic operation performed on the plaintext is equivalent to another (possibly different) algebraic operation performed on the ciphertext. There are two kinds of cryptosystems devised based on correlation of cipher text for algebric operations. One is called partially homomorphic cryptosystems, viz Benaloh's cryptosystem which processes longer blocks of data at once. Fully homomorphic encryption using ideal lattices that is limited to evaluating low-degree polynomials over encrypted data was proposed by Gentry[10].

The reporting platform is a kind of SaaS application which can be deployed on cloud environment. The content typically constitute meta-data with configuration information and data decryption key and a database query which when run against underlying database would fetch select data.

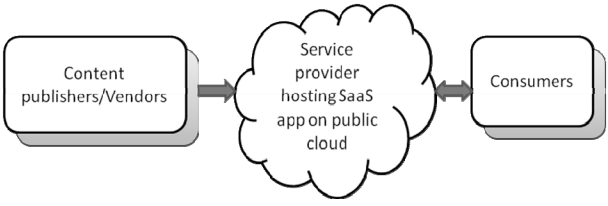


Fig. 1. Interaction of Content publishers with Consumers outlined

The interaction of content publishers with reporting platform as shown in Fig. 1 is not direct but the still content publishers will be able to exercise content access by adding access control headers to the content using KP-ABE scheme. The users take reporting platform help for content processing and getting the results.

2 Related Work

The Attribute based encryption has captured the attention in recent years, which can be regarded as a special type of identity based encryption integrated with flexible access control policy.

In 1985, Shamir [1] first proposed the concept of identity based public key cryptography, in which the public key of an entity can be easily computed from his identity information, and the private key of an entity was generated from his identity information and a master key of a trusted third party called a Private Key Generator (PKG). In 2001, the first practical identity based encryption scheme was presented by Boneh and Franklin [2].

In Eurocrypt 2005, Sahai and Waters [3] first introduced the concept of Fuzzy Identity Based Encryption (Fuzzy IBE), in which identities are regarded as a set of descriptive attributes instead of a string of characters in previous IBE systems. In 2006, Goyal et al. [4] introduced the notion of Key-Policy Attribute-Based Encryption (KP-ABE) for fine-grained sharing of encrypted data and proposed a KP-ABE scheme that allows any monotone access structures. Bethencourt et al. [6] presented the first construction of Ciphertext policy attribute based encryption (CP-ABE). To ensure access structure requirements, in [6] there was proposed a system model using Key Policy-Attribute Based Encryption (KP-ABE) and Proxy Re-Encryption (PRE). Formally, [6] ensures data confidentiality using KP-ABE and sending the data owner delegate computation overload to the proxy using PRE.

In [6], Gentry proposed a fully homomorphic encryption scheme that enables to perform an arbitrary number of arithmetic operations on encrypted data. This increases the computation time but the benefits associated with it are worth the processing overhead. The current system model plans to use a database which supports operations on encrypted data [7]. The queries provided will be encrypted using homomorphic techniques and they are executed as is by the cloud provider.

3 Overview of the Problem

The content publishers create and publish content which is consumed by users. Only the authorized users should be able to use the content and user's actions are guided by permissions given by content publishers. The reporting platform is installed on public cloud, the content confidentiality should be ensured when it is in rest or in motion. The content published should not be modifiable by any intermediate parties. The sensitive data needs to be kept confidential while it is passing between untrusted servers. The users would use reporting platform for storing the content and to perform computations before it is consumed. The cloud resources will be utilized for this and so it should be secured even from cloud provider as they may intercept or decipher the data.

3.1 Participating Entities Description

Content publisher: capable of creating and publishing new content. The newly developed content will be available for download.

User or Reporting platform user: Users are registered with reporting platform before accessing the content. User gets the content published by different content publishers by email or from website. Imports the content into reporting platform and accesses it for day to day operations.

Reporting platform: The reporting platform provides a way to access previously imported content. It manages the user list and access to content is controlled based on inherited from content itself. The database query is executed on underlying database and results are formatted (tables, charts etc) before sent to the user.

Database provider: This is the underlying database upon which the content queries are executed. The reporting platform can have multiple databases configured. The data part of content (sql-query) is executed against database.

4 Design of Secure Content Publishing Network

For achieving a secure, scalable, fine-grained access control, data abstraction for the content that is published on public cloud, we uniquely combine four advanced cryptographic techniques viz. Key-policy Attribute based encryption(KP-ABE), Proxy Re-Encryption(PRE), Lazy Re-Encryption and Homomorphic encryption.

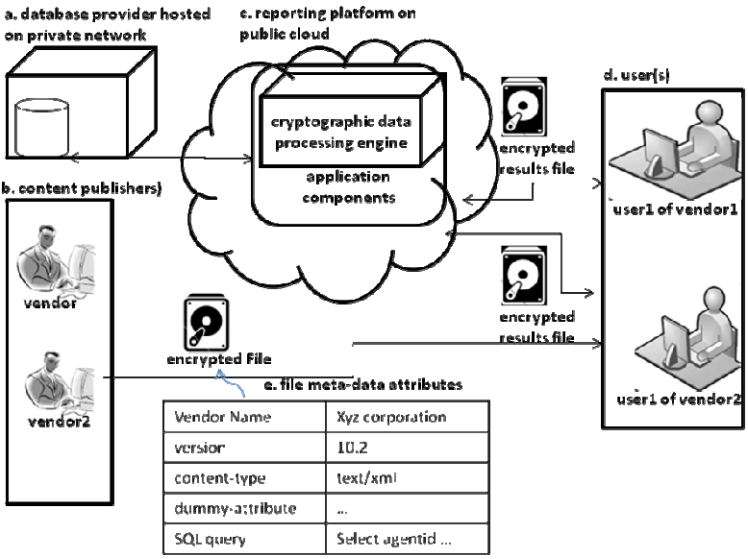


Fig. 2. Reporting platform deployment on public cloud. The figure illustrates different entities of the proposed model and the content passage.

Notation	Description
PK, MK	system public key and master key
T_i	public key component for attribute i
t_i	master key component for attribute i
SK	user secret key
sk_i	user secret key component for attribute i
E_i	ciphertext component for attribute i
I	attribute set assigned to a data file
DEK	symmetric data encryption key of a data file
P	user access structure
L_P	set of attributes attached to leaf nodes of P
Att_D	the dummy attribute
UL	the system user list
AHL_i	attribute history list for attribute i
$rk_{i \leftrightarrow i'}$	proxy re-encryption key for attribute i from its current version to the updated version i'
$\delta_{O,X}$	the data owner's signature on message X

Fig. 3. Notation used in the scheme description

4.1 Proposed Scheme Description

For content encryption a combination of KP-ABE scheme and PRE scheme is used. For further reducing the computation overhead, one can use lazy re-encryption technique and allow cloud provider to "aggregate" computation tasks of multiple system operations. Content is composed of two parts, meta-data and database query. The meta-data comprises of data configuration settings, access control attributes and data decryption key. These are encrypted using the KP-ABE scheme. So only users having secret keys supplied by content publisher can decrypt them and use it.

The detailed interaction between different parties of the content publishing network is outlined below and shown in Fig. 2.

4.1.1 Content Publisher Creates New Content

The content publisher generates a public key and a master key using Setup(k) algorithm and includes the public key in to the content.

4.1.2 Users Imports Content into Reporting Platform

The registered users import the content into reporting platform. While importing the users access details are set based on permissions information present as part of content itself. The reporting platform uses SK for decrypting the header part of content and user access is controlled there after.

The data part of the content contains encrypted database query and it is stored as is in local database for serving future requests.

4.1.3 Users Accessing Content

The user sends request to access content to reporting platform, the request must contain content identifier and publisher identification information.

The reporting platform validates the user details from UL existing user list and it retrieves the encrypted content record (database query) and sends it to database provider for execution. The database query can be encrypted in two ways for security reasons. Homomorphic encryption: The database supports homomorphic query

execution this encryption scheme can be used. One example a CryptDB[9] and there are several other which support this partially. The steps for query interpretation are mentioned in Analysis section.

KP-ABE scheme: If the database do not support homomorphic queries (one reason could be legacy information systems), the secure content execution is done by encrypting using KP-ABE scheme. The content publisher for exchanges private-public key with database provider and based on KP-ABE scheme. During query execute request, the query is decrypted and run. This way security of data is ensured while it is in transit and only authorized access is possible.

Once the results of query execution are received and it does some more transformations on the data (like view creation, formatting), the computing intensive operations done with help from cloud provider.

The final result is encoded using proxy re-encryption scheme and sent back to User. User obtains DEK by using Decryption(P, SK, E) algorithm with secret key $sk`_I$ and access structure P.

4.2 Key Policy Attribute Based Encryption

The KP-ABE scheme used for secret key generation is explained here, the algorithm is sourced from [4] where KP-ABE was originally proposed:

4.2.1 Setup

This algorithm takes as input a security parameter κ and the attribute universe $U = \{1, 2, \ldots, N\}$ of cardinality N. It returns the public key PK as well as a system master key MK as follows

$$PK = (Y, T_1, T_2, \ldots, T_N) \tag{1}$$

$$MK = (y, t_1, t_2, \ldots, t_N) \tag{2}$$

where $T_i \in G_1$ and $t_i \in Z_p$ are for attribute i, $1 \leq i \leq N$, and $Y \in G_2$ is another public key component. We have $T_i = g_{t_i}$ and $Y = e(g, g)_y$, $y \in Z_p$. While PK is publicly known to all the parties in the system, MK is kept as a secret by the authority party.

4.2.2 Encryption

This algorithm takes a message M, the public key PK, and a set of attributes I as input. It outputs the ciphertext E with the following format:

$$E = (I, \tilde{} E, \{E_i\}_{i \in I}) \tag{3}$$

where $\tilde{}E = MY_s$, $E_i = T_{s\,i}$, and s is randomly chosen from Z_p

4.2.3 Key Generation

This algorithm takes as input an access tree T, the master key MK, and the public key PK. It outputs a user secret key SK as follows. First, it defines a random polynomial $p_i(x)$ for each node i of T in the top-down manner starting from the root node r. For each non-root node j, $p_j(0) = p_{parent(j)}(idx(j))$ where $parent(j)$ represents j's parent and $idx(j)$ is j's unique index given by its parent. For the root node r, $p_r(0) = y$. Then it outputs SK as follows.

$$SK = \{sk_i\}_{i \in L} \tag{4}$$

where L denotes the set of attributes attached to the leaf nodes of T.

4.2.4 Decryption

This algorithm takes as input the ciphertext E encrypted under the attribute set I, the user's secret key SK for access tree T, and the public key PK. It first computes $e(E_i, sk_i) = e(g, g)^{p_i(0)s}$ for leaf nodes. Then, it aggregates these pairing results in the bottom-up manner using the polynomial interpolation technique.

5 Analysis of the Approach

We devised a generic mechanism for content publishing using several known encryption schemes. The existing models of content publishing which are kind of application-store model don't support processing of content. The existing models are limited to providing content as-is without any further processing. The scheme which even supports content publishing that includes some processing before it is actually consumed by intended users. The model supports secure computations on the content by using homomorphic encryption scheme[7] mentioned. The data confidentiality is ensured at every stage of content processing. The feasibility of the approach is described along with design considerations and experimental results will be made part of forth coming paper from us.

Processing a query in database which supports cryptographic operations[9] involves following steps:

1. The application issues a query, which the proxy intercepts and rewrites: it anonymizes each table and column name, using master key, encrypts each constant in the query with an encryption scheme suited for desired operation.
2. The proxy checks if the DBMS server should be given keys to adjust encryption layers before executing the query, and if so, issues an UPDATE query at the DBMS server that invokes a UDF to adjust the encryption layer of the appropriate columns.
3. The proxy forwards the encrypted query to the DBMS server, which executes it using standard SQL (occasionally invoking UDFs for keyword search).
4. The DBMS server returns the (encrypted) query result, which the proxy decrypts and returns to the application.

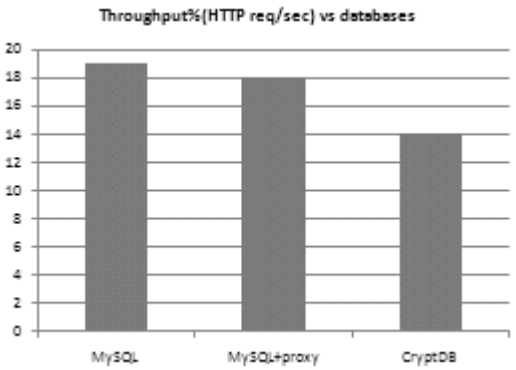


Fig. 4. The throughput of CryptDB compared to MySQL and MySQL+Proxy(the executor using proxy in between)

The analysis as shown in Fig. 4 indicates that throughput is slightly less when compared to standard SQL servers as indicated in [9]. Some techniques like Ciphertext precomputing and caching which caches the results of frequently used keywords encrypted text, Known query set where a content publisher uses specific queries in training mode for adjusting onion layers and avoid runtime learning readjustment.

6 Conclusion

The cloud service provider cannot be totally trusted due to risk of data security and violation of privacy factors. Deploying a reporting platform on public cloud poses challenges. We propose a generic scheme for end-to-end reliable data transfer along with secure computations on a public cloud environment which can be leveraged by any content publishing model. With this model the content privacy is ensured from users and the involving parties viz. reporting platform, cloud provider will not able to learn from anything from the data.

Acknowledgments. The author wish to thank members of Nitte Meenakshi Institute of Technology, Bangalore and CBABU Business Unit, Customer Collaboration (CCG), Cisco Systems, Bangalore for their help in completing this work.

References

1. Shamir, A.: Identity-Based Cryptosystems and Signature Schemes. In: Blakely, G.R., Chaum, D. (eds.) CRYPTO 1984. LNCS, vol. 196, pp. 47–53. Springer, Heidelberg (1985)
2. Boneh, D., Franklin, M.: Identity-Based Encryption from the Weil Pairing. In: Kilian, J. (ed.) CRYPTO 2001. LNCS, vol. 2139, pp. 213–229. Springer, Heidelberg (2001)
3. Yu, S., Wang, C., Ren, K., Lou, W.: Achieving Secure, Scalable and Fine-grained Data Access Control in Cloud Computing. In: IEEE INFOCOM 2010 (2010)
4. Goyal, V., Pandey, O., Sahai, A., Waters, B.: Attribute Based Encryption for Fine-Grained Access Conrol of Encrypted Data. In: ACM Conference on Computer and Communications Security (2006)
5. Lewko, A., Okamoto, T., Sahai, A., Takashima, K., Waters, B.: Fully Secure Functional Encryption: Attribute-Based Encryption and (Hierarchical) Inner Product Encryption. In: Gilbert, H. (ed.) EUROCRYPT 2010. LNCS, vol. 6110, pp. 62–91. Springer, Heidelberg (2010)
6. Wang, G., Liu, Q., Wu, J.: Hierarchical attribute-based encryption for fine-grained access control in cloud storage services. In: Proc. ACM Conference on Computer and Communications Security
7. Gentry, C.: A fully Homomorphic encryption scheme. Stanford University (September 2009)
8. Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-policy attribute based encryption. In: Proc. of IEEE Symposium on S&P (2007)
9. Popa, R.A., Redfield, C.M.S., Zeldovich, N., Balakrishnan, H.: CryptDB: Protecting Confidentiality with Encrypted Query Processing. In: MIT CSAIL
10. Wang, C., Liu, Y.: A Secure and Efficient Key-Policy Attribute Based Key Encryption Scheme. In: 1st International Conference on Information Science and Engineering (ICISE26)

A Semi-Interquartile Min-Min Max-Min (SIM^2) Approach for Grid Task Scheduling

Sanjaya Kumar Panda, Sourav Kumar Bhoi, and Pabitra Mohan Khilar

Department of Computer Science and Engineering
National Institute of Technology, Rourkela, India
{sanjayauce,souravbhoi}@gmail.com,
pmkhilar@nitrkl.ac.in

Abstract. Task scheduling on distributed computing is a NP-Complete problem. It is a great challenge for the system to preserve and enhance its performance. A schedule is said to be optimal if it gives a robust solution by proper utilization of the resources. In our paper, we have proposed Semi-Interquartile Min-Min Max-Min (SIM2) approach, which generates a robust optimal solution for task scheduling. We have used the concept of Semi-Interquartile, Min-Min and Max-Min to minimize the completion time of the tasks. Our experimental analysis shows better results than other conventional algorithms in terms of Makespan, Average Resource Utilization and Average Cycle Time.

1 Introduction

Grids are widely being used for high performance computing (HPC) applications because of high cost of massively parallel processors (MPP) and the wide availability of network workstations [3]. A grid is said to be heterogeneous if all nodes having different architectures and operating systems. A task scheduling algorithm may be local, global, static [2] or dynamic [4], [17]. We cannot guarantee the optimum solution but always find solution which is close to optimum [15]. Each algorithm has their merits and demerits. Scheduling distributed applications is a NP-complete problem [8], [18]. The better use of distributed system, efficient and effective algorithms is required [5].

Scheduling in grid is not limited to resource utilization but can be stretch out to the quality of service, the security, central control in administrative domains and real time scheduling [9], [14], [21]. Single system Image (SSI) is an illusion to the user. It is designed in such a way that appears as a single resource. When the user submits a job, it is the responsibility of grid resource broker to divide the job into various tasks and assigns to several resources [6]. Further, task can be divided into subtasks and it can be scheduled in parallel. Our end objective is to increase the overall throughput and resource utilization [7], [20]. Also, it is required to break resource idle time and balance the load [10], [12].

The rest of the paper is organized as follows: section 2 is presents the related works, section 3 presents the preliminaries. Section 4 proposes the SIM^2 algorithm and the performance metrics. Section 5 elaborates the illustration and experimental analysis. We conclude this study in Section 5.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 415–421.
springerlink.com

2 Related Works

Etminani et al. and Parsa et al. introduced a new scheduling algorithm which takes advantages of two traditional algorithms (min-min and max-min) [5], [16]. It selects one of the algorithms based on standard deviation of the expected completion time [5]. Senthilkumar et al. and Mehta et al. proposed a robust task scheduling for heterogeneous computing system. In this algorithm, each task arrival times and order of the task are not decided previously [13], [18]. Rasooli et al. introduced two new dispatching criteria for the first phase and three new dispatching criteria for the second phase in his rule based algorithm [17]. Sun et al. developed a priority-based task scheduling in which tasks are assigned to resources in priority order [19]. Parsa et al. chooses min-min strategy if available resource is odd. Otherwise, it chooses max-min [16]. Abdi et al. developed a job scheduling policy. In order to improve data access efficiencies, the replica manager is used. For replica selection or deletion, this strategy considered bandwidth between the regions [1].

3 Scheduling Algorithms

There are many task scheduling algorithms and/or heuristics exist in grid computing like min-min, max-min, minimum execution time (MET), minimum completion time (MCT), min-min max-min selective [5], RASA [16], LBMM [8] and many more.

3.1 Min-Min and Max-Min

Min-Min starts with small tasks; before the large one. First Min indicates minimum execution time i.e. small task and second mean indicates minimum completion time i.e. in a resource. Load imbalance is the main drawbacks of this algorithm.

Max-Min starts with large one first instead of small one. It is very similar to the min-min algorithm. Max indicates maximum execution time and min indicates minimum completion time. Load imbalance is reduced in this algorithm. Time complexity of min-min and max-min are O (tr) to assign a task to a resource [5]. In which t denotes number of tasks and r denotes number of resources.

3.2 MET and MCT

MET follows first in first out (FIFO) sequence. The task having less arrival time scheduled first in task queue (TQ). Load imbalance is the main drawbacks of this algorithm.

MCT also follows FIFO sequence. But, it finds the resource which takes less completion time. Again, Load balance is not achieved in this algorithm. Time complexity of MCT and MET are O(r) [5].

3.3 Min-Min Max-Min Selective

It is the combination of min-min and max-min algorithm. It calculates expected completion time (ECT) of all tasks. Then, it computes standard deviation (SD) using ECT. Based on the SD, it selects any one of the algorithm. Time complexity is $O(t^2r)$ [5].

3.4 RASA & LBMM

It is similar to min-min max-min selective algorithm. But, the main difference is instead of calculating SD, it checks number of available resources. Based on resource, it selects any one of the algorithm.

LBMM starts by executing min-min algorithm first. In second, it selects the resource which is heavy loaded and reassigns the tasks to light loaded resources. So, it is called as load balanced min-min algorithm.

4 Proposed Approach

4.1 Description

In our purposed approach, communication time is assumed to be negligible. Each task execution time is taken in seconds. Execution time can calculate easily if speed of a resource, bandwidth, instruction and data are already known previously [16]. Semi-interquartile range (SI) and Interquartile range (IQ) are calculated using a formula shown in equation 1 and 2 respectively.

$$\text{Interquartile range (IQ)} = Q3 - Q1 \tag{1}$$

$$\text{Semi-interquartile range (SI)} = (Q3 - Q1) / 2 \tag{2}$$

where Q1 = First Quartile, Q3 = Third Quartile

4.2 Pseudocode of SIM^2 Approach

1. Sort the meta-tasks in ascending order of their Execution Time (ET).
2. while there are meta-tasks in Task Queue (TQ)
3. for all meta-tasks T_i in TQ
4. for all resources R_j
5. $CT_{ij} = ET_{ij} + RT_j$ // CT = Completion Time, ET = Execution Time, RT = Ready Time
6. end for
7. end for
8. for all meta-taks T_i in TQ
9. Find minimum CT_{ij} and resource R_j that holds it.
10. end for
11. Calculate difference between two consecutive minimum CT_{ij} and Store in DQ // DQ = Difference Queue (DQ)
12. Calculate semi-interquartile range.
13. Find an element e in DQ $\geq$ semi-interquartile range and Store the location l.
14. If $l \geq (P / 2)$ or l = NULL
15. then assign meta-task T_p to resource R_k that holds minimum CT_{pk}.
16. else assign meta-task T_l to resource R_k that holds minimum CT_{lk}.
17. end if
18. Delete the meta-task, update TQ and DQ.
19. end while
20. Calculate makespan, average resource utilization and average idle cycle time.

Time complexity of SIM^2 is $O(t^2r)$.

4.3 Performance Metrics

To evaluate the performance of scheduling algorithms, we use following performance metrics:

4.3.1 Makespan

Makespan is a measure of the throughput of the heterogeneous computing system. It is calculated using a formula shown in equation 3.

$$\text{Makespan (M)} = \max\left(\sum_{i=1 \text{ to } t} CT_i\right) \tag{3}$$

4.3.2 Resource Utilization (RU)

It is the time that the resource is busy. It is calculated using a formula shown in equation 4 and average resource utilization (ARU) is calculated using a formula shown in equation 5.

$$RU(R_r) = \sum_{i=1 \text{ to } t} ETS(T_{ir}) \tag{4}$$

$$ARU = \left(\left(\sum_{i=1 \text{ to } r} RU(R_r)\right) / r\right) \times 100 \tag{5}$$

where R_r = Resource number r, ETS = Execution Time Spent,
t = Number of tasks, r = Number of resources

4.3.3 Idle Cycle

It is the time that the resource is idle. It is calculated using a formula shown in equation 6. Average idle cycle (AIC) is the average of all the resources idle cycle time.

$$\begin{aligned} IC_r &= ((M - RU(R_r))/M) \quad \text{if } M \neq RU(R_r) \\ &= 0 \qquad\qquad\qquad\quad \text{Otherwise} \end{aligned} \tag{6}$$

where M = Makespan, ICr = Idle Cycle of resource r

5 Experimental Results

5.1 Illustration

In Figure 1, first example calculates the IQ = 46.5. So, SI is IQ/2 = 23.25. As it is greater than P/2, max-min algorithm is applied for first iteration. In second example, as it is less than P/2, min-min algorithm is applied for first iteration.

Position (P):	1	2	3	4	5
CT_{ij}:	3	11	21	41	66
				⇧	
				$4 \geq P/2 = 2.5$	
IQ = 46.5, SI = 23.25, P = 5					
Position (P):	1	2	3	4	5
CT_{ij}:	1	55	75	95	120
	⇧				
	$1 < P/2 = 2.5$				
IQ = 79.5, SI = 39.75, P = 5					

Fig. 1. Illustration of SIM^2 approach

5.2 Experimental Analysis

To evaluate and compare SIM^2 algorithm with two existing approach min-min and max-min, we have considered four cases with two resources. The experimental results shows that proposed SIM^2 approach is more efficient than the other scheduling approach in terms of makespan, average resource utilization and idle cycle of resources. Performance metrics of all the approaches for three different cases are shown in figure 2, figure 3 and figure 4 respectively.

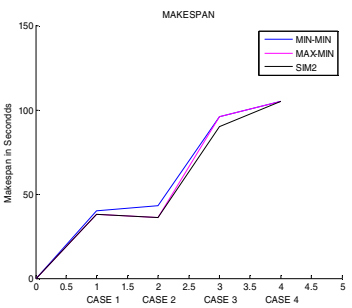


Fig. 2. Comparison of Makespan

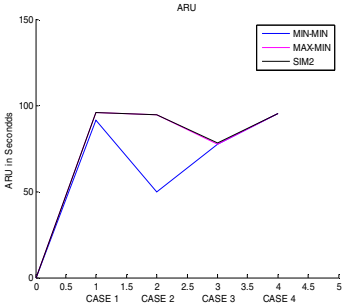


Fig. 3. Comparison of ARU

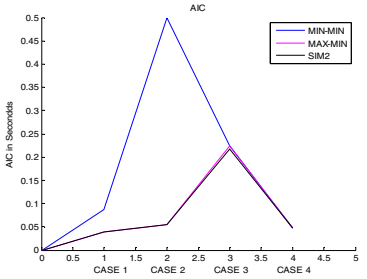


Fig. 4. Comparison of ARU

6 Conclusion

From the above experiments, SIM^2 approach shows better or equal results than other scheduling approach in heterogeneous distributed environment. The main goal of grid task scheduling is to increase the resource utilization (throughput), minimize the makespan and reduce the idle cycle time. By using SIM^2 approach, we are getting better performance in makespan, average resource utilization and average idle cycle. As there is no priority, each task is independent and it can be assigned to any resources at any point of time. In future, we can implement this algorithm in real grid environment. We can extend this approach by adding real time aspects like deadline; divides the task into number of pieces, user defined priority.

References

1. Abdi, S., Pedram, H., Mohamadi, S.: The impact of data replication on job scheduling performance in hierarchical data grid. International Journal on Applications of Graph Theory in Wireless ad Hoc Networks and Sensor Networks 2(3) (September 2010)
2. Braun, T., Siegel, H.J., Maciejewski, A.: Heterogeneous computing: goals, methods and open problems. Springer, USA (2001)
3. Buyya, R.: High performance cluster computing. Pearson Education (2008) ISBN 81-317-1693-7
4. Dong, F., Akl, S.G.: Scheduling algorithms for grid computing: state of the art and open problems. Technical Report No. 2006-504, School of Computing, Queen's University, Kingston, Ontario (January 2006)
5. Etminani, K., Naghibzadeh, M.: A min-min max-min selective algorithm for grid task scheduling. IEEE (2007)
6. Hemamalini, M.: Review on grid task scheduling in distributed heterogeneous environment. International Journal of Computer Applications 40(2), 24–30 (2012)
7. Kamalam, G.K., Bhaskaran, V.M.: New enhanced heuristic min-mean scheduling algorithm for scheduling meta-tasks on heterogeneous grid environment. European Journal of Scientific Research 70, 423–430 (2012)
8. Kokilavani, T., Amalarethinam, D.I.G.: Load balanced min-min algorithm for static meta-task scheduling in grid computing. International Journal of Computer Applications IJCA 20(2), 43–49 (2011)
9. Liang, Y., Jiliu, Z.: The improvement of a task scheduling algorithm in grid computing. In: First International Symposium on Data, Privacy and E-Commerce. IEEE (2007)
10. Liu, K., Chen, J., Jin, H., Yang, Y.: A min-min average algorithm for scheduling transaction-intensive grid workflows. In: 7th Australasian Symposium on Grid Computing and e-Research (2009)
11. Maheswaran, M., Ali, S., Siegel, H.J., Hensgen, D., Freund, R.F.: Dynamic matching and scheduling of a class of independent tasks onto heterogeneous computing systems. University of Manitoba, Purdue University, USA
12. Mansouri, N., Dastghaibyfard, G., Horri, A.: A novel job scheduling algorithm for improving data grid's performance. IEEE (2011)
13. Mehta, A., Smith, J., Siegel, H.J., Maciejewski, A., Jayaseelan, A., Ye, B.: Dynamic resource management heuristics for minimizing makespan with maintaining an acceptable level of robustness in an uncertain. In: 12th International Conference on Parallel and Distributed Systems (January 2006)
14. Munir, E.U., Li, J., Shi, S.: QoS sufferage heuristic for independent task scheduling in grid. Information Technology Journal 6(8), 1166–1170 (2007)
15. Navimipour, N.J., Khanli, L.M.: The lgr method for task scheduling in computational grid. In: International Conference on Advanced Computer Theory and Engineering, pp. 1062–1066. IEEE (2008)
16. Parsa, S., Maleki, R.E.: RASA: A new grid task scheduling algorithm. International Journal of Digital Content Technology and its Applications 3(4), 91–99 (2009)
17. Rasooli, A., Aghatabar, M.M., Khorsandi, S.: Introduction of novel rule based algorithms for scheduling in grid computing systems. In: Second Asia International Conference on Modelling & Simulation, pp. 138–143. IEEE (2008)
18. SenthilKumar, B., Chitra, P., Prakash, G.: Robust task scheduling on heterogeneous computing systems using segmented maxr-minct. International Journal of Recent Trends in Engineering 1(2), 63–65 (2009)

19. Sun, W., Zhu, Y., Su, Z., Jiao, D., Li, M.: A priority-based task scheduling algorithm in grid. In: 3rd International Symposium on Parallel Architectures, Algorithms and Programming, pp. 311–315. IEEE (2010)
20. Tang, M., Lee, B.S., Tang, X., Yeo, C.: The impact of data replication on job scheduling performance in the data Grid. Future Generation Computer System 22, 254–268
21. Xiao, Y.: Security in distributed, grid, mobile and pervasive computing. Auerbach Publications (2007) ISBN-10 0-8493-7921-0, ISBN-13 978-0-8493-7921-5

Simulation Based Performance Comparison of Reactive Routing Protocols in Mobile Ad-Hoc Network Using NS-2

G. Jose Moses[1], D. Sunil Kumar[2], P. Suresh Varma[1], and N. Supriya[1]

[1] Adikavi Nannaya University, Rajahmundry, India
{josemoses,supriyacse}@gmail.com,
vermaps@yahoo.com
[2] Government College, Rajahmundry, India
scientist.sun@gmail.com

Abstract. An ad hoc network is a collection of mobile nodes forming an instant network without fixed topology. In such a network, each node acts as both router and host simultaneously, and can move out or join in the network freely. To facilitate communication within the network a routing protocol is used to discover routes between nodes. The main aim of the routing protocol is to have an efficient route establishment between a pair of nodes, so that messages can be delivered in a timely manner. Routing in the MANETs is a challenging task which has led to development of many different routing protocols for MANETs. In this paper, an attempt has been made to compare two well known reactive routing protocols Ad-hoc On demand Distance Vector (AODV) and Dynamic Source Routing (DSR) by using three performance metrics Packet Delivery Ratio, Average End to End Delay and Routing Load. The comparison is done by varying number of sources and for each pause time. These simulations are carried out using the NS-2 network simulator.

Keywords: MANET, Routing Protocol, Reactive routing protocol, AODV, DSR.

1 Introduction

Wireless networks can be classified into two types: infrastructure network and infrastructure less (ad hoc) networks. Mobile Ad hoc network belongs to the category of infrastructure less network. Nodes in this network are autonomous in itself, that they are not dependent on any infrastructure. In this way, ad-hoc networks have a dynamic topology such that nodes are mobile in nature, so that they can easily join or leave the network at any time. for MANET, various routing protocols are available. Depending upon the nature of application, appropriate routing protocol is implemented. Proactive and reactive protocols are the two classes of MANET routing protocols and each constitute a set of protocols as follows. Proactive routing protocol: Table-driven (Ex: DSDV, FSR, WRP …) Reactive routing protocol: On-demand (Ex: AODV, DSR, TORA …).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 423–428.
springerlink.com

Reactive routing protocol

These protocols are also called on demand protocols since they don't maintain routing information or routing activity at the network nodes if there is no communication. If a node wants to send a packet to another node then this protocol searches for the route in an on-demand manner and establishes the connection in order to transmit and receive the packet. The route discovery usually occurs by flooding the route request packets throughout the network.

2 AODV

Ad hoc On Demand Vector routing protocol is a reactive routing protocol [1]. As the name suggests AODV constructs route from source to destination when ever needed. It allows mobile computers (nodes), to pass messages through their neighbors to nodes with which they cannot directly communicate. AODV discover the routes to all destination nodes by using its immediate neighbors. AODV makes sure these routes do not contain loops and tries to find the shortest route possible. AODV is capable to handle changes in routes and can create new routes if there is an error. Each node broadcasts a HELLO message at regular intervals to keep track of neighbors list. When one node needs to send a message to another node that is not its neighbor node it broadcasts a Route Request (RREQ) message. The RREQ message format is as below.

Source	destination	Broadcast Id	Hop count	Destination seq no	Source seq no	Lifespan

Sequence number serves as a unique Id to indicate roots freshness. As RREQ travels from node to node, it automatically sets up the reverse path from all these nodes back to the source. Each node that receives this packet records the address of the node from which it was received. This is called Reverse Path Setup.

3 DSR

Dynamic Source Routing [2] is a reactive routing protocol that uses source routing to send packets. It uses source routing which means that the source must know the complete hop sequence to the destination. Each node maintains a route cache, where all routes it knows are stored. The route discovery process is initiated only if the desired route cannot be found in the route cache. To limit the number of route requests propagated, a node processes the route request message only if it has not already received the message and its address is not present in the route record of the message. As mentioned before, DSR uses source routing, i.e. the source determines the complete sequence of hops that each packet should traverse. This requires that the sequence of hops is included in each packet's header. A negative consequence of this is the routing overhead every packet has to carry. However, the advantage is that intermediate nodes can learn routes from the source routes in the packets they receive. Since finding a route is generally a costly operation in terms of time, bandwidth and energy, this

is a strong argument for using source routing. Another advantage of source routing is that it avoids the need for up-to-date routing information in the intermediate nodes through which the packets are forwarded since all necessary routing information is included in the packets. Finally, it avoids routing loops easily because the complete route is determined by a single node instead of making the decision hop-by-hop. The protocol is composed of the two main mechanisms of "Route Discovery" and "Route Maintenance", which work together to allow nodes to discover and maintain routes to arbitrary destinations in the ad hoc network. The protocol allows multiple routes to any destination and allows each sender to select and control the routes used in routing its packets, for example, for use in load balancing or for increased robustness.

4 Performance Metrics

Packet Delivery Ratio = No. of packets received successfully/No. of packets sent. Average Routing Load = No. of routing controlled packets/Total simulation time. Average End to End Delay = “Sum (for each i equal to packet number, (packet i received time- packet i sent time))”

Awk Script to Obtain the above Performance Metrics

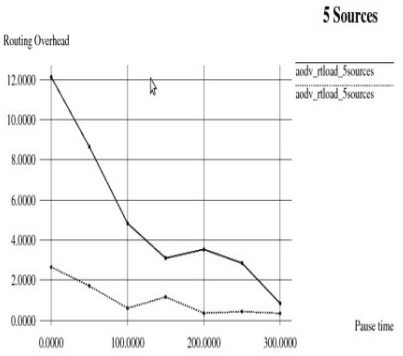
```
BEGIN {
seqno = -1; droppedPackets = 0;
receivedPackets = 0; count = 0;
sentpackets = 0; ctrlpac = 0; pdf = 0;
recvdSize = 0; startTime = 400; stopTime = 0;
} {         event = $1; time = $2; node_id = $3;pkt_size = $8;level = $4;
            if (level == "AGT" && event == "s" && pkt_size >= 512) {
            if (time < startTime) {         startTime = time;     }         }
    if (level == "AGT" && event == "r" && pkt_size >= 512) {
            if (time > stopTime) { stopTime = time;    }
            hdr_size = pkt_size % 512;pkt_size -= hdr_size;recvdSize += pkt_size;}
    if($4 == "AGT" && $1 == "s" && seqno < $6) { seqno = $6;sentpackets++; }
    else if(($4 == "AGT") && ($1 == "r")) { receivedPackets++; }
    else if ($1 == "d" && ($7 == "tcp" || $7 == "cbr") && $8 > 512){ droppedPackets++; }
    else if($4 == "RTR" && ($1 == "s" || $1 == "f") &&  ($7 == "DSR" || $7 == "AODV" || $7 ==
"message" )) {   ctrlpac++; } if($4 == "AGT" && $1 == "s") {     start_time[$6] = $2; }
    else if(($7 == "tcp"|| $7 == "cbr") && ($1 == "r")) {          end_time[$6] = $2;}
    else if($1 == "d" && ($7 == "tcp" || $7 == "cbr")) {          end_time[$6] = -1; }
    }
    END {for(i=0; i<=sentpackets; i++)      {
            if(end_time[i] > 0) {delay[i] = end_time[i] - start_time[i];count++;}
            else        {delay[i] = -1;}}
    for(i=0; i<=seqno; i++) {
            if(delay[i] > 0) {      n_to_n_delay = n_to_n_delay + delay[i]; }}
    n_to_n_delay = n_to_n_delay/count;
    pdf = receivedPackets/(sentpackets)*100;
    print "Packet Delivery Ratio  = " pdf   #"%";
    print "Average End-to-End Delay    = " n_to_n_delay*1000 "ms" ;
    print "Routing Overhead = "ctrlpac/300;
    }
```

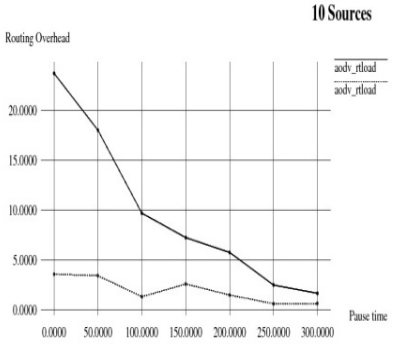
Simulation Parameters

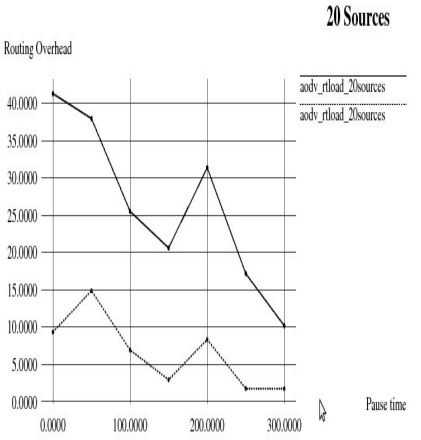
Parameter	Value
Transmission range	250 m
Simulation Time	300 s
Topology size	500m X 500m
No. of Mobile nodes	50
No. of Sources	5, 10, 20, 30
Traffic Type	CBR (Constant Bit Rate)
Packet Rate	5 packets/sec
Packet size	512 bytes
Pause Time	300

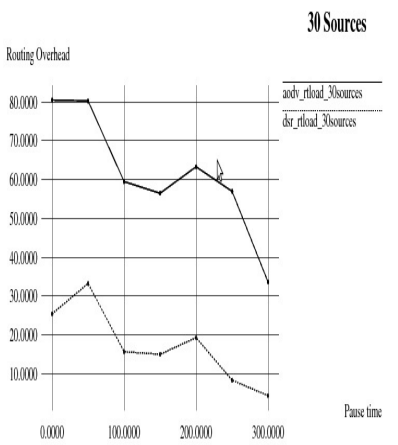
Results

RoutingLoad

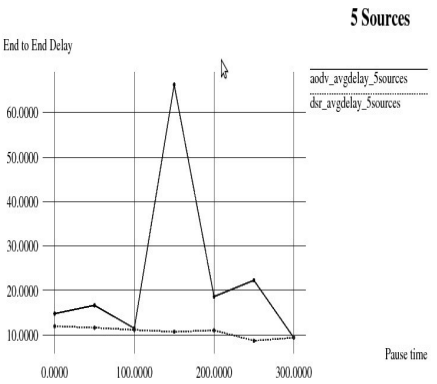


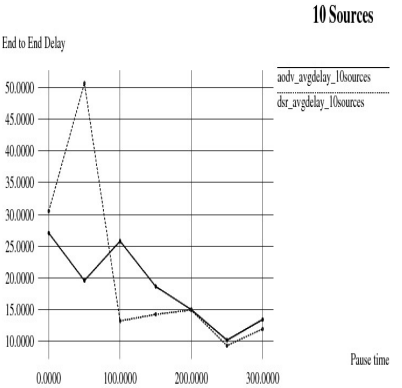


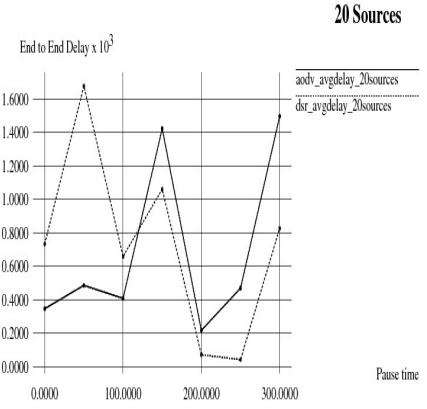


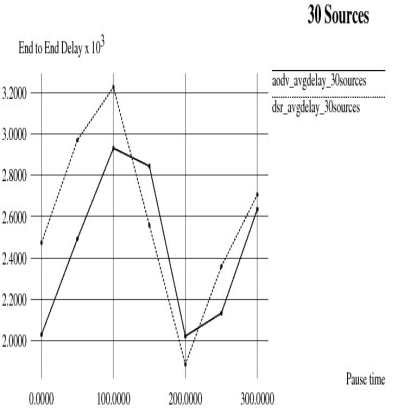


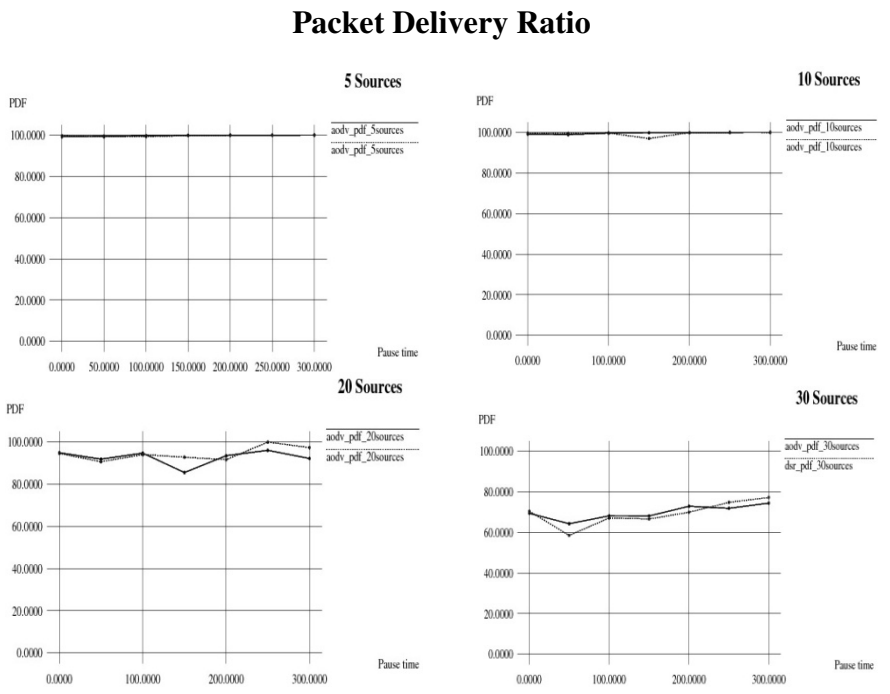
Average End to End Delay











The observations clears that both AODV and DSR protocols acheives maximum Packet Delivery Ratio and almost both are close to each other. The Protocol AODV has more Routing overhead than DSR because AODV periodically sends RREQ, RREP packets for each intermediate node. The average delay for DSR is better at least sources. As the number of sources increase AODV is better for this metric.

5 Conclusion

Comparison of these both reactive routing protocols based on significant parameters like Average End to End Delay, Routing Load and Packet delivery ratio are done. The protocol AODV performs better for high mobility networks where as DSR is better for low mobility networks.

References

1. Perkins, C.E., Royer, E.M.: Ad-Hoc on-Demand Distance Vector Routing. In: Proc. Workshop Mobile Computing Systems and Applications (WMCSA 1999), pp. 90–100 (February 1999)
2. Johnson, D.B., Maltz, D.A.: Dynamic source routing in ad hoc wireless networks. In: Mobile Computing, pp. 153–181. Kluwer Academic Publishers (1996)
3. `http://www.isi.edu/nsnam/ns/tutorial` Marc Greis tutorial on ns2
4. Transier, M.: Ns2 tutorial running simulations
5. Awk - A Tutorial and Introduction - by Bruce Barnett

Fault Tolerance for Large Scale Storage Systems

Pradeep K.R. and George Philip C.

Information Science
MSRIT, Bangalore
{pradeep.aarya,georgephilipc}@gmail.com

Abstract. With exponential increase in the digital information the amount of data that enterprises should manage is very tedious and storage systems are frequently scaled to meet the industry demand and service level agreements. As the storage systems are scaled, validating the storage systems installation becomes very desirable as this help in notifying the reliability, availability, performance issues, analyzing the downtime, taking backup, snapshots, mirroring the image, making decisions to avoid single point of failure …etc. This task of validating the storage systems installation becomes very tedious as the number of wiring configurations increases exponentially when they are scaled and process of validation needs automation as it can lead to misconfigurations even at maximum attention if done manually. In this paper we present a solution to ensuring a high degree of availability and reliability in large scale storage systems.

Keywords: HA(High Availability), SPOF(Single Point of Failure), ASUP(Auto Support).

1 Introduction

Storage systems are very crucial in the current scientific or business world they have become inherent part of any organization or any business unit. Due to the scientific advancements in various domains the amount of information that is generated day by day is exponentially increasing and people are finding it very easy and cost effective to store the information in the digital format[3]. The digital information generated by the people on daily basis is very huge, maintaining such a large amount of data is very difficult. So these organizations need to scale their storage systems and also maintain them every day in order to meet the demand and service level agreements, if not the customers might migrate to the competitor thus loosing huge amount of business.

Storing huge amount of data is very difficult and as the information increases day by day organizations need to scale their storage systems. Since it is not very efficient or recommendable to store all information in single storage[4], it is usually stored in large no of storage disks that are stacked in the shelves as recommended by storage experts in order to operate them without any problems. Many such storage disks are controlled by one controller and controllers are in-turn connected to many other storage controllers and also to the network[5][8]. For an enterprise the average no of disks can be around thousands and the controllers can be around hundreds. Assuming this condition the amount of wiring that could be done to connect all these controllers

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 429–434.
springerlink.com

and to the network becomes very complex and it is very difficult for any human being to understand these connections when scaling the storage systems. So organization should employ large no of staff in order to maintain these systems. This not only increases the operational cost but also the staff training cost and business loss if any controllers are down in the system.

In order to solve the above said storage problems, there is huge demand for a storage system installation validation tool in the market, which checks the wiring configuration in the storage systems and also notifies the problems that could occur if certain configurations are not properly set .Such a tool should not only notify the problems, it should also notify how reliable the system is and also check if it is giving the desired performance or not.

2 Need for Installation Validation

High-end enterprise storage systems currently deployed in production environments have several symmetrical multiprocessors, multiple internal fabrics, and large cache memories [9]. In Storage systems, with the quantity of information exploding, how to offer stable, continuous and secure data store and retrieve services is a basic but critical problem. With large explosion of information and demand to achieve very high availability and reliability of data, large enterprises are purchasing huge amount of storage systems in order to scale it the market demand [1][2] .In order to achieve very high availability and reliability these scaling storage systems connected to the network should be validated very frequently in order to avoid single point of failure .The figure below gives a clear picture of, to what extent does the storage system can scale and also the connection complexity that increases with increase in the no of storage controllers .Hence validation of connections can be to check whether the HA pairs are available, or does HA pair have equal no of disks and shelves as that of its partner node etc..

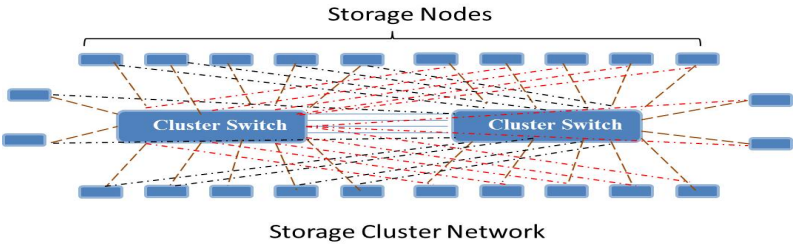


Fig. 1.

This Installation tool should be run once the storage system has been setup for the first time to check the installation problems so that user can be very confident and it should also be able to run by customer whenever he wants to check the reliability of the storage system. the system should be able check the system configuration faults and inform the user if there are any problems immediately ,so that immediate action can be taken against those problems reported by the tool ,so that there will be no busi-

ness loss or data failure in the network. This assures the user or customer some confidence that the storage systems have the correct configuration. Thus assuring high availability, performance, etc..

3 Architecture of Installation Validation

Architecture of the Installation Validation consists of following four important steps

1. Data collection
2. Parsing
3. Analyzing
4. Results Display

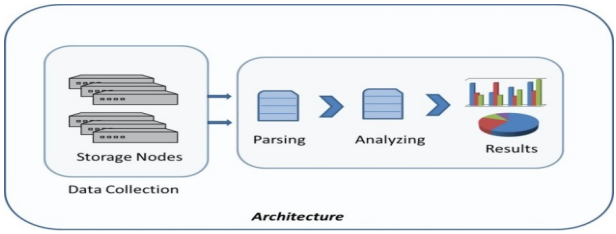


Fig. 2.

Data Collection: This is the primary step that is most essential in the process of Installation validation, this process involves collection of data from all the storage devises by running some commands, this data is collected and stored in the some generic data format so that it is independent of any language or technology.

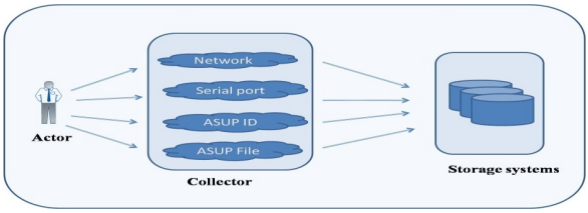


Fig. 3.

Data collection can be done in 4 different ways,

1. Through Network.
2. Serial port (Directly connecting with storage devices).
3. Through Auto-support Identifier.
4. Through Auto-support file.

Parsing: This step involves the process of parsing the data that was collected from storage devices and using local data structures to store the data so that this data can be used very efficiently while analyzing storage output. This process involves writing large number of regular expressions that parse necessary data that is needed for validation.

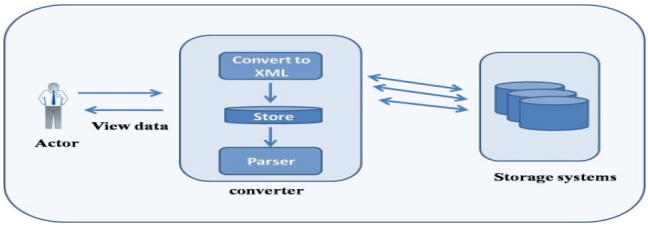


Fig. 4.

Analyzing: This is the most important step in the process of installation validation as this involves analyzing structured data that was collected from and parsed. Analyzing is done by writing rules, where each rule analyzes some particular fault in the installation.

E.g. 1. Rule that validates weather inter-switch links are up or down.
2. Rule that validates weather wiring is done on proper ports.

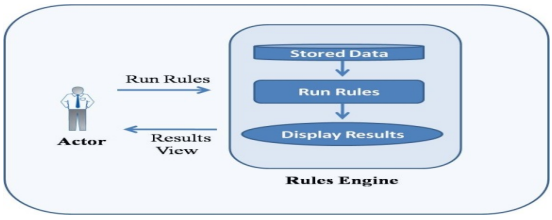


Fig. 5.

Result Display: This step involves portraying the results of analysis in an intuitive way so that they can be easily understood by the customer with minimal difficulty. Results can be diagrammatic representation of stacks, disks, shelves, .etc.

Sequence diagram below clearly describes the workflow of the project from the point the user stars interacting with the UI to the final step of producing the results.

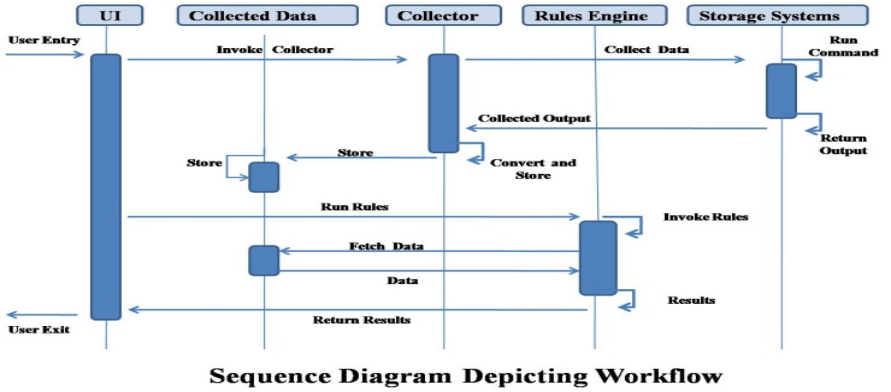


Fig. 6.

4 Statistics and Results

The toll was used to test a storage system setup which consisted of two storage controllers(nodes) and each node consists of 2 shelves each having 24 disks of capacity 1TB each. Many rules were programmed to analyze the storage wiring configuration. This tool was run when the installation was done for the first time and the tool helped us in finding many wiring misconfigurations listed below

- Disks were placed in a disproportionate number.
- Dual power supply was not configured.
- Disks were not connected with fiber optic cable to get maximum bandwidth.
- Inter node links were misconfigured.
- High Availability (HA) pairs were not configured properly to avoid single point of failure (SPOF).

The figure shows the snapshot of the results view of the tool when it was run for the first time(after fresh Installation of storage nodes).

Impact Level	Rule Name	Rule Description	Status	Details
●	DualPowerCheck	Checks the dual power supply for all nodes	Fail	Node 2 Power supply is off ,immediate action required
■	InterNodeLinks	Check the Inter Node links	Warning	Inter Node link are down on port 0a,0c.
★	CabellingCheck	Checks the cabling configurations with that of ports	Warning	Wrong cables are connected for ports 0b, 0c,0f

Fig. 7.

5 Conclusion

This paper presents an effective method of determining the reliability in storage network. Using this process we could determine the total number of misconfigurations in node-pair network for every run of a validation tool thus gaining some confidence of how reliable and available is the system. Using this process we can even determine faults whenever the system is scaled and plan on how to tolerate such faults, hence improving the reliability. System reliability as we have seen can be refined on periodic validation of the storage network installation hence obtain precise value with repetition of the validation process.

References

1. Bobbio, A., Ferraris, C., Terrugia, R.: New Challenges in Network Reliability Analysis. In: CNIP 2006, Rome, pp. 554–564 (2006)
2. Pinheiro, E., Weber, W.-D., Barroso, L.A.: Failure Trends in a Large Disk Drive Population. In: Proc. Fifth Conf. File and Storage Technology (February 2007)
3. Nastase, M., Dobre, C., Pop, F., Cristea, V.: Fault Tolerance using a Front-End Service for Large Scale Distributed Systems. IEEE (2009)
4. Zhou, S., Chen, L.: Fault Tolerant Maximal Local Connectivity of Alternating Group Networks. IEEE (2010)
5. Yang, X., Zhu, S.: Fragment Maintenance in Distributed Storage Systems. IEEE (2010)
6. Rao, K.K., Hafner, J.L., Golding, R.A.: Reliability for Networked Storage Nodes. IEEE (2011)
7. Weber, W.-D., Barroso, L.A.: Failure Trends in a Large Disk Drive Population. In: Proc. Fifth Conf. File and Storage Technology. Proc. ACM/IEEE Conf. Supercomputing (February 2007)
8. Coit, D.: System Reliability Confidence Intervals for Complex Systems with Estimated Component Reliability. IEEE Trans. on Reliability 46(4), 487–493 (1997)
9. Song, P., Sun, J.-L., et al.: Survey of Fault Tolerant Technology Based on Quorum Systems. Journal of Computer Research and Development, 513–522 (April 2004)

Real Time Electro-Oculogram Driven Rehabilitation Aid

Anwesha Banerjee[1], Pratyusha Das[2], Shounak Datta[1], Amit Konar[1], R. Janarthanan[3], and D.N. Tibarewala[1]

[1] Jadavpur University, Kolkata, West Bengal India
[2] Institute of Engineering & Management, Kolkata, West Bengal, India
[3] Jaya College of Engineering, Chennai, Tamilnadu, India
anwesha.banerjee@ymail.com,
{pratyushadas01,srmjana_73}@yahoo.com,
{shounak.jaduniv,biomed.ju}@gmail.com,
konaramit@yahoo.co.in

Abstract. Human computer interfacing technology based rehabilitation aids have shown a new horizon towards intelligent systems to improve the quality of life of physically challenged people. Research is going on to utilize biosignals to interface the movement based signals with machines. Electro-oculogram is the signal to detect eye ball movements and can be used to control mobility aids. Electro-oculogram is the potential difference around the eyes due to movement of the eye balls in different directions. In this study an acquisition system for electro-oculogram is designed to collect the desired signal with low noise and then signal processing is done for control application. The contribution of this paper lies in the development of two new strategies to use electrooculographic signal based control of motors in real time.

1 Introduction

An efficient alternative way to communicate without speech and hand movements is important to increase the quality of life for patients suffering from neural diseases or other illnesses or congenital problem or age which destroys proper limb and facial muscular responses. Hence, the area of study related to the Human Computer Interface (HCI) is very important to help such severely paralyzed patients. According to WHO, there are almost 650 million people or more who are solely physically challenged [1]. Given the growth in life expectancy in the world (in the countries of the Organization for Economic Cooperation and Development (OECD) it is expected that a large part of its population will experience functional problems. On the other hand, there are some diseases like Amyotrophic Lateral Sclerosis (ALS), brain or spinal cord injury, cerebral palsy, muscular dystrophies, Guillain-Barre syndrome, some rare cases of Parkinson disease, etc. which leads to a condition called locked-in state (LIS) when the patient's peripheral and central motor system gets completely destroyed but sensory or cognitive functions remain active [1]. These diseases impair the neural pathways that control muscles or impair the muscles themselves.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 435–440.
springerlink.com

The HCI systems translate biopotentials (e.g. EEG, EMG, ECG, EOG etc.) into electrical signals that control external devices. [1]. To develop eye movement controlled HCI, gaze detection can be done using many techniques such as Infrared Video System (IRVS), Infrared oculography (IROG), Search Coil (SC), Optical-type Eye Tracking System, Purkinje dual-Purkinje-image (DPI) and Electro-oculography (EOG) [2]. EOG is the simplest method among all of them. Electroencephalogram (EEG) or Electromyogram (EMG) has a complex signal to be processed and acquiring Electrocardiogram (ECG) 24×7 is always not possible. Sometimes, people are largely paralyzed by massive brainstem lesions cannot move their muscles but are able to control their eye movement. Moreover, EOG system is fairly easy to construct and easy to work in real time which makes EOG a better option over other biopotentials [3].

Many approaches have been experimentally done to control wheelchairs using EOG [4, 5, 6, 7, 8, 9, 10, 12]. In our study, a data acquisition system for EOG is designed. The acquired EOG signal has been applied to control the movement of motors according to the direction of eye ball movement. It is helpful to produce an EOG controlled HCI.

2 Electro-Oculogram

Electro-Oculogram (EOG) signal is actually the corneo-retinal potential resulting from a dipole (eye ball), generated between the cornea and the retina. The amplitude of the signal remains in the range of few micro volts. This potential is produced due to the movement of the eye ball and can be acquired noninvasively by placing electrodes in the surrounding region of eye.

The signal shows a particular pulse shape for eye ball movement in either direction. EOG has pulse duration of approximately 200ms on average and the signal magnitude changes from 5-20 micro volts for a degree of eye ball movement [4]. The amplitude of the EOG signal changes depending on the angle through which the eyeball was moved. When eye ball is moved one side the voltage remains positive (or negative) and returns to zero when looking straight. The pulse produced by leftward movement is nearly the same as produced by rightward movement in both amplitude and pulse duration. The signal potential remains the same even with the eyes closed. One problem of EOG signal is the head or body movement alters the DC level of the signal.

The main application of EOG signal is the detection and assessment of the degenerative muscular disorders like laziness of the eye in tracking moving objects. Analysis of EOG helps to track the progress of many ophthalmological diseases such as retinitis pigmentosa and neural diseases (e.g. Parkinson's, Alzheimer's' etc.) [8]. EOG is also used for drowsiness detection [9] and cognitive process modeling [10]. An important utilization of the EOG signal is in producing eye movement controlled human computer interfaces.

3 Acquisition of EOG

EOG signal is available in the frequency range of 0.1 to 20 Hz and the amplitude lies between 100-3500 micro volts. A voltage gain of minimum 2000 is needed to process the signal further [11]. EOG signal is acquired using Ag-Agcl disposable electrodes. The placement of the electrodes can be seen in the figure 1.

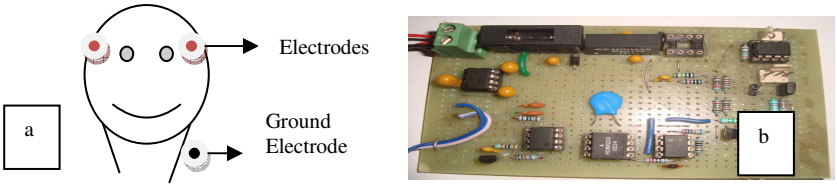


Fig. 1. Placement of electrodes for the acquisition of the signal(a) and designed data acquisition circuit for EOG(b)

The acquired signal from the electrodes is fed to instrumentation amplifier (implemented using IC AD620) having high input impedance and CMRR followed by a second order low pass filter with a cut off of 20HZ and a high pass filter of 0.1Hz cut off to eliminate unwanted data. For filter designing IC OP07s are used. Gain is applied in various stages. Amplifier has a gain of 200 and 10 gain is provided by the filters. Thus an overall gain of 2000 is reached. For biopotential signal acquisition isolation is an important factor to be considered for patient's safety as well as for instrument's safety. Power isolation is provided by the use of a dual output hybrid DC-DC converter (MAU 108) and signal isolation is obtained by optically coupling the amplifier output signal with the next stage. HCNR 200 is used to achieve this. The EOG signal is observed in LabVIEW 2009 environment.

4 Real Time Control Strategy

The ultimate aim of this undertaking is to facilitate eye movement based control of rehabilitation aid. For realization of the same, a microcontroller based strategy is implemented.

At first we are applying the amplified sensor output (analog EOG signal) to the op-amp level shifter circuit to make the negative input a positive one. Here our input can vary from -2V to +2V.So we are shifting the level by 2volts.

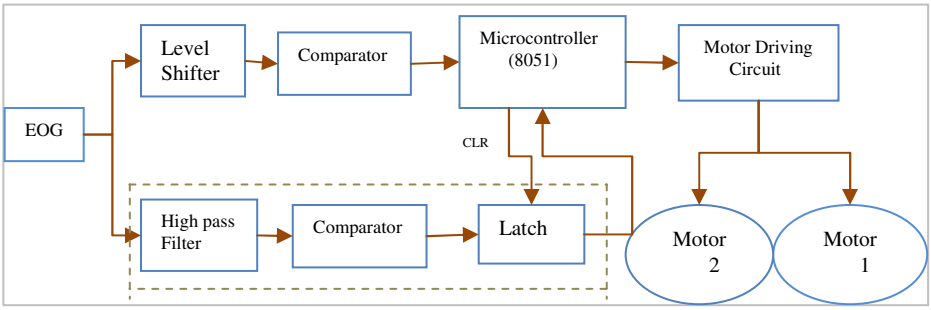


Fig. 2. Block diagram representation of the 8051 based control method

Then we are applying the level shifted output to the comparator as input. We apply the reference voltage of the comparator as per our input. From that we get

- ✓ 00 as output in case of eye movement in left direction
- ✓ 11 as output in case of eye movement in right direction
- ✓ 01 if there is no movement or eye is in middle position

Table 1. Movement Criteria Considered for the Motors According to Eye Movement

Eye Movement	Corresponding Motor Movement
Looking Straight	No Movement
One time left then straight	Left turn
One time right then straight	Right turn
Right-straight-right-straight	Forward
Left-straight-left-straight	Stop

After that we apply the comparator output to the microcontroller as input. Here we are using the port 1 as input port of microcontroller ATMEL89C51 and port2 as output port. As we are applying two digit input, so here we are using specifically p1.0 and p1.1as input pin. After that we are applying the microcontroller output to the motor driver as input so that we can drive the motor at 12V. For level shifting IC 741, comparator IC LM324 and motor driving IC L293D are used.

The developed system has some limitations and there is a chance to improve its functionality. The system is always on and it would rotate the motors even for normal eye movement of the user which can create a major problem. To solve this problem turning ON or OFF the motor rotation system should be controlled by the user. In this paper we propose a scheme to utilize the EOG signal for eye blink to turn ON-OFF the system. The EOG signal collected for eye blink can be quantized by feeding it to a high pass filter and comparator and then to a latch. The input of the latch is given high and the clock is the EOG signal for blink. Whenever it gets a blink it will send high input to the micro controller and simultaneously the microcontroller resets the output of the latch. Microcontroller counts the blink signal and when it gets two of them it

turns on the motor controlling system and then only according to the predefined commands the system would work. The proposed system is shown in the figure 3(the dotted portion).

5 Observation

EOG signal is acquired using disposable electrodes from the surrounding region of eye. Then processing the signal motors were moved according to the eye ball movement. The observed pattern of EOG signal is shown below (figure 3).

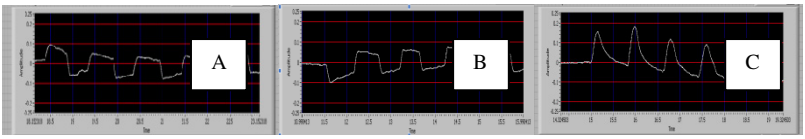


Fig. 3. EOG signal patterns for right(A), left(B) eye ball movement and eye blink(C) as observed through LabVIEW

Using the microcontroller system, when the user moves his/her eyes according to the previously mentioned commands the motors move forward, right and left. The motors are mounted in a toy car and connected with the output port of the microcontroller. Movement of the motors is shown in figure 4.

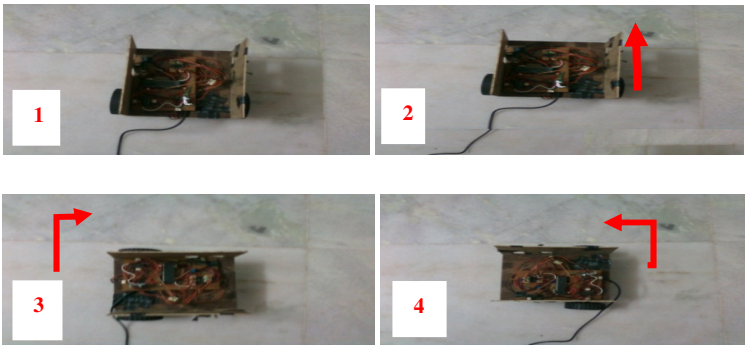


Fig. 4. Movement of the motors using microcontroller based system (1-initial position, 2- forward movement, 3- right turn, 4- left turn)

6 Conclusion and Future Work

In this study, Electrooculogram (EOG) signal is used to control motors. 8051 microcontroller has been used to implement eye movement based motor control. The method performs well while having some lack of precision. The microcontroller based system is good enough to be used in wheelchairs. In this way rehabilitation aids can

be controlled using subject's own EOG which will help severely paralyzed patients. On the other hand, proposed system is comparatively cheap and easy to design.

For improvement of the system a scheme is proposed which is to be done in further work. Other method also can be adapted to improve the accuracy of the system. One important factor to be considered for real time implementation is that the user should be trained properly.

References

1. Cincotti, F., Mattia, D., Aloise, F., et al.: Non-invasive braincomputer interface system: towards its application as assistive technology. Brain Research Bulletin 75(6), 796–803 (2008)
2. Deng, L., Hsu, C., Lin, T., Tuan, J., Chang, S.: EOG-based Human–Computer Interface system development. Expert Systems with Applications, 3337–3343 (2010)
3. Barea, R., Boquete, L., Mazo, M., Lopez, E.: System for Assisted Mobility Using Eye Movements Based on Electrooculography. IEEE Transaction on Neual Systems and Rehabilitation Engineering 10(4), 209–218 (2002)
4. Barea, R., Boquete, L., Mazo, M., López, E., Bergasa, L.M.: Wheelchair guidance strategies using EOG. Journal of Intelligent and Robotic Systems 34, 279–299 (2002)
5. Barea, R., Boquete, L., Mazo, M., López, E., Bergasa, L.M.: E.O.G. guidance of a wheelchair using spiking neural networks. In: European Symposium on Artificial Neural Networks, Belgium, pp. 233–238 (2000)
6. Rajan, A., Shivakeshavan, R.G., Vijay Ramnath, J.: Electrooculogram based Instrumentation and Control System (IC System) and its applications for Severely Paralysed Patients. In: Intl. Conf. on Biomedical and Pharmaceutical Engineering (2006)
7. Lv, Z., Wu, X., Li, M.: Implementation of the EOG-based Human Computer Interface System. In: 2nd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2008, pp. 2188–2191 (2008)
8. Ding, Q., Tong, K., Li, G.: Development of an EOG (Electro- Oculography) Based Human-Computer Interface. In: Proceedings of the 2005 IEEE EMBS 27th Annual Conference, Shanghai, China, September 1-4, pp. 6829–6831 (2005)
9. Liu, C., Chaovalitwongs, W., Pardalos, P., Seref, O., Xanthopoulos, P., Sackellares, J., Skidmore, F.: Quantitative analysis on electrooculography (EOG) for neurodegenerative disease
10. Roy Choudhury, S., Venkataramanan, S., Nemade, H.B., Sahambi, J.S.: Design and Development of a Novel EOG Biopotential Amplifier. IJBEM 7(1) (2005)
11. Fkirin, M.A., Badawy, S., El-Sherbeny, A.S.: Driving a DC Motor by Numerically Manipulated Eye Signal Captured by EOG. The Online Journal on Electronics and Electrical Engineering (OJEEE) (2009)
12. Usakli, A.B., Gurkan, S., Aloise, F., Vecchiato, G., Babiloni, F.: On the Use of Electro Occulo Gram for Efficient Human Computer Interfaces. In: Corporation Computational Intelligence and Neuroscience. Hindawi Publishing (2010)

A Quantitative Approach Using Goal-Oriented Requirements Engineering Methodology and Analytic Hierarchy Process in Selecting the Best Alternative

Vinay S.[1], Shridhar Aithal[2], and Sudhakara G.[3]

[1] NMAMIT,
Nitte, India
vinaymanyan@gmail.com
[2] TAPMI,
Manipal, India
saithal@tapmi.edu.in
[3] MIT,
Manipal, India
sudhakaraadiga@yahoo.co.in

Abstract. Decision making in Software Engineering plays an important role at different stages of Software development life cycle. In this paper we consider the case study of selecting one among the three Content Management Systems (CMS) for a university website. We use our Goal-Oriented Requirements Engineering (GORE) method to identify the soft goals which play a vital role in deciding which CMS is chosen. Analytic Hierarchy Process (AHP) is then used to prioritize the soft goals. The output of GORE and AHP are combined in order to produce a metric which decides the best alternative among the candidates.

Keywords: GoalOriented Requirements Engineering, Analytic Hierarchy Process, Soft goals.

1 Introduction

It is well acknowledged in Software Engineering that while functional requirements are important, eliciting and capturing the non-functional requirements (NFR) during the requirements engineering phase becomes even more important [1]. Goal-oriented requirements engineering (GORE) approaches make a good attempt to address the essential quality characteristics which are commonly known as non-functional requirements [2, 3]. NFRs play a major role in coming up with alternative system configurations for a given functionality.

Gunther Rahe in his paper [6] highlights the importance of Decision Support system (DSS) in Software Engineering. Decisions are the driving engines for all stages of software development and evolution. Decisions can be related to resources and

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 441–454.
springerlink.com

software technologies (methods, tools, and techniques). Decisions related to the 'How'? 'How good'? 'When'? 'Why'? and 'Where'? questions in the use of Software technologies is too often still based on simplistic rules of thumb. What is needed is a sound methodology and with links to validated models and experience. The importance of decision making techniques is also addressed in [7, 8 and 9]. The importance of stakeholders in decision making is done in [10]. Architecture decision making is closely linked to requirements engineering and the aspects related to this are addressed in [11] and [12]. A survey of various requirements prioritization techniques is undertaken in [13]. Decision-Making in Software Engineering is extremely challenging because of a dynamically changing environment, conflicting stakeholder objectives, constraints, coupled with a high degree of uncertainty and vagueness of the available information.

The Analytic Hierarchy Process (AHP) is based on the experience gained by its developer, T.L. Saaty [25], while directing research projects in the US Arms Control and Disarmament Agency. It was developed as a reaction to the finding that there is a miserable lack of common, easily understood and easy-to-implement methodology to enable the taking of complex decisions. Since then, the simplicity and power of the AHP has led to its widespread use across multiple domains in every part of the world. The AHP has found use in business, government, social studies, R&D, defence and other domains involving decisions in which choice, prioritization or forecasting is needed [24]. Combining our GORE method and AHP in decision making provides adequate rationale for the decision arrived at.

In this paper we consider the case study of selecting one among the three Content Management Systems (CMS) for a university website. We use our Goal-Oriented Requirements Engineering (GORE) method to identify the soft goals which play a vital role in deciding which CMS is chosen. Analytic Hierarchy Process (AHP) is then used to prioritize the soft goals. The output of GORE and AHP are combined in order to produce a metric which decides the best alternative among the three. In Section 2 and 3, we discuss our proposed approach and highlight how our GORE approach is used for identifying soft goals (non-functional requirements), their contribution links to each of the alternative. Section 4 discusses how AHP is used to prioritize the soft goals. Combining the output of GORE and AHP is discussed in Section 5. The effectiveness of the proposed method is discussed in Section 6.

2 Proposed Approach

The major steps involved in our proposed approach are shown in Table 1:

Table 1. Steps involved in our approach

Step	Description	Method	Output
1	Identify the alternatives	Apply our GORE method to identify alternatives to achieve a hard goal or functional requirement. The task here is to decide on the best alternative to be selected among three CMS for designing a website.	We consider the following CMS as candidate choices: i. Drupal 4.6.5 ii. e Z Publish 3 iii. MD Pro 1.0.76
2	Identify soft goals (non-functional requirements)	Apply our GORE method to identify soft goals which affect decision making	Table 3 and 4. Discussed in section 3
3	Identify contribution links	Apply our GORE method to identify contribution links between soft goals and each alternative	Table 3. Discussed in section 3
4	Calculate priority of soft goals	Use AHP to calculate priority of soft goals. If it is a multi-level hierarchy, we need to proceed from root to leaf level soft goals in calculating priorities of soft goals using AHP. Local weights are calculated among same level goals using AHP. Global weights are the product of all local weights proceeding from leaf to root.	Table 5. Discussed in section 4
5	Evaluation of each alternative	Use Quality Function Deployment to convert contribution link to numeric values. This value is multiplied with the global weights for each alternative found in the previous step. Convert the absolute value to relative value. This step is repeated for other alternatives	Table 6. Discussed in section 5
6	Calculation of effectiveness of each alternative	The effectiveness of an alternative depends on what extent the alternative fulfills the soft goals which defines quality. This is done by matrix multiplication of relative values of each alternative with the global weights of soft goals. Convert the absolute value to relative value.	Fig. 1. Discussed in section 5
7	Ranking of alternatives	The alternatives are ranked in terms of increasing order of their relative values.	Table 7. Discussed in section 5

3 Identifying Soft Goals and Contribution Links Using GORE Methodology

Some of the prominent GORE methods include: Keep All Objectives Satisfied (KAOS) method [13], Tropos [14], NFR framework [2, 15]. Our GORE method retains the generic features of other techniques in extracting hard and soft goals, decomposing the goals and refining them. In addition to these, we give more emphasis on identifying the contribution links. Many methods have taken this approach of representation. In our approach we propose the following links: ++, +, -- or -. The symbols and their meanings are shown in first two columns Table 2. The significance of the values mentioned in the third column is discussed in section 5. The method of identifying these contribution links which could be either quantitative or qualitative can be found in our earlier work [16].

Table 2. Meaning of Contribution Links

Symbol	Meaning	Value
++	A hard goal requirement is fully supported by the soft goal.	9
+	A hard goal requirement is partially supported by the soft goal.	6
-	A hard goal requirement is supported very minimally by the soft goal.	3
--	A hard goal requirement is not supported by the soft goal.	0

In our case study the hard goals are the three CMS which are the candidate choices for designing University website. The first task in our proposed approach is to identify all the soft goals which play a vital role in deciding the best alternative. The soft goals identified for this particular scenario are shown in the first three columns of table 3. Each of the soft goal could be further AND decomposed into sub-goals. For example, Security has 2 sub-goals and each of these sub-goals have been decomposed into 2 sub-goals each. Table 4 contains description about all the identified leaf level soft goals.

Table 3. Soft goals and Contribution links for each of the alternative

Goal (L1)	Sub goal (L2)	Sub goal (L3)	Drupal 4.6.5	e Z Publish 3	MD Pro 1.0.76
1.Security	1.1 Application Security	1.1.1 Human vs PC Verification	--	--	++
		1.1.2 Authentication extensibility	++	++	++
	1.2 Data Security	1.2.1 Support SSL Protocol	+	++	--
		1.2.2 SQL Security	++	++	++

Table 3. (*Continued*)

2.Management	2.1 Style Management	2.1.1 Web-based style	++	++	++
		2.1.2 Multiple templates per site	++	++	++
		2.1.3 Multiple menu types	+	+	+
		2.1.4 Dynamic Menus	--	--	--
		2.1.5 Multilingual contents	++	++	++
	2.2 Web-Statics	2.2.1 Visitor tracking	++	--	++
		2.2.2 Contents tracking	++	--	++
		2.2.3 Log-in History	++	--	--
3. Ease of Use	3.1 Drag and Drop		++	--	--
	3.2 Preview		++	++	++
	3.3 Spell Checker		++	--	--
	3.4 Undo (upto 10 levels)		++	++	--
	3.5 Image Resizing		++	++	--
	3.6 File type Conversion		++	--	--
4. Efficiency	4.1 Static Content export		--	++	--
	4.2 Page Caching		++	++	--
5. Help and Support	5.1 Manuals		++	++	++
	5.2 Online Help		++	--	--
	5.3 Videos and Demos		+	--	--
	5.4 Mailing list		+	--	--
	5.5 Public forum		+	--	--
6. Richness of built-in tools	6.1 Blog		++	++	--
	6.2 Chat		--	--	--
	6.3 Forum		++	++	--
	6.4 Image Gallery		--	++	--
	6.5 Graph and chart		--	--	--
	6.6 Search Engine		++	++	++

Table 4. Description of leaf level soft goals

Leaf Level Soft Goal	Description
1.1.1 Human vs PC Verification	The ability to determine whether the user is a human or a machine. The supportability to Captcha challenge-response protocol.
1.1.2 Authentication extensibility	The ability to integrate additional authentication mechanisms beyond the proprietary authentication protocols.
1.2.1 Support SSL Protocol	The ability of the system to work with a secure socket layer (SSL) certification on the web server
1.2.2 SQL Security	The supportability of encryption capabilities within the database
2.1.1 Web-based style	The ability to create, upload, and delete templates via a web browser
2.1.2 Multiple templates per site	The ability to choose a different template for each page
2.1.3 Multiple menu types	The ability to choose from different menu types
2.1.4 Dynamic Menus	The ability to create menus that are dynamically updated based on the site-map
2.1.5 Multilingual contents	The ability to create sites with multilingual contents
2.2.1 Visitor tracking	The ability to report the number of visitors per time period
2.2.2 Contents tracking	The ability to report on the number of downloads per time Period
2.2.3 Log-in History	The ability to keep track of who logged in, when, and what is his/her IP address.
3.1 Drag and Drop	The ability to position contents in a drag-and-drop fashion
3.2 Preview	The ability to preview contents online before publishing
3.3 Spell Checker	The ability to check spelling by a built-in spell checker
3.4 Undo (upto 10 levels)	The ability to undo performed operations
3.5 Image Resizing	The ability to resize images within articles without affecting the original stored image
3.6 File type Conversion	The ability to convert an image from one format to another
4.1 Static Content export	The ability of the system to export its contents as static HTML so it may be served by static HTML servers
4.2 Page Caching	The ability to cache pages so as to save the time needed for creating it when it is requested again.
5.1 Manuals	The availability and quality of additional books and manuals to explain the system installation process
5.2 Online Help	The availability and quality of the online installation help
5.3 Videos and Demos	The availability and quality of videos explaining the system installation process

Table 4. (*Continued*)

5.4 Mailing list	The availability of a mailing list service that includes latest news, updates, etc to the system
5.5 Public forum	The availability and quality of a public forum or discussion board for the system
6.1 Blog	The availability of a blog facility
6.2 Chat	The availability of an real-time online chat facility
6.3 Forum	The availability of a message board creation and management Utility
6.4 Image Gallery	The ability to create a gallery-page with thumbnails of images stored in the database
6.5 Graph and chart	The ability of the system to create graphs and charts based on some data sets
6.6 Search Engine	The availability of an integrated search engine for searching and indexing contents. The users can then use this engine to search the contents

4 Prioritizing Soft Goals Using AHP

We then apply step 4 of our approach which involves prioritizing the soft goals. We make use of AHPs pair wise comparison technique for this purpose. In case of a multi-level hierarchy, this is done by applying AHP starting from root level goals to leaf level goals. Local weights are calculated among same level goals using AHP. Global weights are the product of all local weights proceeding from leaf to root. The result of this step is shown in table 5.

Table 5. Calculation of Local and Global weights

Goal (L1)	**Local Weight**	**Sub goal (L2)**	**Local Weight**	**Sub goal (L3)**	**Local Weight**	**Global Weight**
1.Security	.023	1.1 Application Security	.167	1.1.1 Human vs PC Verification	0.875	0.0032
				1.1.2 Authentication extensibility	0.125	0.0004
		1.2 Data Security	.833	1.2.1 Support SSL Protocol	0.833	0.0159
				1.2.2 SQL Security	0.167	0.0031

Table 5. (*Continued*)

2.Management	.032	2.1 Style Management	0.833	2.1.1 Web-based style	0.512	0.0136
				2.1.2 Multiple templates per site	0.063	0.0166
				2.1.3 Multiple menu types	0.261	0.0695
				2.1.4 Dynamic Menus	0.129	0.0343
				2.1.5 Multilingual contents	0.033	0.0087
		2.2 Web-Statics	0.167	2.2.1 Visitor tracking	0.657	0.0034
				2.2.2 Contents tracking	0.146	0.0007
				2.2.3 Log-in History	0.196	0.0010
3. Ease of Use	.519	3.1 Drag and Drop	.228			0.1183
		3.2 Preview	.228			0.1183
		3.3 Spell Checker	.228			0.1183
		3.4 Undo (upto 10 levels)	.228			0.1183
		3.5 Image Resizing	.044			0.0228
		3.6 File type Conversion	.044			0.0228
4. Efficiency	.115	4.1 Static Content export	0.25			0.0287
		4.2 Page Caching	0.75			0.0862
5. Help and Support	.207	5.1 Manuals	0.506			0.1047
		5.2 Online Help	0.130			0.0269
		5.3 Videos and Demos	0.273			0.0565
		5.4 Mailing list	0.031			0.0064
		5.5 Public forum	0.060			0.0124
6. Richness of built-in tools	.101	6.1 Blog	0.348			0.0351
		6.2 Chat	0.072			0.0072
		6.3 Forum	0.045			0.0045
		6.4 Image Gallery	0.142			0.0143
		6.5 Graph and chart	0.045			0.0045
		6.6 Search Engine	0.349			0.0351

5 Combining GORE and AHP

The step 5 involves combining the outputs of GORE and AHP. The output of GORE is a set of hard and soft goals with contribution links. The output of AHP will be the global weights of all the leaf level soft goals. The first task here is to convert the contribution links into a numerical value.

Quality function deployment (QFD) is a "method to transform user demands into design quality, to deploy the functions forming quality, and to deploy methods for achieving the design quality into subsystems and component parts, and ultimately to specific elements of the manufacturing process [17]. Many QFD techniques [18, 19 and 20] employ a 4-point or 5-point scale to convert the qualitative links to quantitative value. To our approach, we consider a 4-point scale for converting the contribution links to a numerical value. Column three of Table 2 specifies the value which we have chosen in our approach.

We take each alternative and multiply the global weight of the soft goal with the numerical value of the contribution link which gives us the absolute value. This absolute value is then converted into relative value. This calculation for the first alternative Drupal 4.6.5 is shown in table 6. The same steps are applied for the other two alternatives. Due to space constraint, the calculations for other two alternatives are not shown.

Table 6. Evaluation of Drupal CMS

Soft goal	Global Weights(1)	Contribution Link(2)	Absolute Value(1*2)	Relative Value
1.1.1 Human vs PC Verification	0.0032	0	0	0
1.1.2 Authentication extensibility	0.0004	9	0.0036	0.000585
1.2.1 Support SSL Protocol	0.0159	6	0.0954	0.015501
1.2.2 SQL Security	0.0031	9	0.0279	0.004533
2.1.1 Web-based style	0.0136	9	0.1224	0.019888
2.1.2 Multiple templates per site	0.0166	9	0.1494	0.024275
2.1.3 Multiple menu types	0.0695	6	0.417	0.067755
2.1.4 Dynamic Menus	0.0343	0	0	0
2.1.5 Multilingual contents	0.0087	9	0.0783	0.012722

Table 6. (Continued)

2.2.1 Visitor tracking	0.0034	9	0.0306	0.004972
2.2.2 Contents tracking	0.00077	9	0.00693	0.001126
2.2.3 Log-in History	0.00103	9	0.00927	0.001506
3.1 Drag and Drop	0.1183	0	0	0
3.2 Preview	0.1183	9	1.0647	0.172995
3.3 Spell Checker	0.1183	0	0	0
3.4 Undo (upto 10 levels)	0.1183	9	1.0647	0.172995
3.5 Image Resizing	0.0228	0	0	0
3.6 File type Conversion	0.0228	0	0	0
4.1 Static Content export	0.0287	0	0	0
4.2 Page Caching	0.0862	9	0.7758	0.126054
5.1 Manuals	0.1047	9	0.9423	0.153107
5.2 Online Help	0.0269	9	0.2421	0.039337
5.3 Videos and Demos	0.0565	6	0.339	0.055082
5.4 Mailing list	0.0064	6	0.0384	0.006239
5.5 Public forum	0.0124	6	0.0744	0.012089
6.1 Blog	0.0351	9	0.3159	0.051328
6.2 Chat	0.0072	0	0	0
6.3 Forum	0.0045	9	0.0405	0.006581
6.4 Image Gallery	0.0143	0	0	0
6.5 Graph and chart	0.0045	0	0	0
6.6 Search Engine	0.0351	9	0.3159	0.051328

Next step involves calculation of effectiveness of each alternative in fulfilling soft goals which defines quality. This is done by matrix multiplication of relative values of each alternative with the global weights of soft goals as shown in fig 1.

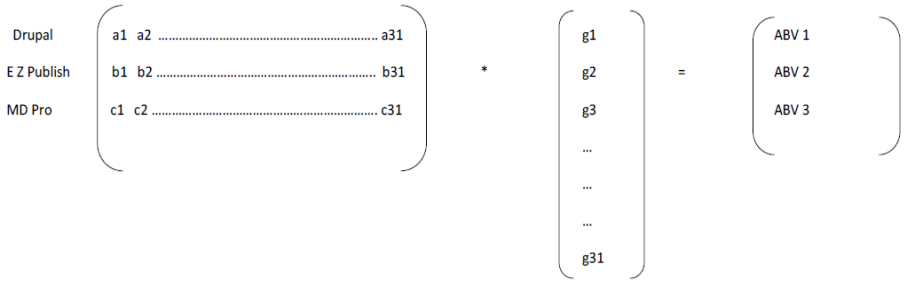


Fig. 1. Calculation of effectiveness of each alternative

a1 to a31 represents the relative values of Drupal 4.6.5 for identified soft goals.
b1 to b31 represents the relative values of e Z Publish 3 for identified soft goals.
c1 to c31 represents the relative values of MD Pro 1.0.76 for identified soft goals.
g1 to g31 represents the global weights of leaf level soft goals as shown in Table 5.
ABV1 to ABV3 represents the absolute value of alternatives Drupal 4.6.5, e Z Publish 3 and MD Pro 1.0.76 respectively. The absolute value obtained is then converted to relative value as shown in table 7.

Table 7. Relative values of each alternative

Alternative	Absolute Value	Relative Value	Ranking
Drupal	0.0995	0.42601	I
E Z Publish	0.06025	0.257949	III
MD Pro	0.07381	0.316039	II

The relative values shown in table 7 indicate that Drupal is the best alternative followed by MD Pro and E Z Publish which fulfills the set of soft goals identified for the problem.

6 Discussion

The steps proposed for integrating AHP and SQFD into the proposed method has been validated with case studies as explained earlier. Here are the following observations about the results obtained. The success of this approach depends on the following factors:

i. Clear identification of hard goals and soft goals.
ii. The identification of qualitative and quantitative criteria is an important step in the process since they directly contribute to identification of contribution links. These contribution links play a major role in choosing the best alternative.
iii. Appropriate stakeholders' needs to be identified in prioritizing soft goals.

AHP is a technique for computing priorities which is widely used in many domains. The data entered by a stakeholder is checked for consistency by measuring Consistency Index or Ratio (CI / CR). If CI > 0.1, then it implies that the priorities given by a stakeholder are not consistent. In the existing GORE literature, there exists technique which makes use of formal techniques [3, 14] in choosing the best alternative. They make use of temporal logic and label propagation algorithms. Our approach differs in adopting a quantitative way of evaluating the alternative using AHP.

7 Future Work

The future work will be carried out in the following areas:

- We have assumed the contribution links as either contributing positively or negatively. There are situations when the soft goals conflict with one another. For example, enhancing security can compromise on the performance of the system. This aspect will be considered in our future work.
- While converting the contribution links to a numerical value, we have adopted a linear scale. We will be studying the effect of using a non-linear scale of values on the final outcome.
- Topsis can be considered as a technique for prioritizing soft goals.
- Comparison of our approach with other techniques by identifying suitable metrics.
- Technique can be used in negotiating SLA (Service Level Agreement) for Cloud Computing where service quality plays a vital role.
- A tool to support the entire process.

8 Conclusion

In this paper, we have made an attempt to integrate GORE method with AHP in arriving at a decision. The method gives consistent results which depend on the quality of the goals identified and their priority. The same process might yield a different outcome if we are going to select a CMS for an e-commerce site. Security becomes more important and the global weights for the soft goals will change as a result. We have identified further work which will be carried out in the near future.

References

1. Doerr, J., Kerkow, D., Koenig, T., Olsson, T., Suzuki, T.: Non-Functional Requirements in Industry–Three Case Studies Adopting an Experience-based NFR Method. In: Proceedings of the 2005 13th IEEE International Conference on Requirements Engineering (1995)
2. Mylopoulos, J., Chung, L., Nixon, B.: Representing and Using Nonfunctional Requirements: A Process-Oriented Approach. IEEE Transactions on Software Engineering 18(6), 483–497 (1992)
3. Van Lamsweerde, A.: Goal-Oriented Requirements Engineering: A Guided Tour. In: Proceedings of the 5th IEEE International Symposium on Requirements Engineering. IEEE Computer Society, Washington (2001)
4. Ruhe, G.: Software Engineering Decision Support and Empirical Investigations–A Proposed Marriage. In: Workshop on Empirical Studies in Software Engineering, WSESE (2003)
5. Omasreiter, H.: Balanced Decision Making in Software Engineering–General Thoughts and a Concrete Example from Industry. In: First International Workshop on the Economics of Software and Computation, ESC (2007)
6. Easterbrook, S., Singer, J., Storey, M.-A., Damian, D.: Selecting Empirical Methods for Software Engineering Research. In: Guide to Advanced Empirical Software Engineering, Section III, pp. 285–311 (2008)
7. Hannay, J.E., Sjøberg, D.I.K., Dyba, T.: A Systematic Review of Theory Use in Software Engineering Experiments. IEEE Transactions on Software Engineering 33(2) (February 2007)
8. Price, J., Cybulski, J.: The Importance of IS Stakeholder Perspectives and Perceptions to Requirements Negotiation. In: AWR, Adelaide, Australia (2006)
9. Ivanović, A., America, P.: Information Needed for Architecture Decision Making. In: Proceedings of the ICSE Workshop on Product Line Approaches in Software Engineering (2010)
10. Lakshminarayanan, V., Liu, W., Chen, C.L., Easterbrook, S., Perry, D.E.: Software Architects in Practice: Handling Requirements. In: Proceedings of the Conference of the Center for Advanced Studies on Collaborative Research, CASCON (2006)
11. Hasan, M.S., Mahmood, A.A., Alam, M.J., Hasan, S.N., Rahman, F.: An Evaluation of Software Requirement Prioritization Techniques. International Journal of Computer Science and Information Security (IJCSIS) 8(9) (December 2010)
12. Saaty, T.L.: Decision making with the analytic hierarchy process. Int. J. Services Sciences 1(1) (2008)
13. Haag, S.E., Hogan, P.: Research issues in software quality function deployment: A new beginning for software engineering methodologies. In: Proceedings of Decision Sciences Institute 1992, San Francisco, California, DSI, Atlanta, Ga, November 23-25, pp. 926–928 (1992)
14. Castro, J., Kolp, M., Mylopoulos, J.: Towards Requirements-Driven Information Systems Engineering: The Tropos Project. Information Systems 27(6) (September 2002)
15. Kaiya, H., et al.: Identifying Stakeholders and Their Preferences about NFR by Comparing Use Case Diagrams of Several Existing Systems. In: 12th IEEE International Requirements Engineering Conference, RE 2004 (2004)
16. Vinay, S., Aithal, S., Adiga, S.: A Goal-Oriented Requirements Engineering Method for Analysing Conflicts. In: ICCANA, Nitte (January 2011)

17. Akao, Y.: Development History of Quality Function Deployment. In: The Customer Driven Approach to Quality Planning and Deployment, Asian Productivity Organization, Minato, Tokyo 107 Japan, p. 339
18. Herzwurm, G., et al.: QFD for customer focused Requirements Engineering. In: 11th IEEE International Requirements Engineering Conference (2003)
19. Hierholzer, A., Herzwurm, G., Schlang, H.: Applying QFD for Software Process Improvement at SAP AG, Walldorf, Germany. In: Proceedings of the Third Workshop on Software Quality. ACM (2003)
20. De Felice, F., Petrillo, A.: A multiple choice decision analysis: an integrated QFD–AHP Model for the Assessment of Customer Needs. International Journal of Engineering, Science and Technology 2(9), 25–38 (2010)

Multilanguage Based SMS Encryption Techniques

M. Rajendiran, B. Syed Ibrahim, R. Pratheesh, and C. Nelson Kennnedy Babu

Chettinad College of Engineering Technology
{mrajendiran,ibzz82,pratheeshr.nair}@gmail.com,
cnkbabu63@yahoo.in

Abstract. In our day to day life all of us are sending and receiving SMS to our friends, relatives and our loved ones. The frequent problem we find out here is lack of user applied encryption techniques to send their messages in a secure manner. In this paper we proposed a simple Multilanguage Encryption Technique for SMS (METSMS) applications based on symmetric key. In this METSMS technique plain text Alphabets are replaced by Multilanguage alphabets to generate the cipher text. And the cipher text is block cipher in nature. Due to this block ciphering nature we are transmitting less amount of cipher alphabets, thus the transmitting message length is small and secure. With this METSMS method, it is possible to do the encryption and decryption process for any language in the world which have Unicode.

1 Introduction

The encryption of information for the purpose of encoding, processing and sharing by using the cryptography method [5] was introduced only for a few languages of the world. It is also difficult to write cryptography by using only one algorithm. Though different methods of cryptography were introduced / used by different countries at various periods of time in the past, it is quite interesting to know that in India the cryptography method has been in use since 1600 BC[7]. In India 1.20 Billion people speak in 337 of the 348 languages and the rest 11 languages went out of usage by the passage of time. Around 8 million people speak in Tamil language, which is an oldest South Indian language. Tamil is the Administrative language in Tamil Nadu, Singapore, Malaysia and Sri Lanka. In this language, during the periods of 1600 BC , a cryptography method known as 'Porulkoal' (பொருள்கோள் *in Tamil*) was in existence. In the Ancient and Medieval Tamil grammar books 'Tholkappiam' (தொல்காப்பியம் *in Tamil*) [4] and 'Nannool' (நன்னூல் *in Tamil*), we can understand the meaning only by applying the 'Porulkoal' system. In Thirukkural, which is an ancient treasure of wisdom and also regarded as a Common Veda for the world by the Tamil people, also we can understand the meaning by using the 'Porulkoal' method. This 'Thirukkural' consists of 1330 couplets and over 100 couplets can be understood with the help of 'Porulkoal' method. 'Porulkoal' explains how the meaning of a poem or a literary work, with a secretly changed words, should be obtained.

After these periods, Caesar introduced the world famous character level cryptography method. By applying the shifting of words system, a sentence or a word can be secretly changed so that other do not understand the meaning. The Quantum

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 455–460.
springerlink.com

Cryptography method [6], which came into use afterwards, has several limitations. During 20th century period a number of cryptography methods were developed by using ASCII code system. But since the ASCII [2] numbers are made up of 64 to 127, which is shorter in length, one can easily understand them by applying the Brute force method. This method is developed by using only the English word and characters. Hence, no encryption can be made for other languages by applying this method.

In MULET [1] type encryption, the cipher text which is derived by applying the multi language contains not only a shorter multi language characters but also two types of Replacement method are explained in them. Hence there are two shortcomings in this process, But the METSMS method, which is now introduced here is free from the above two types of shortcomings and also the cipher text which we get at the end is also a Block cipher, which is a notable feature. This apart, the mapping array used here is made up with the help of Three Dimensional Multi language method and hence it is very difficult for others to understand the information even by attempting the Brute force attack or any similar types of attack.

2 Choice of Unicode Standard

In the ASCII code and EBCDIC code which were introduced earlier encompassed only the English language. Other languages cannot be encoded and processed with the ASCII and EBCDIC codes [3]. It was very difficult to represent the characters of all the world languages with the help of a single encoding method. Evidently many types of encoding methods were used to encode the languages of the European Union. The encoding methods that came into existence afterwards were used for languages which are being written from left to right. For languages which are written from Right to left, for example Arabic and Hebrew, there was no solution from the above encoding method. But through a single encoding system known as 'Unicode' method, we can make encoding, processing and sharing of information of all classical languages of the world. Normally it is very difficult to represent a language with 256 characters, like Japanese, with the previous methods of encryption. But with the help of Unicode [8] method it has now become easier for us to encrypt these languages also. We can represent any of the classical languages of the world (i.e. we can encode, process and share the information) through this Unicode method.

3 Existing Method

In MULET, the plain text is first converted to Unicode. Then it is converted as cipher text by using the Multilanguage character and Multilanguage numerical table. The cipher we get as above is known as stream cipher. The length of the cipher we got as above is too lengthier and also requires two types of Replacement method. The step by step procedure of MULET is clearly explained in Table 1.

Table 1. Characterwise Encryption Using One Dimensional Array Substitution

Plain Text	*G*	*O*	*D*	*I*	*S*	*G*	*R*	*E*	*A*	*T*
Unicode value (U)	*71*	*79*	*68*	*73*	*83*	*71*	*82*	*69*	*65*	*84*
Mapping Constant (M)	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*	*3*
Quotient value(Q)	*23*	*26*	*22*	*24*	*27*	*23*	*27*	*23*	*21*	*28*
Remainder Value(R)	*2*	*1*	*2*	*1*	*2*	*2*	*1*	*0*	*1*	*0*
Stream Cipher	ङ	ञ	ङ	ञ	ङ	ॅ	ञ	ञ	ञ	ञ

For the plain text English letters – 'God is great', first we have to write the equivalent Unicode. Assuming that the mapping constant M = 3, then we have to divide all the Unicode value by M. The stream cipher is made up with the resultant Quotient and Remainder value after the Division. By using Table 2, the Multilanguage character equivalent for the Remainder is written. Suppose a Remainder is repeated for one or more time, it is replaced with the help of Table 3.

Table 2. Mapping Array With M=3

Index	*0*	*1*	*2*
Character (Multilanguage)	ञ	ञ	ङ

In Table 1, since the remainder '2' for the letters 'S' and 'G' comes immediately after one another, the first '2' is replaced with the help of Table 2 and the second '2' is replaced with the help of Table 3.

Table 3. Multilanguage Set

Index	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Numerical (Multilanguage)	ऍ	ॾ	ॅ	ॸ	ॾ'	ॅ	ॾ२	ॅ	ऽ	ॾ

On the whole, the stream cipher is finally constructed through two Replacements. The cipher is also too lengthier compared to the cipher which we get through the proposed method. The mapping array in Table 2 is constructed with the help of Hindi language. This apart, Multilanguage numerical is used in Table 3. Here, it is easier to know the information because the number of Multilanguage characters used as mapping array in Table 2 is very less.

4 Proposed Method

The METSMS encryption method is clearly explained in Table 4. According to this method, the plain text – "God is great" is changed as Unicode. Let us assume the same mapping constant (i.e. M = 3). But, here the mapping constant array is created of a Three Dimensional value. The Unicode value is divided with M = 3 and then we get the Remainder and the Quotient value. The Remainder value is then grouped as x y z (3 x 3 groups). Then, for each group a Three Dimensional multi language

character from Table 5 is written. The cipher we get here is known as Block cipher. The Three Dimensional mapping array in Table 5 is created with the help of multi language characters. The x here denotes the Table number. y and z denote the Column and Row.

For example the mapping array constant for Hindi, Tamil and Telugu languages are given in Mapping Table 0, Mapping Table 1 and Mapping Table 3 respectively. Likewise all the languages of the world for which Unicode is available can be created with the mapping array constants. Since the last letter 'T' in Table 4 is short/ less by two pairs, two zero padding are created for it. That is why there shall be no difficulty in the process of encryption and decryption. Table 6 clearly explains the METSMS decryption process.

The MULET process gives the cipher which is lengthier when compared to the Block cipher derived from the METSMS process. Apart from this, the other notable feature is that one cannot understand the information even by attempting the Brute force method because the mapping array used in the METSMS process is of Three Dimensional multi language character.

Table 4. METSMS Encryption

Plain Text in English	G	O	D	I	S	G	R	E	A	T		
Multilanguage Characters Unicode	71	79	68	73	83	71	82	69	65	84		
Three Dimensional Index value = M	3	3	3	3	3	3	3	3	3	3		
Quotient =Q=U/M	**23**	**26**	**22**	**24**	**27**	**23**	**27**	**23**	**21**	**28**		
Reminder =R=U mod M	2	1	2	1	2	2	1	0	1	0	0	0
X,Y,Z	2,1,2			1,2,2			1,0,1			0,0,0		
Cipher Text	ల			ఆ			ళ			ఀ		

Table 5. Three Dimensional Mapping Array with M=3

Mapping Table 0

Row/ Column	0	1	2
0	ए	ओ	ॐ
1	व	ङ	ल
2	र	अ	आ

Mapping Table 1

	0	1	2
0	ఏ	ఓ	ళ
1	ఖ	ఙ	ఱ
2	ఋ	అ	ఆ

Mapping Table 2

	0	1	2
0	ఏ	ఓ	న
1	వ	ఘ	ల
2	ర	అ	ఆ

X
Z
Y

Table 6. METSMS Encryption

Cipher Text	ల			ఆ			ళ			ఀ		
X,Y,Z	2,1,2			1,2,2			1,0,1			0,0,0		
Reminder=R	2	1	2	1	2	2	1	0	1	0	0	0
Quotient=Q	**23**	**26**	**22**	**24**	**27**	**23**	**27**	**23**	**21**	**28**	**28**	**28**
Three dimensional index value=M	3	3	3	3	3	3	3	3	3	3	3	3
U=M*Q + R	71	79	68	73	83	71	82	69	65	84	84	84
Plain text	G	O	D	I	S	G	R	E	A	T		

5 Performance Analysis

A detailed comparison between the MULET and METSMS process methods are well explained with the help of the graph given above. It shows how the cipher in the METSMS method is shorter and how the cipher in the MULET method is lengthier. The Quotient value repeated in the MULET method is re-arranged with the help of Table IV. But there is no need for such types of replacements or re-arrangements in the METSMS method. Since the number of mapping array constants are higher in METSMS method when compared to the MULET method, the METSMS method is proved to be more secured.

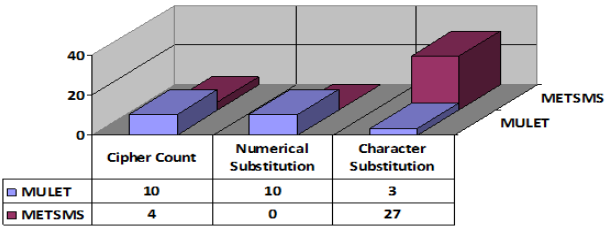


Fig. 1. MULET Vs METSMS

6 Conclusion

The METSMS encryption method is created with the help of the Unicode and hence is very easier for encrypting the world languages which have Unicode. That is why this method can be called as Universal Cryptographic Algorithm. This method also ensures better security, since the encryption is made by applying the Three Dimensional mapping array. Hence it is a very difficult task for others to understand the information. In order to decrypt the Block cipher with due process, the 'Q' value becomes an important factor. Hence, by using the advanced stenography method, the information can be safely passed on to others. In future, it shall become very easier to convert the cipher text in the METSMS process to Binary code.

References

1. Praveen Kumar, G., Parajuli, A.K.B., Choudhury, P.: MULET: A Multilanguage Encryption Technique. In: IEEE Seventh International Conference on Information Technology, pp. 779–782 (2010)
2. Shankar, T.N., Sahoo, G.: Cryptography by Karatsuba Multiplier with ASCII Codes. International Journal on Computer Applications, 53–60 (2010)
3. Sweetman, W.: The Bibliotheca Malabarica- An 18th century library of Tamil Literature. In: World Classical Tamil Conference (WCTC), part 10, Coimbatore (2010)
4. Shanmugam, S.V.: Tholkappiyar concept of Semantics. In: Published in World Classical Tamil Conference (WCTC), part 9, Coimbatore (2010)

5. Paar, C., Pelzl, J., Preneel, B.: Understanding Cryptography. A Text book for student and Practitioners. Springer (2010)
6. Sakthi Vignesh, R., Sudharssun, S., Jegadish Kumar, K.J.: Limitations of Quantum & The Versatility of Classical Cryptography: Comparative Study. In: Second International Conference on Environmental and Computer Science, pp. 333–337 (2009)
7. Farmer, S., Sproat, R., Witzel, M.: The collapse of the Indus-script thesis: The myth of a literate Harappan civilization. Electronic Journal of Vedic Studies 11, 19 (2004)
8. Unicode Character form, http://www.unicode.org

High Speed Low Power VLSI Architecture for SPST Adder Using Modified Carry Look Ahead Adder

Narayan V.S., Pratima S.M., Saroja V.S., and R.M. Banakar

B.V. Bhoomaraddi College of Engineering and Technology,
Vidyanagar, Hubli
narayansugur@yahoo.in
smpratima@gmail.com
{sarojavs,banakar}@bvb.edu

Abstract. Now a day various techniques have been developed for reducing the power consumption of VLSI designs, such as pipelining and parallel processing, reducing the dynamic power, voltage scaling, clock gating etc. To increase the processing speed of the silicon IC, logic gates are made using CNT FETs and designed using the VLSI technology. Lowering down the power consumption and enhancing the processing speed of IC designs are undoubtedly the two important design challenges in designing ICs.

The objective of a paper is to provide, high speed and low power adder. In this paper, a VLSI designed low power; high speed adder is proposed using the SPST approach. This adder is designed by applying the Spurious Power Suppression Technique (SPST) on a modified Carry look ahead adder, which is controlled by a detection unit using an AND gate. The proposed architecture is synthesized. In Xilinx RTL, chip XC5VLX50TFF1165 Vertex 5 series is selected for benchmarking. The timing report shows that to perform 16 bit addition the minimum period required for CLA adder is $14.685\,ns$ and MCLA adder requires $10.003\,ns$. The proposed adder requires 9.147 ns. An improvement of 37.71% is achieved in speed when compared to SPST with CLA adder and an improvement of 9.147% is achieved in speed when compared to CLA SPST with MCLA adder. The SPST adder implementation with AND gates have an extremely high flexibility on adjusting the data asserting time. This facilitates the robustness of SPST can attain 30% speed improvement.

Keywords: SPST Technique, Detection unit, MCLA Adder.

1 Introduction

Adders are commonly found in the critical path of many building blocks of microprocessors and digital signal processing chips. A fast and accurate operation of a digital system is greatly influenced by the performance of the resident adders. The most important for measuring the quality of adder designs in the past were propagation delay, and area. There are many different approaches to consider when designing a high performance adder.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 461–466.
springerlink.com

In array processing and in multiplication and division, multi-operand addition is often encountered. More powerful adders are required which can add many numbers instead of two together. One such design of a high-speed multi-operand adder is Carry-Look Ahead Adder (CLA).It can prevent time-consuming carry propagation and speed up computation.

There have been a number of studies developed to reduce the dynamic power by minimizing the switching activity [3] [4]. This paper presents a low power adder design with Spurious Power Suppression Technique (SPST) and Modified CLA adder (MCLA) in which there will be enhancement in the speed and reduction in power dissipation by minimizing the switching activity. The rest of the paper is organized as follows. In section 2 Spurious Power Suppression Technique is explained. Section 3 describes the architecture design of 32 bit SPST adder. In section 4 MCLA adder is discussed. Sections 5 pave a way to the implementation and results followed by conclusion in section 6.

2 Spurious Power Suppression Technique

The SPST uses a detection logic circuit to detect the effective data range of arithmetic units, e.g., adders or multipliers [1]. When a portion of data does not affect the final computing results, the data controlling circuits of the SPST latch this portion to avoid useless data transitions occurring inside the arithmetic units. This data controlling unit brings evident power reduction [2][5].

To illustrate the SPST, five cases of a 16-bit addition are explained as shown in Figure. 1. The 1st case illustrates a transient state in which the spurious transitions of carry signals occur in the MSP though the final result of the MSP are unchanged. The 2nd and 3rd cases describe the situations of one negative operand adding another positive operand without and with carry from LSP, respectively. Moreover, the 4th and 5th cases respectively demonstrate the conditions of two negative operands addition without and with carry-in from LSP. In those cases, the results of the MSP are predictable, therefore the computations in the MSP are useless and can be neglected. Eliminating those spurious computations will not only save the power consumed inside the SPST adder/subtractor but also decrease the glitching noises which will affect the next arithmetic circuits[1][5].

CASE 1:

(28)	**0000000000**111000
+(64)	0000000001000000
(92)	**0000000001**111000

(-28)	**1111111111**001000
+(92)	**0000000001**111000
(64)	**0000000001**000000

CASE 2:

(-61)	**1111111111** 000011
+(51)	000000000 0110011
(-10)	**1111111111**110110

CASE 3:

(60)	**0000000000**111100
+(-52)	**1111111111**001100
(+8)	**0000000000** 001000

CASE 4:

(252)	**0000000**11111100
+(-52)	**1111111**11001100
(200)	0000000011001000

CASE 5:

(-78)	**1111111**110110010
+(-22)	**1111111**111101010
(-100)	1111111110011100

Fig. 1. Example for SPST addition

3 Architecture of 16 Bit SPST Adder

In Figure 2 the architecture of SPST is illustrated through a low power adder/subtractor design. The adder/subtractor is divided into two parts, the most significant part (MSP) and the least significant part (LSP). The MSP of the original adder/subtractor is modified to include detection logic circuits, data controlling circuits, sign extension circuits, logics for calculating carry in and carry out signals. The LSP adder is implemented using MCLA adder. When a portion of data effect the final computing results, the data controlling circuits of the SPST latch this portion to avoid useless data transitions occurring in the arithmetic unit. The detection logic is used to detect the spurious activity using MSP bits of the numbers and the carry from the LSP adder. The outputs from the detection logic are close, carry_ctrl, sign. The generations of the three control signals is illustrated in the below mentioned equations.

$$A_{MSP}= \text{A [15:0]} \quad ; \quad B_{MSP}=\text{B [15:0];} \tag{1}$$

$$A_{and}= \text{A [15]*A [14]*.....A [0];} \tag{2}$$

$$B_{and}= \text{B [15]*B [14]*.....B [0];} \tag{3}$$

$$A_{nor}= \overline{\text{A [15]} + \text{A [14]} + \cdots \text{A [0]};} \tag{4}$$

$$B_{nor}= \overline{\text{B [15]} + \text{B [14]} + \cdots \text{B [0]};} \tag{5}$$

$$Close= \overline{(A_{and} + A_{nor}) * (B_{and} + B_{nor});} \tag{6}$$

Where A_{MSP} and B_{MSP} denote the MSP part of A and B, i.e. 9th to16th bit. When the bits of A_{MSP} and/or B_{MSP} are all ones, the value of A_{and} and/or that of B_{and} respectively becomes one, while the bits in A_{MSP} and/or B_{MSP} are all zeros, the value of A_{nor} and/or B_{nor} turn into one. Being one of the outputs of the detection logic unit, close denotes whether the MSP circuits can be neglected or not. When the two input operand can be classified into one of the five classes as shown in Figure 1, the value of the close become zero which indicates that the MSP circuits can be closed. Figure also we derive the Karnaugh maps which lead to the Boolean equations (7) and (8) for carry_ctrl and sign signals respectively.

$$\text{Carry\_ctrl= (CLSP} \oplus A_{and} \oplus B_{and})*(A_{and}+A_{nor})*(B_{and}+B_{nor}) \tag{7}$$

$$\text{Sign= } \overline{CLSP}*(A_{and}+B_{and}) + \text{CLSP}*A_{and}*B_{and} \tag{8}$$

In this example, the 16 –bit adder is divided into MSP and LSP at the place between 8th and 9th bit. Latches implemented by simple AND gates are used to control the input data of the MSP. When the MSP is necessary, the input data of MSP remain the same as usual, while the MSP is negligible, the input data of the MSP become zeros to avoid switching power consumption. From the equations (1) and (8), the detection logic unit of the SPST is designed as shown in figure which can determine whether the input data of MSP should be latched or not.

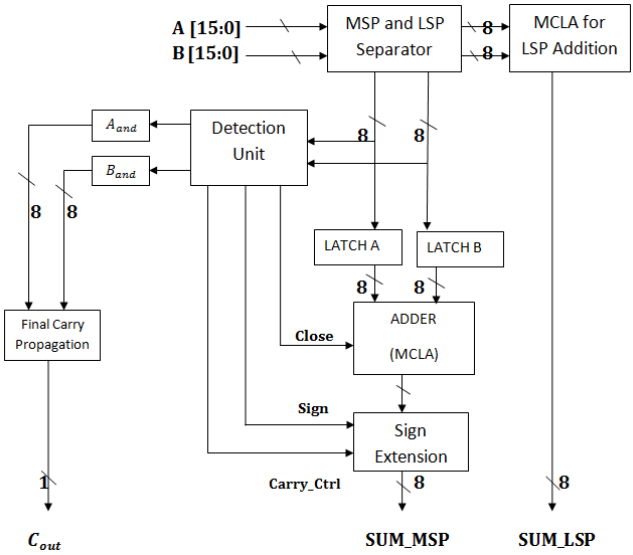


Fig. 2. Architecture of Low Power adder

The detection logic circuit is shown in Figure 3. The three output signals of the detection logic are close, control, sign.LSP adder is Modified carry look ahead adder.

4 Modified CLA Adder

The main concept of MCLA is to use NAND gates to replace AND and NOT gates of CLA adder. The design of MCLA adder consist of two parts
Arithmetic adder circuit

- Carry look ahead circuit

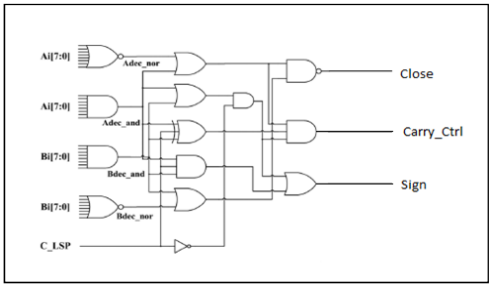


Fig. 3. Detection logic using AND gate

In MCLA a new full adder called metamorphosis of PFA is used. Here the generator g_i is produced by using NAND gate. The carries of the next stage are also generated using NAND gate and is written as

$$c_{i+1} = \overline{\bar{g}_i \left(\overline{p_i c_i}\right)} \tag{9}$$

Where C_{i+1} is the next stage carry, C_i is the previous carry, g_i is the generate logic and p_i is propagate logic. In metamorphosis PFA the signal $\bar{g}_i$ is implemented with NAND gate. It is faster than the g_i of PFA implemented with AND gates. The above generated carries are implemented in carry look ahead circuit. The AND gate is used to generate g bit. This can be simplified using NAND gate and a NOT gate to replace AND gate of the previous level. With this logic one gate delay is increased due to one NOT gate. To reduce this in MCLA NOT gate of previous stage is canceled with NOT gate of present stage. With the cancellation of NOT gate, one gate delay is reduced at every stage. This again adds to the improvement of the speed.

5 Implementation and Results of SPST Adder

To prove the architectural concept after functional validation, the proposed architecture is synthesized. In Xilinx RTL, chip XC5VLX50TFF1165 Vertex 5 series is selected for benchmarking. The tool does a high level of optimization and generates net list.

Table 1. Performance analysis of SPST adder using MCLA

Adder Type	CLA adder[1]	CLA adder + MCLA[2]	Proposed Adder[3]	%improvement	
				1&3	2&3
Delay	14.685 ns	10.003 ns	9.147 ns	37.71	8.55

Table 1 shows the performance analysis of the SPST adder when compared to other adder. The timing report shows that to perform 16 bit addition the minimum period required for CLA adder is 14.685 ns and MCLA adder requires $10.003\,ns$. The proposed adder requires 9.147 ns. An improvement of 37.71% is achieved in speed when compared to SPST with CLA adder and an improvement of $9.147\,\%$ is achieved in speed when compared to SPST with MCLA.

The gate utilization summary is presented in Table 2. The SPST adder is implemented using basic gates like AND, Inverter and OR. It uses 2 inputs, 3 inputs, 4 inputs, 5 input and 6 input LUT's. As shown in the table a total of 896 gates are required to implement this adder.

Table 2. Gate utilization summary of SPST Adder

No. of I/P LUTS						
TYPE OF GATES	LUT2	LUT3	LUT4	LUT5	LUT6	TOTAL
Inverter	0	42	0	104	114	260
AND	0	58	7	239	173	477
OR	1	25	4	79	50	159
Total						896

6 Conclusion

In this paper we propose a high speed low power VLSI Architecture for SPST adder using Modified Carry Look Ahead Adder. The SPST adder implementation with AND gates have an extremely high flexibility on adjusting the data asserting time. This facilitates the robustness of SPST can attain 30% speed improvement. The proposed architecture is synthesized in Xilinx RTL; chip XC5VLX50TFF1165 Vertex 5 series is selected for benchmarking. The timing report shows that to perform 16 bit addition the minimum period required for CLA adder is 14.685 ns and MCLA adder requires 10.003ns. The proposed adder requires 9.147 ns. An improvement of 37.71% is achieved in speed when compared to SPST with CLA adder and an improvement of 9.147% is achieved in speed when compared to CLA SPST with MCLA adder.

References

[1] Marimuthu, C.N., Thangaraj, P.: Low Power High Performance Multiplier. In: ICGST PDCS, vol. 8(1) (December 2008)

[2] Lakshmi Narayanan, G., Venkataramani, B.: Optimization Techniques for FPGA-Based Wave Pipelined DSP Blocks. IEEE Trans. Very Large Scale Integr (VLSI) Syst. 13(7), 783–792 (2005)

[3] Chen, K.H., Chao, K.C., Guo, J.I., Wang, J.S., Chu, Y.S.: An Efficient Spurious Power Suppression Technique (SPST) and its Applications on MPEG-4 AVC/H.264 Transform Coding Design. In: Proc. IEEE Int. Symp. Low Power Electron. Des., pp. 155–160 (2005)

[4] Benini, L., Micheli, G.D., Macii, A., Macii, E., Poncino, M., Scarsi, R.: Glitching power minimization by selective gate freezing. IEEE Trans. Very Large Scale Integr (VLSI) Syst. 8(3), 287–297 (2000)

[5] Praveen Kumar, M.: A Spurious-Power Suppression Technique For Multimedia/DSP Applications. International Journal of Advanced Engineering Scinces and Technologies 11(1), 035–051

ASIC Primitive Cells in Modified Gated Diffusion Input Technique

R. Uma and P. Dhavachelvan

Department of Computer Science, School of Engineering,
Pondicherry University, Puducherry, India
{uma.ramadass1,dhavachelvan}@gmail.com

Abstract. Power dissipation has become a prime constraint in high performance applications, especially in clocked devices like microprocessor and portable devices. Optimizations for the ASIC cells are crucial in order to improve the performance of various low power and high performance devices. The design criterion of primitive cells is usually multi-fold. Optimization of several devices for speed and power is a significant issue in low-voltage and low-power applications. These issues can be overcome by incorporating Gated Diffusion Input (GDI) technique. This paper mainly presents the design of primitive cells like AND, OR, NAND, NOR, MUX, XOR and XNOR cell in Modified Gate Diffusion Input Technique. This technique allows reducing power consumption, delay and area of digital circuits, while maintaining low complexity of logic design. Delay and power has been evaluated by Tanner simulator using TSMC 0.250 technologies considering minimum power design. The simulation results reveal better delay and power performance of proposed primitive cells as compared to existing GDI cell and CMOS at 0.250µm CMOS technologies.

Keywords: Gate Diffusion Input, ASIC, Full Adder Cell, Primitive cell.

1 Introduction

As VLSI circuits continue to evolve and technologies progresses, the level of integration is increased and higher clock speed is achieved. For such submicron CMOS technology area, topology selection, power dissipation and speed are very important aspect especially for designing Clocked Storage Element (CSE), adder circuits and MAC unit for high-speed and low-energy design like portable batteries and microprocessors.

The overall performance of a design depends on the logic technique used in terms of primitive cells that has been used in the hierarchy of the design. Therefore, careful design and analysis is required for construction of primitive cells like AND, OR, NAND, NOR, MUX, XOR and XNOR to reduce power, delay and area of larger design unit. Several optimization techniques for primitive cell design are reported in the literature [2-10]. Among Gate Diffusion Input (GDI) is a lowest power design technique which offers improved logic swing and less static power dissipation. Using this technique several logic functions can be implemented using less number of transistor counts. This method is suitable for design of fast, low-power circuits, using a reduced number of transistors (as compared to TG and CMOS).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 467–474.
springerlink.com

The main contribution of this paper presents the design of modified primitive cells of OR, AND, NAND, NOR at the circuit level designed based on the GDI technique. The modified primitive cells are constructed and its significant variation between CMOS and conventional GDI are compared. Though GDI technique offers low power, less transistor count and high speed, the major challenges occurs in the fabrication process. The GDI technique requires twin-well CMOS or Silicon on Insulator (SOI) process to realize a chip which increases the complexity as well as cost of fabrication.

The organization of the paper is as follows: The section 2, describes the basics of GDI. Section 3, presents the implementation of modified primitive cell of AND, OR, NAND, NOR, XOR, XNOR and MUX. Section 4 presents simulation result using Tanner EDA and it is compared with modified GDI and CMOS logic. Finally the conclusion is presented in section 5.

2 Basics of GDI Technique

The basic primitive of GDI cell consists of nMOS and pMOS as shown in Fig 1. A basic GDI cell contains four terminals – G (common gate input of nMOS and pMOS transistors), P (the outer diffusion node of pMOS transistor), N (the outer diffusion node of nMOS transistor), and D (common diffusion node of both transistors) [5]. Table 1 shows how a simple change of the input configuration of the simple GDI cell corresponds todifferent Boolean functions. Referring to Table1 most of the functions are realized using the function F1 and F2 since they are possible to realize using CMOS p-well process.

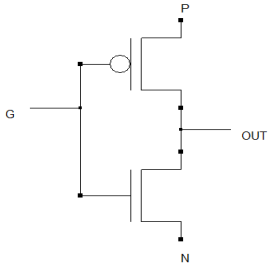


Fig. 1. Basic GDI cell

Table 1. Logic function implemented with GDI Technique

N	P	G	OUT	Function
'0'	*B*	*A*	$\overline{A}B$	*F1*
B	*'1'*	*A*	$\overline{A}+B$	*F2*
'1'	*B*	*A*	*A+B*	*OR*
B	*'0'*	*A*	*AB*	*AND*
C	*B*	*A*	$\overline{A}B+AC$	*MUX*
'0'	*'1'*	*A*	$\overline{A}$	*NOT*

The structure uses 3-inputs instead of 2-input in CMOS logic in order to attain implementation of complicated logic function with less number of transistors. Normally in CMOS the pMOS is connected to VDD and nMOS is connected to VSS. But in GDI technique both pMOS and nMOS are given with independent inputs so as to accommodate more logic function thereby minimizing transistor count as well as power dissipation. Most of these functions presented in Table1 require 6–12 transistors when it is implemented with CMOS, Transmission Gate, but they are simple in GDI design

logic since each design implementation requires only a minimum of two transistors. GDI enables simpler gates, lower transistor count, and lower power dissipation. Multiple-input gates can be implemented by combining several GDI cells. The buffering constraints, due to possible threshold (V_t) drop, are described in detail in [5], as well as technological compatibility with CMOS and SOI. The primitive cell construction using GDI technique is shown in Fig 2. The primitive cells are simulated using Tanner EDA with BSIM3v3 250nm technology with supply voltage ranging from 1V to 2V in steps of 0.2V which listed in Table 2.

Table 2. Delay and power of primitive cell in GDI Technique

Primitive Cell	Switching Delay Of GDI gates (ps)	Transistor Count for GDI cell	Avg Power in GDI Technique (µW)
2-input AND	0.200	2	1.286
2-input OR	0.280	2	1.30
3-input AND	0.500	4	1.45
3-input OR	0.503	4	1.55
2-input NAND	0.520	4	0.657
2-input NOR	0.540	4	0.680
2-input XOR	0.545	4	1.48
2-input XNOR	0.540	4	1.50
3-input XOR	0.432	6	1.5

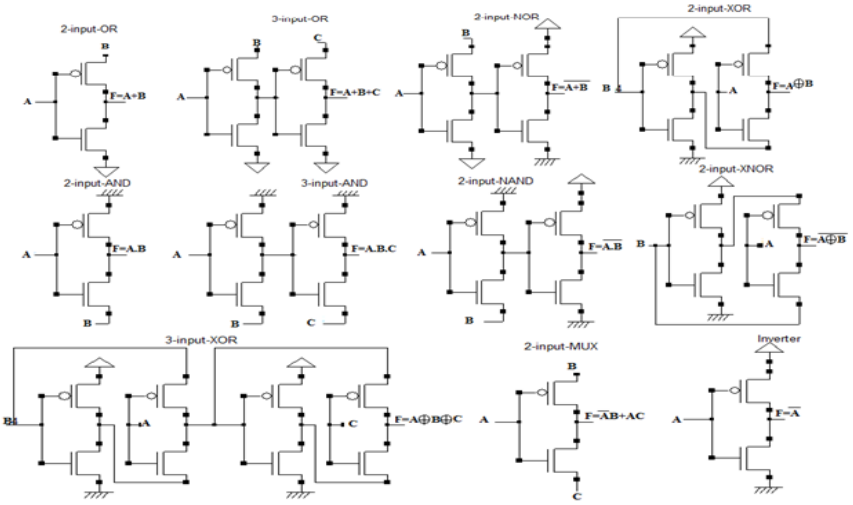


Fig. 2. Primitive cells in GDI technique

3 Modified GDI Primitive Cells

In this work a modified primitive logic gates have been implemented in 0.250nm technology and it is compared with CMOS logic. Fig 3 shows the construction of modified basic gates of AND, NOR, AND, NAND, XOR, XNOR and MUX. The modified GDI primitive logic function (MGDI) is shown in Table 3.

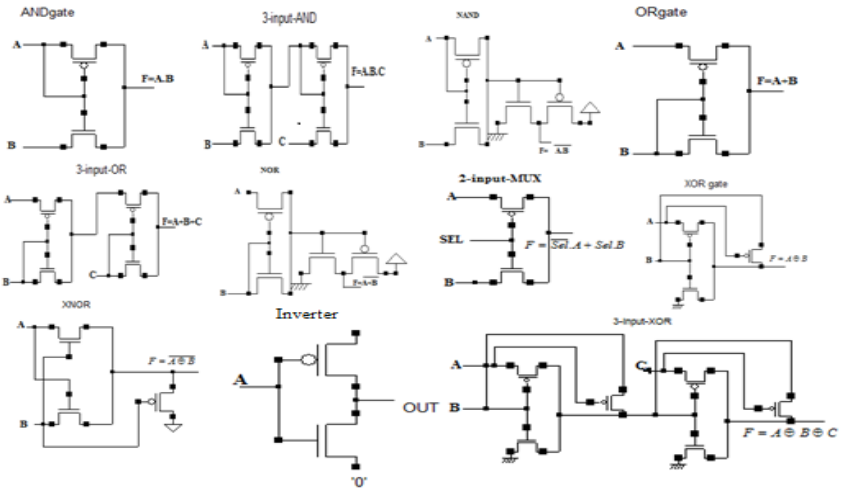


Fig. 3. Catalogue of modified GDI logic gates

As an example the operation of AND gate is elucidated. For AND gate the drain of pMOS is connected with input 'A' and the source of nMOS is connected with input 'B'. The gate terminal G is connected with 'A'. When both the inputs are zero then pMOS will operates in linear whereas nMOS in cut-off. While A='1' and B = '0' then pMOS in cut-off and nMOS in cut-off. Similarly for A='0' and B = '1' then pMOS in cut-off and nMOS in linear. Therefore for A='1' and B='1' pMOS in saturation and nMOS in linear thereby producing the output as 1. The switching characteristics of AND is shown in Fig 4. The logical level for different input combination will be:

For A=0 and B=0: pMOS in Linear: $V_{in} - V_{tp} < V_{out} < V_{DD}$
nMOS in Cut-off: $V_{in} < V_{tn}$
For A=1 and B=0: pMOS in Cut-off: $V_{in} > V_{DD} + V_{tp}$
nMOS in Cut-off: $V_{in} < V_{tn}$
For A=0 and B=1: pMOS in Cut-off: $V_{in} > V_{DD} + V_{tp}$
nMOS in linear: $0 < V_{out} < V_{in} - V_{tn}$
For A=1 and B=1: pMOS in linear: $V_{in} - V_{tp} < V_{out} < V_{DD}$
nMOS in linear: $0 < V_{out} < V_{in} - V_{tn}$

Table 3. Logic function implemented with MGDI Technique

N	P	G	OUT	Function
A	*B*	*B*	*A+B*	*OR*
B	*A*	*A*	*AB*	*AND*
B	*A*	*C*	$\overline{C}A+CB$	*MUX*
'0'	*'1'*	*A*	$\overline{A}$	*NOT*

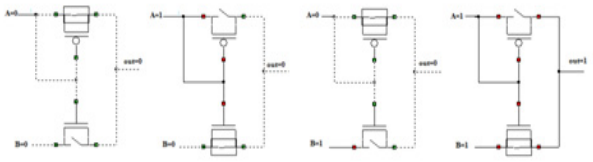


Fig. 4. Switching Characteristics of MGDI AND gate

For OR gate the drain of pMOS is connected with input 'A' and the source of nMOS is connected with input 'B'. The gate terminal G is connected with 'B'. When both the inputs are zero then pMOS will operates in linear whereas nMOS in cut-off. While A='1' and B = '0' then pMOS in linear and nMOS in cut-off. Similarly for A='0' and B = '1' then pMOS in cut-off and nMOS in linear. For A='1' and B='1' pMOS in saturation and nMOS in linear thereby producing the output as 1. The switching characteristics of OR is shown in Fig 5. The logical level for different input combination will be:

For A=0 and B=0: pMOS in Linear: $V_{in} - V_{tp} < V_{out} < V_{DD}$
nMOS in Cut-off: $V_{in} < V_{tn}$
For A=1 and B=0: pMOS in linear: $V_{in} - V_{tp} < V_{out} < V_{DD}$
nMOS in Cut-off: $V_{in} < V_{tn}$
For A=0 and B=1: pMOS in Cut-off: $V_{in} > V_{DD} + V_{tp}$
nMOS in linear: $0 < V_{out} < V_{in} - V_{tn}$
For A=1 and B=1: pMOS in linear: $V_{in} - V_{tp} < V_{out} < V_{DD}$
nMOS in linear: $0 < V_{out} < V_{in} - V_{tn}$

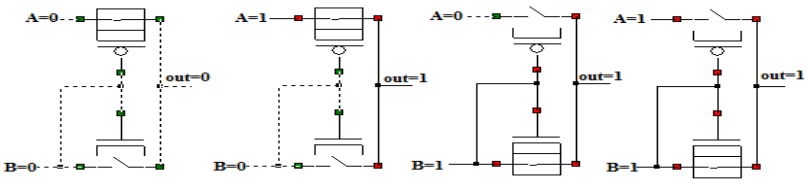


Fig. 5. Switching Characteristics of MGDI OR gate

The performance analysis of MGDI and CMOS logic is presented in Table 4. The performance evaluation is made with respect to switching delay, transistor count and average power consumed by MGDI and CMOS logic. From this analysis it is observed that the modified GDI performance is better when comparing to CMOS logic. In CMOS the number of transistor used to realize a function is twice that of MGDI. All the transistors used to design XOR and XNOR has only three transistors in MGDI whereas it is 8 in CMOS logic. The power consumed by CMOS is slightly higher than MGDI.

Table 4. Delay and power of primitive cell in MGDI and CMOS logic

Primitive Cell	Switching Delay Of MGDI gates (ps)	Switching Delay Of CMOS gates (ps)	Transistor Count for MGDI cell	Transistor Count for CMOS cell	Avg Power in MGDI Technique (µW)	Avg Power in CMOS Technique (µW)
2-input AND	0.180	0.240	2	6	0.986	1.698
2-input OR	0.180	0.270	2	6	1.20	1.550
3-input AND	0.490	0.542	4	8	1.34	1.872
3-input OR	0.350	0.492	4	8	1.33	1.723
2-input NAND	0.242	0.280	4	4	0.540	0.604
2-input NOR	0.280	0.300	4	4	0.654	0.756
2-input XOR	0.362	0.567	3	8	1.23	1.5
2-input XNOR	0.363	0.567	3	8	1.23	1.5
3-input XOR	0.432	0.678	6	12	1.5	1.75

4 Simulation and Performance Analysis of MGDI

The modified GDI primitive cells are simulated using Tanner EDA with BSIM3v3 250nm technology with supply voltage ranging from 1V to 2V in steps of 0.2V. All the MGDI cells are simulated with multiple design corners (TT, FF, FS, and SS) to verify that operation across variations in device characteristics and environment. The W/L ratios of both nMOS and pMOS transistors are taken as 2.5/0.25µm. To establish an unbiased testing environment, the simulations have been carried out using a comprehensive input signal pattern, which covers every possible transition for primitive cells. The performances of these primitive cells are reported in Table 5.

The performance of these primitive cells has been analyzed in terms of delay, transistor count and power dissipation with respect to MGDI, CMOS and GDI technique. It is observed that modified cells have least delay and power consumption when compared to CMOS and GDI technique. The design of XOR and XNOR has only 3 transistors when compare to GDI logic. The overall performance of MGDI, GDI and CMOS logic is shown in Table 5. From the performance analysis the modified primitive cell has the minimum delay and power dissipation when compare to conventional GDI technique. The performance analysis of MGDI, GDI and CMOS in terms delay, gate count and power dissipation is shown in Fig 6.

Table 5. Delay and power of primitive cell in MGDI, GDI and CMOS logic

Primitive Cell	Delay Of MGDI gates (ps)	Delay Of GDI gates (ps)	Delay Of CMOS gates (ps)	Gate Count for MGDI Cell	Gate Count for GDI cell	Gate Count for CMOS cell	Avg Power in MGDI Technique (µW)	Avg Power in GDI Technique (µW)	Avg Power in CMOS Technique (µW)
AND2	0.180	0.200	0.240	2	2	6	0.986	1.286	1.698
OR2	0.180	0.280	0.270	2	2	6	1.20	1.30	1.550
AND3	0.490	0.500	0.542	4	4	8	1.34	1.45	1.872
OR3	0.350	0.503	0.492	4	4	8	1.33	1.55	1.723
NAND2	0.242	0.520	0.280	4	4	4	0.540	0.657	0.604
NOR2	0.280	0.540	0.300	4	4	4	0.654	0.680	0.756
XOR2	0.362	0.432	0.567	3	4	8	1.23	1.35	1.5
XNOR2	0.363	0.456	0.567	3	4	8	1.23	1.36	1.5
XOR3	0.432	0.523	0.678	6	8	12	1.5	1.25	1.75

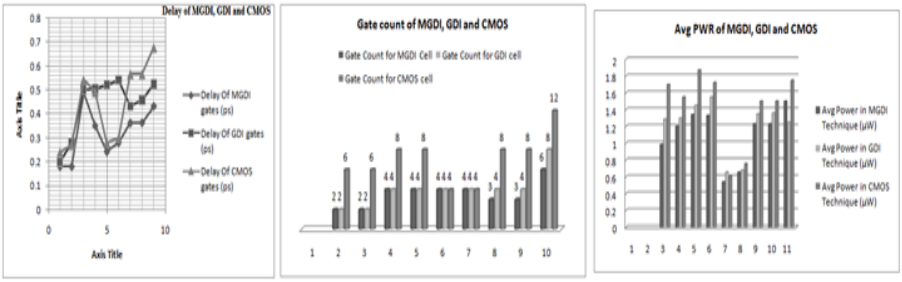


Fig. 6. Performance Comparison of MGDI, GDI and CMOS logic

From the delay graph it is noticed that the switching delay of NAND and NOR gate is much less when compare to AND and OR gate for all MGDI, GDI and CMOS logic. Similarly the average power dissipation of NAND and NOR are comparatively less when compare to other gates. From the overall analysis the modified primitive cell has low area, delay and power when compare to conventional GDI and CMOS logic. So if these primitive cells are incorporated in larger hierarchy design the overall performance will be superior when compare to GDI and CMOS logic.

5 Conclusion

An extensive performance analysis of modified primitive cells of AND, OR, NAND, NOR, MUX, XOR and XNOR has been presented. The performance of these MGDI was analyzed in terms of transistor count, delay and power dissipation using Tanner EDA with TSMC MOSIS 250nm technology and it is compared with conventional GDI and CMOS logic. The simulation results reveal better delay and power performance of proposed primitive cells as compared to existing GDI cell and CMOS at 0.250µm CMOS technologies. The work presented in this paper gives more insight and deeper understanding of GDI technique and provides scope to include this MGDI in larger design to enhance the performance in terms of area, delay and power consumption.

References

[1] Agrawal, A.K., Wairya, S., Nagaria, R.K., Tiwari, S.: A New Mixed Gate Diffusion Input Full Adder Topology for High Speed Low Power Digital Circuits. World Applied Sciences Journal (Special Issue of Computer & IT) 7, 138–144 (2009)

[2] Zimmermann, R., Fichtner, W.: Low-power logic styles: CMOS versus pass-transistor logic. IEEE J. Solid-State Circuits 32, 1079–1090 (1997)

[3] Calhoun, B., Cao, Y., Li, X., Mai, K., Pileggi, L., Rutenbar, R., Shepard, K.: Digital Circuit Design Challenges and Opportunities in the Era of Nanoscale CMOS. Proceedings of the IEEE 96(2), 343–365 (2008)

[4] Al-Assadi, W., Jayasumana, A.P., Malaiya, Y.K.: Pass-transistor logic design. Int. J. Electron. 70, 739–749 (1991)

[5] Morgenshtein, A., Fish, A., Wagner, I.A.: Gate-Diffusion Input (GDI): A Power-Efficient Method for Digital Combinatorial Circuits. IEEE Transactions on Very Large Scale Integration (VLSI) Systems 10(5) (October 2002)
[6] Uma, R., Dhavachelvan, P.: Performance of Full Adders with Skewed Logic. In: International Conference on Advances in Computing and Communications, ACC 2012 (2012)
[7] Balasubramanian, P., John, J.: Low Power Digital design using modified GDI method. IEEE (2006)
[8] Uma, R., Dhavachelvan, P.: Analysis on Impact of Behavioral Modeling in Performance of Synthesis Process. In: Third International workshop on VLSI, VLSI 2012 (2012)
[9] Uma, R., Dhavachelvan, P.: Performance of Adders with Logical Optimization in FPGA. In: First Interantional Conference on Signal, Image Processing and Pattern Recognition, SPRR 2012 (2012)
[10] Morgenshtein, A., Shwartz, I., Fish, A.: Gate Diffusion Input (GDI) Logic in Standard CMOS Nanoscale Process. In: 2010 IEEE 26th Convention of Electrical and Electronics Engineers in Israel (2010)
[11] Agrawal, A.K., Mishra, S., Nagaria, R.K.: Proposing a Novel Low-Power High-Speed Mixed GDI Full Adder Topology. IEEE (2010)

Latent Dirichlet Allocation Model for Recognizing Emotion from Music

S. Arulheethayadharthani and Rajeswari Sridhar

Anna University
{heetha,rajisridhar}@gmail.com

Abstract. The recognition of emotion has become a multi-disciplinary research area that has received great interest. Recognizing emotion of audio data will be useful for content-based searching, mood detection etc. The goal of this paper is to elaborate a system that automatically recognizes the emotion of the music. We present a technique used for document classification, Latent Dirichlet allocation (LDA) for the purpose of identifying emotion from music. The recognition process consists of three steps. In the first step, extractions of ten distinct features from music are performed followed by Clustering of values of these features, and finally in the third step an LDA model for each of the emotions is constructed. After constructing the LDA the emotion of the given music is identified. This model was tested on South Indian film music to recognize 6 emotions happy, sad, angry, love, disgust, fear and achieved an average accuracy of 80%.

1 Introduction

The word emotion includes a wide range of observable behaviours, expressed feelings, and changes in the body state. Emotion is a feeling that is private and subjective. There are eight basic emotions happy, sadness, acceptance, disgust, anger, fear, surprise, anticipation [1]. All emotions are a combination of these basic emotions. With the recent advances in the field of music information retrieval, there is an emerging interest in analyzing and understanding the content of music which includes emotion, genre, lyricist, etc.

Due to the diversity and richness of music content, many researchers have been pursuing a multitude of research topics in this field, ranging from computer science, digital signal processing, mathematics, and statistics applied to musicology and psychology [2].Music is not only a set of sounds; it evokes emotions based on listeners' perspective. Music emotion plays an important role in music retrieval, mood detection and other music-related applications.

The goal of this paper is to develop a music emotion recognition system for Tamil songs. This emotion recognition system considered both vocal and instrumental sounds of the given music piece.

In the following section, we present a brief overview of existing work related to emotion recognition process. Section 3 gives insight into various feature vectors that were tried and about implementation of LDA. Section 4 gives evaluation of a system. Concluding remark and references follow in section 5.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 475–481.
springerlink.com

2 Related Works

Music emotion plays an important role in music retrieval, mood detection and other music-related applications. Many researchers have explored models of emotions and factors that give rise to the perception of emotion in music. Some researchers investigate the problem of automatically recognizing emotion in music by proposing algorithms for the same based on music and signal features that convey emotion [2].

An emotion recognition system called SMERS - SVR based music emotion recognition system has been developed by Byeong-jun Han et al [2].The authors have extracted seven different features like pitch, tempo, loudness, tonality, key, rhythm, harmonics and mapped them into eleven categories of emotion: angry, bored, calm, excited, happy, nervous, peaceful, pleased, relaxed, sad and sleepy. In which three types of classifiers were employed to compare their performance SVR (Support Vector Regression), SVM (Support Vector Machine), GMM (Gaussian Mixture Model). Since SVM is a binary classifier, in SVMs-based classification, one-to-one training policy was employed. Authors have trained two regression functions to represent arousal and valence respectively. Finally, GMM was trained using 7 Gaussian models for arousal and valence sets. Each GMM is trained using the Expectation Maximization (EM) algorithm. This system had a maximum accuracy of 91.52% (151 of 165 samples). By changing coordinate system into polar, the accuracy was increased to 94.55% (156 of 165 samples) using SVR and 92.73% (153 of 165 samples) using GMM.

In another system proposed by Alicja Wieczorkowska et al, the authors' elaborated a tool for content - based searching of music files [1]. A Set of descriptors like frequency, level, tristimulus, brightness, irregularity, even harm and odd harm were extracted and labelled with 8 classes of emotion. The purpose of their research was to perform parameterization of audio data for the purpose of automatic recognition of emotion in music. The authors have employed K-NN (k-nearest neighbours) algorithm for classifying emotion. In this algorithm, the class of unknown sample is assigned on the basis of k nearest neighbours of known origin.

Ashutosh kulkarni [3] extracted features like MFCC, Spectral Flux, Spectral centroid, zero crossing rate, Average energy, Spectral roll off and proposed a novel algorithm to segment an audio piece into structural components. The author have classified each frame of the song into three classes Non vocal, Vocal or Silence using multinomial softmax regression. Then they have used a Hidden Markov Model to smooth the previous output as well as enforce the time dependent structuring.

In another work proposed by Qi Lu[4] the authors have applied AdaBoost algorithm to integrate MIDI, audio and lyrics information and proposed a two-layer classifying strategy called Fusion by Subtask Merging for 4-class music emotion classification.

In the work proposed by Erik M.Schmid, [5] the authors have presented a system linking models of acoustic features and human data to provide estimates of the emotional content of music according to the arousal-valence space. They have extracted features like MFCC, Chroma, and Statistical Spectrum Descriptors (SSD).

Though they are many classification algorithms like SVM, GMM, HMM for emotion recognition, each one has its own disadvantages as well as advantages. In SVMs-based classification, one-to-one training policy was employed for training

single emotion, since SVM does not support multi-classification [1]. When we consider the next classification algorithm HMM, to arrive better emotion recognition we need a large training corpus. On the other hand, GMM is based on clustering using the features set based on mean and variance of the feature values.

Considering the above factors we wanted to verify the LDA classification technique to recognize emotion of the given music which was initially proposed for document classification. LDA is a probabilistic, generative model for discovering latent semantic topics in large collections of text data. We extended the basic concepts of topic modeling to construct our emotion model using LDA. In a work proposed, Diane J. Hu [6] considered LDA as a statistical approach to document modeling that discovers latent semantic topics in large collection of text documents. Latent topics are discovered by identifying groups of words in the corpus that frequently occurs together within documents.

In another work by David M.Blei, [7] they have described a new model for collections of discrete data that provides full generative probabilistic semantics for documents. Documents are modeled via a hidden Dirichlet random variable that specifies a probability distribution on a latent, low dimensional topic space. In our proposed system, we show use of LDA for recognizing emotion, where the LDA is constructed as a two level step based on the features that convey emotion.

3 Emotion Recognition System

3.1 System Description

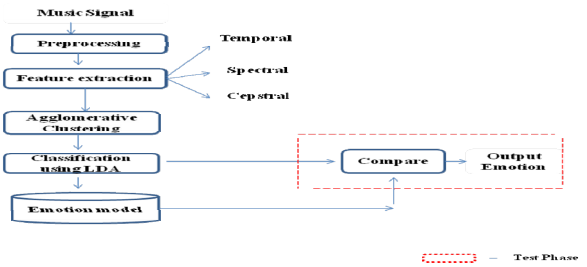


Fig. 1. Overall System Design

First, we extracted seven different music features, such as Zero Crossing rate, short term energy, energy entropy, spectral centroid, spectral flux, spectral roll off and MFCC from segmented music signal of size two seconds. Along with these seven distinct features three more additional features were chosen to enhance the performance of emotion recognition system. The additional features are rhythm, tempo and harmonic.Then, the features values are clustered using agglomerative clustering. In our proposed system clustering of emotion takes place in bottom up fashion. Finally a LDA is implemented to classify emotion of a song.

The emotion recognition process involves two phase: Training phase and testing phase. In Training phase, we extract features from segmented music signal. Then we perform clustering of the extracted feature. By using this features generate the clusters using agglomerative clustering then construct the model using LDA classifier. By performing classification, we construct emotional model to differentiate various emotions. In Testing phase, the same features were extracted from the music signal and by comparing the constructed emotion model (training phase) and the test input, our system determine the emotion.

3.2 Agglomerative Clustering

Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. In our proposed system clustering of emotion takes place in bottom up fashion. Features as indicated in Section 3.1 of each segment of a music signal are clustered using k-means algorithm. Then, the clusters are merged from generalized emotion to specific emotion using agglomerative clustering.

3.3 Construction Using LDA

LDA posits that words carry strong semantic information, and documents discussing similar topics will use a similar group of words [6]. Latent topics are thus discovered by identifying groups of words in the corpus that frequently occur together within documents. In this way, LDA models documents as a random mixture over latent topics, with each topic being characterized by its own particular distribution over words [7]. In this paper, we show that LDA is not only useful in the text domain, but can also find application in music domain. In one of the work for Raga identification LDA has been modelled [8]. In this work, we discuss algorithms that extend LDA to accomplish the task of emotion recognition of music signal.

Topic models provide a simple way to analyze large volumes of unlabeled text. A "topic" consists of a cluster of words that frequently occur together. Using contextual clues, topic models can connect words with similar meanings and distinguish between uses of words with multiple meanings. We have extent this topic model concept for music domain, by considering words as features and topic as emotions and is given in Figure 2. While constructing the emotion recognition system using LDA, the parameters that we have considered are given below:

α – Probability distribution of a feature in emotion set
β– Probability distribution of a feature in particular emotion
ϕ – Emotion weight vector

Initially, we assign random value for α and β during the first iteration. During the training process the values α and β will be refined in the subsequent iteration. We create a feature matrix based on the features extracted from the previous module. For each emotion, the feature values, α and β values will be stored. These values will change for each iteration while training the new segment of music. The distribution for each emotion will be identified based on the value which was stored under each emotion.

The step to find out the distribution for each emotion is given below:

STEP 1: p(ø/ α) – choose most likely emotions ø

STEP 2:
2a. P(E/ø) – choose Emotion Z
2b. P(F /Z, β)- choose features F

STEP 3: Inference - Determine posterior distribution
p(ø , E/F, α, β)= p(ø , E, F/ α, β)/p(F/ α, β)

At the end of training process, the model is constructed for each emotion. During the test phase, we compare the emotion model constructed with the feature values of the new song given for testing and recognize the emotion.

4 Evaluation

Emotion recognition of Tamil music involves recognition of emotion based on the Features collected from the training samples. For evaluation of the emotion recognition system, music samples were collected from Tamil film song portal. 260 songs (100 songs in sad, 60 in joy, 70 in love, 10 in angry 10 in disgust, 10 in anger) from different emotions collected to train our system. Songs with vocal and instrumental characteristic were considered in both training and testing phase.

During the evaluation step we have compared the accuracy of the emotion recognition using 7 features and 10 features. This comparison helps to find the below points,

1. To know whether the features we are calculating necessary features or the features that we are calculating are do not have any influence in emotions
2. If the number of features increased whether the accuracy of emotion recognition also increase or not
3. What are the frequent false recognitions when we consider less amount of features

Table 1. Evaluation

Emotion	7 Features	10 Features
Happy	85%	86%
Sad	82%	83%
Love	80%	83%
Angry	80%	82%
Fear	70%	70%
Disgust	79%	81%

Based on the above table we observe the following inference,

When we considering 7 features, we got some false emotion recognitions like LOVE songs are classified under SAD emotion, ANGRY songs are classified under HAPPY emotion, DISGUST songs are classified under SAD emotion.

Hence during this analysis we found that increasing the number of features increases the quality of the emotion recognition and all the 10 features that we considered are having considerable impact on the emotion recognition.

We further analyzed the results to identify the reason behind the increase in accuracy of emotion recognition due to the additional 3 features and observed the following:

- When we consider happy emotion songs, most of the songs are having tempo and rhythm value as high. When we consider the seven features we have not included these features which influence in happy emotion.
- Similarly when we consider love songs, rhythm, harmonics have more influence when compared to other features.
- However, even after increasing the number of features to 10, the accuracy of emotion recognition of songs whose emotion is 'fear' are not up to the expected level.

Therefore, in order to increase the accuracy we need to consider additional features of the signal for constructing the LDA.

5 Conclusion and Future Work

Thus an emotion recognition system for Tamil songs has been developed based on the ten different features and clustering them. Based on the values in a particular cluster we have found the range of values for particular emotion. Then the features are classified and are used to construct emotion recognition model using LDA in training phase of the project. The use of LDA technique improves the accuracy of recognition process greatly and thus enables making of an emotion recognition system. In testing phase, the model that we constructed was validated against the huge number of manually evaluated songs and the accuracy of the proposed system is considerably good for songs.

However, our proposed system recognized emotion of a song with high level of accuracy, the accuracy of emotion recognition for a particular instrument is low since we considered only ten features. Consideration of more features will improve the accuracy of the emotion recognition system which could be based on instrument.

References

[1] Wieczorkowska, A., Synak, P., Lewis, R., Ras, Z.: Extracting Emotion from Music Data. In: IEEE ICDM 2006 (2006)

[2] Han, B.-J., Rho, S., Dannenberg, R.B., Hwang, E.: SMERS: Music Emotion Recognition using Support Vector Regression. In: 10th International Society for Music Information Retrieval Conference 2009, pp. 651–656 (2009)

[3] Kulkarni, A., Iyer, D., Sridharan, S.R.: Audio Segmentation. In: IEEE, International Conference on Data Mining, ICDM (2001)
[4] Lu, Q., Chen, X., Yang, D., Wang, J.: Boosting for multi-modal music emotionclassification. In: 11th International Society for Music Information Retrieval Conference (ISMIR 2010), pp. 105–110 (2010)
[5] Schmidt, E.M., Turnbull, D., Kim, Y.E.: Feature Selection for Content-Based, Time-Varying Musical Emotion Regression. In: International Conference on Data Mining (ICDM), pp. 267–273 (2010)
[6] Hu, D.J.: Latent Dirichlet Allocation for Text, Images and Music. University of California, San Diego (2009)
[7] Blei, D.M., Ng, A.Y., Jordon, M.I.: Latent Dirichlet Allocation. University of Californiya, Berkeley (2009)
[8] Sridhar, R., Subramanian, M., Lavanya, B.M., Malinidevi, B., Geetha, T.V.: Latent Dirichlet Allocation Model for raga identification of Carnatic Music. Journal of Computer Science, 1711–1716 (2011)

Investigations on the Routing Protocols for Wireless Body Area Networks

Jayanthi K. Murthy[1], Thimmappa P.[1], and V. Sambasiva Rao[2]

[1] BMS College of Engineering, Bangalore
[2] PES Institute of Technology
jayanthi.ece@bmsce.ac.in,
thimmappa.p@gmail.com, vsrao@pes.edu

Abstract. In the last few years, wearable health monitoring systems have gained the attention of various researchers in order to cope with the rising cost of the health care systems. This paper addresses the use of standard protocols, particularly IEEE 802.15.4 and ZigBee, which are capable of supporting the Quality of Service (QoS) requirements in Wireless Body Area Networks. Simulation study on the routing protocols was done to investigate the protocol suitable for Body Area Networks and AODV was found to meet the requirements of energy and QoS.

Keywords: Wireless Body Area Networks, health monitoring, IEEE, 802.15.4, ZigBee Routing Protocols.

1 Introduction

Technological advances in wireless communications, microelectronics and physiological sensing allow miniature, lightweight, low power, intelligent monitoring devices. A number of these devices can be integrated into a Wireless Body Area Network (WBAN), a new enabling technology for health monitoring. Thus, the ubiquitous healthcare system focuses on prevention and early detection of chronic diseases, provide a cheap and smart way to manage and care for patients suffering from age-related chronic diseases, such as heart disease which require continuous, long-term monitoring rather than sporadic assessments. These days, continuous health monitoring system are wearable and easy to use consisting of tiny wireless sensors, strategically placed on the human body, creating a WBAN that monitors vital parameters and provide real-time feedback to the user and medical personnel. When integrated into a telemedical system, these systems can even alert medical personnel about life-threatening changes. In addition, the wearable systems can be used for health monitoring of patients in ambulatory settings [1]. A WBAN consists of multiple sensor nodes, each capable of sampling, processing, and communicating one or more vital signs (heart rate, blood pressure, ECG, EEG, oxygen saturation,).These sensors are placed strategically on the human body as tiny patches or hidden in users' clothes allowing ubiquitous health monitoring for extended periods of time. These sensor nodes sample vital signs and transfer the relevant data to a personal server using ZigBee (802.15.4) or Bluetooth (802.15.1). A personal server sets up and controls

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 483–490.
springerlink.com

the WBAN by transferring the information about status of health to the medical server through the Internet or mobile telephone networks and provides graphical or audio interface to the user. The electronic medical records of registered users are maintained by the medical server which provides various services to the users and medical personnel. It is also responsible for authenticating users, accepting health monitoring session uploads, formatting and insertion of data into corresponding medical records, analyzing the data patterns and recognizing serious health anomalies to help contact emergency services, or forward new instructions to the users. The physician can access the data from his/her office via the Internet examine the reports to ensure the patient is within expected health metrics , ensure that the patient is responding to a given treatment or that a patient has been performing the prescribed exercises. The server agent is allowed to inspect the uploaded data and create an alert in the case of an emergency medical situation.

The sensor nodes used to monitor the vital statistics of the patient operate on batteries. Thus providing long battery life is the most critical parameter to be considered during design. WBAN protocols can be divided in intra- body and inter-body communication. In the former, the information handling between the sensors or actuators and the sink is controlled, in the latter, communication between the sink and an external network is catered to. In this paper investigations are done on the various routing protocol to determine the most suitable routing protocol considering both the energy and quality of service. The next section deals with the ZigBee (802.15.4) protocol. Section 3 deals with the overview of the routing protocols. In section 4 the simulation model and definitions are highlighted. The results are analyzed and the protocol suitable for WBANs is determined in section 5. Finally, we present our conclusion of the study in the last section.

2 Overview of ZigBee

The ZigBee standard which is based on the IEEE 802.15.4 LR-WPAN standard has been proposed to interconnect simple, low rate, and battery powered wireless devices [2].The ZigBee specification establishes the framework for the Network and Application Layers based on the PHY and MAC layers [3]specified by IEEE 802.15.4 WPAN standard [4].The PHY layer defines a total of 27 channels: 16 channels at a maximum rate of 250 kbps in the ISM 2.4 - 2.4835 GHz band, 10 channels at 40 kbps in the ISM 902 - 928 MHz band, and one channel at 20 kbps in the 868.0 - 868.6 MHz band. At the MAC layer beaconless and beaconed modes access the radio channel using Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) or the optional slotted CSMA/CA mechanism. Two device types are specified within the IEEE 802.15.4 framework: full function device (FFD) and reduced function device (RFD). An FFD maintains routing tables, participate in route discovery and repair, maintains beaconing framework, and handle node joins. It also has the capability of communicating with any other devices within its transmission range. An RFD simply maintains the minimum amount of knowledge to stay on the network, and it does not participate in routing. RFDs can only associate and communicate with FFDs. FFDs and RFDs can be interconnected to form star or peer-to-peer networks.

3 Rouing Protocols

The WBANs are similar to MANETS in the sense that they have mobile nodes which need to reorganize themselves. But they differ in the number of nodes (usually only 10-15 nodes are present in WBANs) and the mobility speed. Thus routing protocols similar to the ones used in MANETs can be used. Two nodes communicate directly if they are within the transmission range of each other, else communicate via a multi-hop route.Routing protocols can be divided into proactive and reactive. Proactive routing approach attempts to maintain routing information to all nodes in the network in the form of routing tables even before it is needed. They require periodical update, which causes overhead. In contrast to this, in reactive protocols routes are built as and when required. Two routing schemes are available in ZigBee networks, namely mesh routing and tree routing. The mesh routing scheme is similar to the Adhoc On Demand Vector (AODV) routing algorithm [5], while the tree routing scheme resembles the cluster tree routing algorithm. Here a few of the routing protocols have been considered.

A) ***Ad hoc On-demand Distance Vector (AODV):*** This is a reactive routing algorithm, [5] where intermediate node decides how the routed packet should be forwarded next. AODV is a variant of classical distance vector routing algorithm based on DSDV and DSR. On the one hand DSR's on-demand characteristic discovers the route using route discovery process, on the other hand traditional routing tables with one entry per destination containing three essential fields: a next hop node, a sequence number and a hop count is used. Similar to DSDV, AODV provides loop free routes but in contrast does not require global periodic routing. It allows periodic neighbor detection packets in its routing mechanism. At each node, AODV maintains a routing table. All packets destined to the destination are sent to the next hop node. The sequence number acts as a form of time stamping and is a measure of the freshness of a route. The hop count represents the current distance to the destination node.

B) ***Zone Routing Protocol (ZRP):*** This is a hybrid protocol with the advantages of both the reactive and proactive protocol. Each node proactively maintains route to the destination with a local neighborhood called routing zone. The size of the zone depends on the zone radius.

C) ***Inter Zone Routing Protocol (IERP):*** This is responsible for reactively discovering routes to the destination beyond the routing zone. It is used if destination is not available within the routing zone. The route request packets are transmitted to all the border nodes which again forward the request if destination is not found in the routing zone. It is different from the standard flood search algorithm that it uses broadcasting. Here broadcast Resolution Protocol is used for the packet delivery.

D) ***Dynamic Manet On-demand Routing Protocol (DYMO):***The basic operation of DYMO is Route Discovery and Route Maintenance. The route discovery is responsible for identifying the appropriate route, including Route Request (RREQ) and Route

Reply (RREP). The route maintenance is responsible for maintaining an established route, including Route Error (RERR) the path accumulation function of DYMO includes source routing characteristics, thereby allowing nodes listening to routing messages to acquire knowledge about routes to other nodes without initiating route request discoveries themselves. As a result, this path accumulation function can reduce the routing overhead, although the packet size of the routing packet is increased [6].

4 Simulation Environment

The main goal of this simulation is to analyze the performance of ZigBee using static IEEE 802.15.4 star topology for different existing routing protocols that can be used for Wireless Body Area Networks. Star topology with one PAN coordinator with a network of 15 nodes which are placed randomly. PAN is static mains powered device placed at the centre of the simulation area. The transmission range of devices is one hop away from PAN Coordinator in star topology. The fact that BO (Beacon order) = SO (super frame order) assures that no inactive part of the super frame is present [1]. A low value of this parameter implies a great probability of collisions of beacon frames as they would be transmitted very frequently by coordinators .On the contrary, a high value of the BO (beacon order) introduces a significant delay in the time required to perform the MAC association procedure since channel duration which is a part of association procedure is proportional to BO (beacon order). The table below shows the simulation parameters.

Table 1. Simulation parameters

Routing protocols	AODV,DYMO, IERP & ZRP
Radio type	802.15.4
Channel frequency	2.4GHz
No. Of Channels	One
Path loss model	Two ray
Mobility speeds	None Random Way Point 0 to 5mps
Battery model	Mica Motes
Simulation area	500cm X500cm
Number of nodes	15
Simulation time	150 sec
Simulator	QualNet 5.0.2

Here, we consider the following five metric to determine the suitability if the protocol for Wireless Body Area Networks.

A). ***Packet Delivery Ratio (PDR):*** is the rate of successfully delivering the data packets to the sink. It is denoted as PRD = (D/S) *100, Where D is the number of packets received by the destination and S the number of packets sent by the source node.

B) ***Throughput:*** is the number of bits passed through a network in one second. It measures how fast data can pass through an entity (such as a point or a network). The throughput of a node is measured by counting the total number of data packets successfully received at the node and computing the number of bits received, which is finally divided by the total simulation runtime.

Throughput of a Node = (Total Data Bits Received) / (Simulation Runtime).

The throughput of the network is defined as the average of the throughput of all nodes involved in data transmission. Network Throughput = (Total throughput of nodes involved in data transmission) / (Number of nodes).

C) ***Energy Consumed:*** Energy is consumed in the active state when the nodes either transmit or receive and in the idle mode. Here the total energy consumed is the sum of transmitted and received energy.

D) ***Jitter:*** Jitter refers to a variation in packet delay, resulting in differing packet interarrival times or out-of-sequence packets or both. It is often known as a measure of the variability over time of the packet latency across a network. A network with constant latency has no jitter. Packet jitter is expressed as an average of the deviation from the network mean latency.

E) ***Average End to End Delay****:* indicates the length of time taken for a packet to travel from the CBR (Constant Bit Rate) source to the destination. The average end-to-end delay of a packet depends on delay at each hop comprising of queuing, channel access and transmission delays and route discovery latency.

Packet Delay= (Receive time at destination) – (Transmit time at source) Average Delay= (Sum of all packet delays) / (Total number of packets received)

5 Simulation Results

The performance of the above mentioned algorithm have been extensively studied but previous evaluation studies are mostly IEEE 802.11 centric which consider all participating nodes to be capable of routing. However, under the innate properties of IEEE 802.15.4 and ZigBee networks (i.e. the addressing structure and service assumptions), the performance of ZigBee mesh routing is expected to be different. Figure 1 shows the packet delivery ratio (PDR) of AODV, DYMO, ZRP, IERP with and without mobility. The mobility considered here is low since movement is not at high speeds in WBANS.

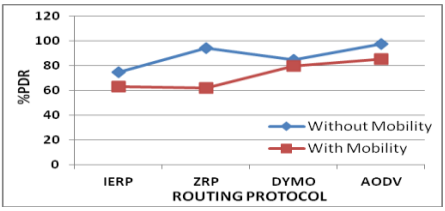


Fig. 1. Comparison of % PDR

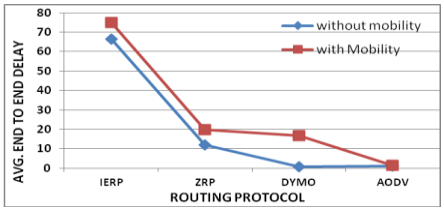


Fig. 2. Comparison of Average end To end delay

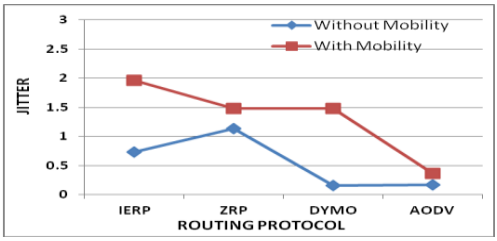


Fig. 3. Comparison of jitter

The average end to end delay which is a very important parameter to be considered in WBANs as critical data is transmitted to the medical server. The figure 2 show that even when there is mobility, AODV protocol provides the best results followed by DYMO. The figure 3 demonstrates that mobility has a lot of effect in routing protocols operating in IEEE 802.15.4.Again AODV proves to be the one which provides minimum jitter, another important parameter in QoS of WBANs.

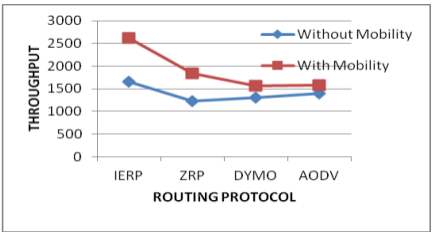


Fig. 4. Comparison of throughput

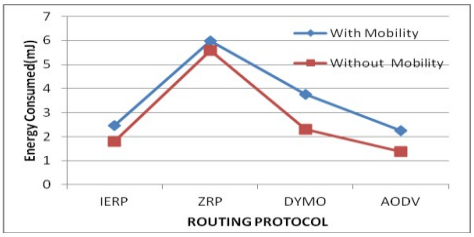


Fig. 5. Energy consumed by various Protocols

The energy consumed is Figure 4 illustrates that throughput IERP is the highest both with and without mobility followed by ZRP minimum in AODV followed by IERP as shown in figure 5.

Table 2 and 3 gives the comparison of the various protocols without and with mobility and based on the reading obtained we can say that AODV is the best protocol nodes operating on ZigBee standard IEEE 802.15.4 for Wireless Body Networks.

Table 2. Comparison of parameters without mobility

ROUTING PROTOCOL	THROUGHPUT	DELAY	JITTER	PDR	TO TAL ENERGY CONSUMED
IERP	*2620*	*6 6 .48*	*0 .73*	*74 .9*	*2 .45*
ZRP	*18 45*	*12 .05*	*1.139*	*9 4 .4*	*5.976*
DYMO	*1561*	*0 .89*	*0 .154*	*85.0*	*3 .76*
AODV	*1583*	*0 .96*	*0 .17*	*9 7.5*	*2 .24*

Table 3. Comparison of parameters with mobility

ROUTING PROTOCOL	THROUGH PUT	DELAY	JITTER	PDR	TOTAL ENERGY CONSUMED
IERP	*1655*	*74 .89*	*1.9 61*	*6 3 .0 8*	*1.791*
ZR P	*1221*	*19 .95*	*1.4 78*	*6 2 .4 1*	*5.57*
DYMO	*1301*	*16 .72*	*1.4 78*	*8 0 .13*	*2 .29*
AODV	*1390*	*1.47*	*0 .3671*	*8 5.4 5*	*1.378*

6 Conclusion

WBANs play an important role in the deployment of wearable/ mobile pervasive computing systems. The performance evaluation of AODV IERP DYMO and ZRP routing protocols for stationary and mobile nodes are done by using CBR application in ZigBee network having static IEEE 802.15.4 star topology using QualNet 5.0.2 network simulator. From the results it can be observed that routing protocol AODV is suited for applications where like WBAN, where limited energy resources are available, making it impossible to recharge or replace the batteries. Energy performance is analyzed and it is observed that AODV performs better than

IERP, ZRP and DYMO. ZRP, IERP and DYMO being the protocols which need sufficient time to establish route discovery and route maintenance, for large range mobile applications they are best suited, where traffic is random and sporadic between several nodes rather than being almost exclusively between a small specified set of nodes.

References

[1] Istepanian, J., Zhang, Y.T.: Guest Editorial Introduction to the Special Section on M-Health: Beyond Seamless Mobility and Global Wireless Health-Care Connectivity. IEEE Transactions on Information Technology in Biomedicine 8(4), 405–414 (2004)

[2] Kaushik, P., Bhatia, A.: A cluster based minimum battery cost AODV routing using multi-path route for Zigbee. Networks, 1–72 (2008) ISSN: 1556-6463

[3] Mohanty, S.: Energy Efficient Routing Algorithms for Wireless Sensor Networks and Performance Evaluation of Quality of Service for IEEE 802.15.4 Networks, NIT Rourkela (2010)

[4] IEEE 802.15.4: Mac and physical specifications for lR-PANS (2003), http://www.ieee802.org/15/pub/TG4.html

[5] Das, S., Perkins, C.E., Royer, E.M.: Ad Hoc On demand Distance Vector Routing (AODV). IETF RFC 3561 (July 2008)

[6] Tiwari, S., Raghuvanshi, A.S.: DYMO as routing protocol for IEEE-802.15.4 enabled Wireless Sensor Networks. IEEE Wireless Communication and Sensor Networks (WCSN), 1–6 (2010)

Effect of Idle Mode on Power Saving in Mobile WiMAX Network

Thontadharya H.J., Shwetha D., Subramanya Bhat M., and Devaraju J.T.

Dept. of Electronic Science, Jnana Bharathi, Bangalore University, Bangalore
{thontadharya,devarajujt}@bub.ernet.in

Abstract. IEEE 802.16e is an emerging standard for mobile wireless broadband access systems. In any mobile networks, power saving is one of the most important features for the extension of devices' lifetime. To manage power usage in a more efficient way, the IEEE 802.16e standard specifies two mechanisms, sleep mode and idle mode. Idle mode allows the mobile station (MS) to conserve power and resources by restricting its activity to scanning at discrete intervals and thus eliminates the active requirement for handover operation and other normal operations. On the base station (BS) and network side, idle mode provides a simple and timely method for alerting the MS for pending downlink (DL) traffic directed to the MS and thus eliminates air interface and network handover traffic from essentially inactive MSs. An attempt made in this paper to evaluate the performance of idle mode in terms of power saving in MSs for long battery life.

1 Introduction

For the past few years, the mobile hand-held devices including cellular phones have become very popular. Currently, to provide both voice and high-bandwidth data services, new systems are being developed. Originally, IEEE 802.16 [1] has been designed for fixed subscriber stations (SSs). On the other hand, the recently developed IEEE 802.16e [2] standard is an extension targeting at the service provisioning to the Mobile Subscriber Stations (MSs). Mobile Worldwide Interoperability for Microwave Access (WiMAX) based on IEEE 802.16e standard enables high speed data communications anywhere and anytime. In any mobile networks, power saving is one of the most important and crucial features for the handheld mobile devices' operating lifetime.

To support battery-operated portable devices, mobile WiMAX has power saving features that allow portable subscriber stations to operate for longer durations without having recharge. Power saving is achieved by turning off the power parts of the MS in a controlled manner when it is not actively transmitting or receiving data. The standard IEEE 802.16e [2] defines two new power saving modes for the MSs, *viz.*, the sleep mode and the idle mode in order to have power efficient MS operation and a more efficient handover. Mobile WiMAX defines signalling methods that allow the MS to retreat into a sleep mode or idle mode when inactive. Sleep mode is a state in which the MS effectively turns itself off and becomes unavailable for predetermined periods. The periods of absence are negotiated with the serving Base Station (BS).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 491–499.
springerlink.com

Idle mode allows even greater power savings and support for it is optional in WiMAX.

Idle mode allows the MS to completely turn off and not to be registered with any BS and yet receive downlink (DL) broadcast traffic. When DL traffic arrives for the idle mode MS, the MS is paged by a collection of BSs that form a paging group (PG). The MS is assigned to a PG by the BS before going into idle mode and the MS periodically wakes up to update its PG. Idle mode saves more power than sleep mode, since the MS does not even have to register or do handoffs. Idle mode also benefits the network and BS by eliminating handover traffic from inactive MSs.

In this paper an attempt has been made to evaluate the performance of idle mode in terms of power saving in MSs. The rest of the paper is organized as follows: Section 2 explains the idle mode and paging operation in mobile WiMAX. Section 3 contains the results of simulation and discussion followed by conclusion in Section 4.

2 Overview of Idle Mode and Paging Operation in Mobile WiMAX Networks

Idle mode is intended as a mechanism to allow the MS to become periodically available for DL broadcast traffic messaging without registration at a specific BS as the MS traverses an air link environment populated by multiple BSs, typically over a large geographic area. Idle mode benefits MS by removing the active requirement for hand over (HO) and all normal operations. By restricting MS activity to scanning at discrete intervals, idle mode allows the MS to conserve power and operational resources. For idle mode operation, the BSs are divided into logical paging groups called PGs. The purpose of these groups is to offer a contiguous coverage region in which the MS does not need to transmit in the uplink (UL), yet can be paged in the DL if there is traffic targeted at it. The PGs should be large enough so that most MSs will remain within the same PG most of the time and small enough so that the paging overhead is reasonable [3].

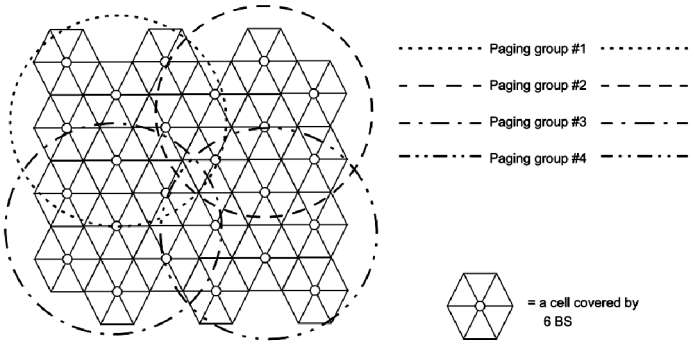


Fig. 1. Paging Groups

Figure 1 shows an example of four PGs defined over multiple BS arranged in a hexagonal grid. A BS may be a member of one or more PGs comprised of differing groupings of BS, of varying cycles and offsets. This provides support for geographic requirements of idle mode operation along with differentiated and dynamic quality of service (QoS) requirements and scalable load-balancing distribution. Upon entering idle mode, the MS relinquishes all of its connections and states associated with the BS it was last registered with. The idle MS is tracked by the network at the granularity of a group of BSs (*i.e.*, PG) as opposed to a non-idle MS which is tracked at the granularity of a BS. While in idle mode the MS periodically listens to the radio transmissions for paging messages, in a deterministic fashion that is decided a priori between the network and itself. The period for which the MS listens to paging messages is known as "paging listen interval" (PLI) and the period for which the MS powers off its radio interface is known as the "paging unavailable interval" (PUI). The operation of idle mode and paging, in mobile WiMAX networks, is summarized in following section.

2.1 Maintaining the Location Information of an Idle-Mode MS

The location information of an idle MS is achieved by logically dividing the network coverage area into different PGs. A PG refers to the coverage area of one or more base stations (BSs). A Paging Controller (PC) administers one or more PGs. There could be one or more PCs in the network. When an MS goes to idle mode, a PC, referred to as anchor PC, creates an entry in its database noting the PG where the MS is initially located. When the MS moves from one PG to another, it updates the location with the anchor PC. Therefore, while in idle mode the location of an MS is known up to the granularity of one PG.

2.2 Paging an Idle Mode MS

When the network wants to locate an idle mode MS, or has incoming data buffered for it, or for administrative purposes, the PC initiates paging the MS by broadcasting mobile paging advertisement (MOB-PAG-ADV) message to all the BSs in the PG; the BSs in turn broadcast this message on the airlink. This is because when the location information stored at a PC is correct; the MS is expected to reside in the coverage area of at least one of these BSs. If the paging advertisement happens during the PLI of the MS, it is expected to receive the page and perform network re-entry or location update in response to the page.

2.3 Paging Architecture

Figure 2 depicts a representative network reference model [4] used to describe the idle mode operation in WiMAX networks. It consists of the three PGs (PG1, PG2, and PG3) and two PCs (PC1 and PC2). PC1 manages PG1 and PG2, PC2 manages PG3. PG1 comprises three BSs, PG2 comprises one BS and PG3 comprises two BSs. Each PC maintains a location database that keeps information about all the MSs that have gone into idle mode in the PG(s) managed by that PC.

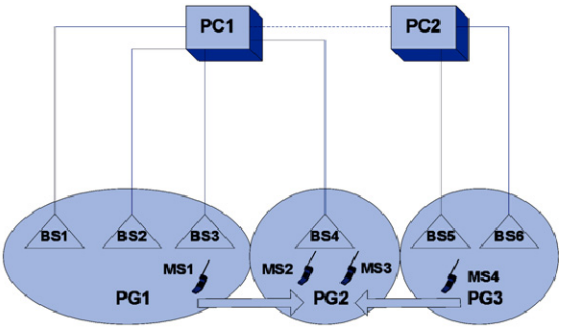


Fig. 2. Network Reference Model

While only four idle mode MSs are shown in the figure 2, there may be several more MSs (both idle mode and active mode) in real deployments in BS4 coverage area. The amount of power saving that can be achieved by MS in idle mode is tightly coupled to the duty cycle of the MS, which is the ratio of paging listen to the paging unavailable interval. One paging unavailable interval and one paging listening interval constitute a paging cycle as shown in Figure 3. Therefore, once in every paging cycle interval the idle mode MS wakes up and listens for paging messages. When traffic arrives for the idle mode MS the network performs paging to locate the MS and to bring it back to active mode [5].There are three main parameters in idle mode operation, *viz.*, PG identifier, paging cycle, and paging offset. They are determined at the initiation by exchanging the messages (DREG-REQ/RSP) as shown in Figure 3. PG identifier is shared by every member BS and included in every paging message (MOB-PAG-ADV) to inform the idle MSs of the PG that they are located. Two remaining parameters, paging cycle and paging offset, are used for determining the starting point of each paging interval. They are also shared by every member BS so that MSs are able to receive the paging messages from any BS in the PG.

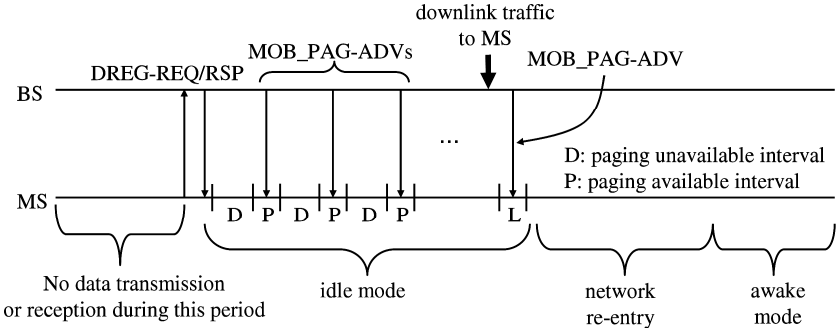


Fig. 3. Idle mode operation

The entire idle mode operation in IEEE 802.16e based mobile WiMAX networks can be divided into following stages: idle mode initiation, idle mode entry, operation during idle mode and idle mode exit.

2.3.1 Idle Mode Initiation

The idle mode can be initiated either by the BS or MS when the MS does not have any ongoing traffic. In case of MS initiated idle mode, the MS sends a deregistration request (DREG-REQ) message to the BS [1]. Similarly, in the case of a BS-initiated idle mode, the BS sends a deregistration command (DREG-CMD) message to the MS. When the MS receives the DREG-CMD message, it sends the DREG-REQ message to the BS [1]. In each case, the BS receives a DREG-REQ message from the MS.

2.3.2 Idle Mode Entry

When a BS receives the DREG-REQ message from one of its MSs, it sends a message to the anchor PC. This message contains certain MS service and operation information referred to as idle mode retain information (IMRI). IMRI can be used to expedite the MS's network re-entry from idle mode. PC stores the MS IMRI and transmits a backbone message to BS that includes numerical values for PAGING CYCLE, PAGING OFFSET, and MS Paging Listening Interval (PLI) for the MS [1]. Note that PC may use its own algorithm or negotiate with the BS and/or MS to decide the numerical values of PAGING CYCLE and PLI. On the other hand, it can determine the PAGING OFFSET using its own algorithm. Once BS receives the backbone message from the PC, it sends the DREG-CMD message to the MS that includes the idle mode entry time (IMET), PAGING CYCLE, PAGING OFFSET, and PLI values. The MS enters into idle mode at IMET.

2.3.3 Idle Mode Operation

While in idle mode the MS alternates between PUI and PLI. The idle mode operation of two MSs (MS1 and MS2) is illustrated in Figure 4. In this case both MS1 and MS2 have the same PAGINGCYCLE and PLI, which is the case in most network deployments. However, MS1 and MS2 have different PAGING OFFSETs of T1 and T2 respectively. Therefore, when the network wants to page MS1 and MS2, it does so through two different MOB-PAG-ADV messages at different times. It may be noted that the network needs to send two different MOB-PAG-ADV messages although it wants to page these two idle mode MSs at the same time because the MSs have non-overlapping paging listening intervals.

2.3.4 Idle Mode Exit

An MS in idle mode exits from idle mode if it has data to send to the BS or if there is downlink traffic addressed to a MS, every member BS in the PG pages the MS at the very next paging interval. Unlike sleep mode, a MS cannot be registered to any BS. Therefore, a few BS-specific parameters used at the entering of idle mode will not be valid any more if the MS's current attachment BS has been changed. Therefore, when a MS terminates idle mode, it has to always perform the process called network

re-entry to obtain, negotiate, adjust, and update the BS-specific parameters. At this juncture, the MS terminates idle mode operation and carries out network re-entry procedures as specified in IEEE 802.16e [1]. As a part of network re-entry the idle mode MS may perform contention based initial ranging. In another instance the BS may assign dedicated ranging region to the MS for initial ranging.

2.3.5 Location Update (LU)

A MS in idle mode may travel outside the current PG. It is known to MS by either missing the paging message at the expected paging interval due to the changed paging cycle and offset or the PG identifier in the paging message if it happens to receive the paging message. In such a case, the MS is needed to update the values of PG identifier, paging cycle, and paging offset, which is referred to as location update (LU) process. After the location update process, idle mode continues. There are two kinds of location updates, *viz.*, secure and unsecure location updates. The secure location update process may be simpler than the network re-entry process while the unsecure location update requires the same procedures as the network re-entry process. LU may be triggered by a timer. Even if there is no change in PG, an idle MS has to perform LU before the timer is expired.

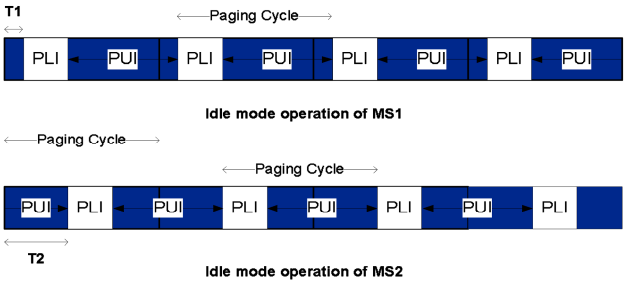


Fig. 4. Comparison of Idle mode operation among MSs

3 Simulation and Results

In this paper the effect of idle mode on the battery performance of MSs in mobile WiMAX network is studied using QualNet 5.0.2 simulator [6]. The scenario designed for this simulation study consists of three adjacent WiMAX cells working at the frequency of 2.4GHz. To assess the idle mode performance of MSs, the number of MSs in one of the WiMAX cells is varied from one to ten. As the idle mode effectively saves power while MS is in handover, mobility is given to MSs in such a way that they traverse through all the cells causing handover.

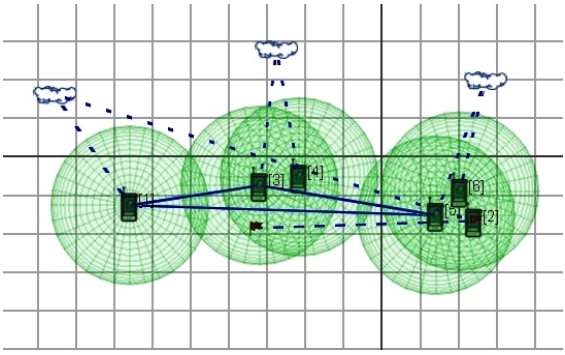


Fig. 5. Snapshot of the simulated scenario

The representative snapshot of the scenario consisting of three WiMAX cells with one BS and one MS each is shown in figure 5. The performance study is carried out by considering the MSs with enabled and disabled idle mode. The performance metrics considered are total charge consumed, energy consumed in transmit mode and energy consumed in receive mode.

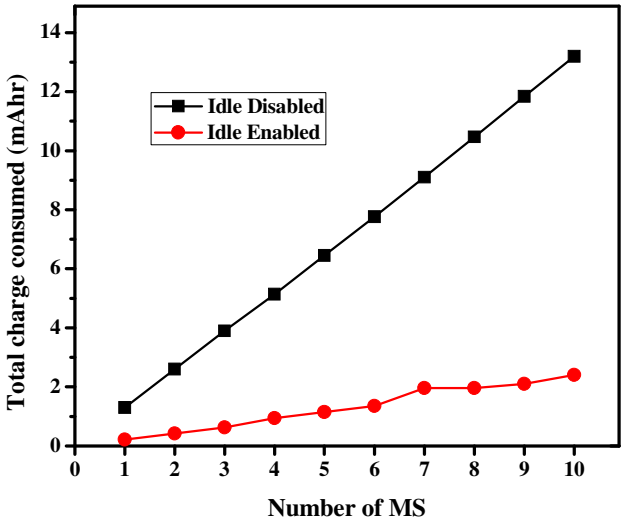


Fig. 6. Total charge consumed for varying number of MSs

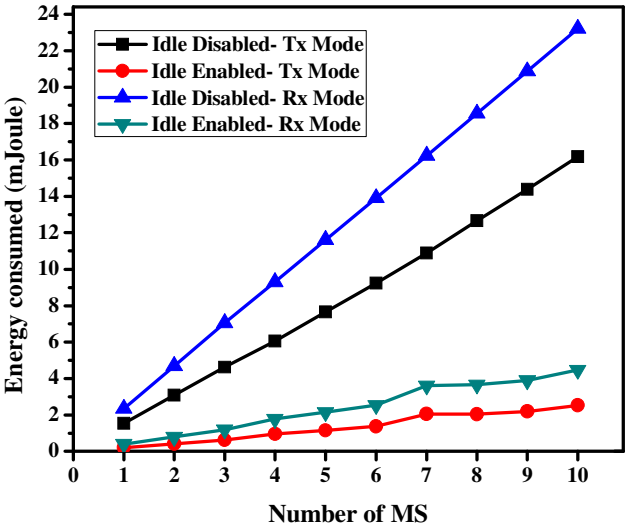


Fig. 7. Energy consumed in transmit and receive mode for varying number of MSs

Figure 6 and 7 show the total charge consumed, energy consumed in transmit & receive mode respectively for varying number of MSs. From figure 6 it is observed that as the number of MSs increases, the energy consumed is also increases for both idle enabled and idle disabled modes. It is apparent that the energy consumed by the idle enabled MSs is less compared to that of idle disabled MS. In idle mode, the MS has no connection to any BSs, does not transmit any management messages until the wakeup process, which allows higher power savings for longer operating life. Figure 7 depicts that as the number of MSs increases, the energy consumed by transmit & receive mode increases for both idle enabled and idle disabled modes.

4 Conclusions

In mobile wireless access networks, battery life and handoff are essential criteria for mobile applications. Hence mobile WiMAX supports power saving modes (sleep and idle modes) to extend battery life of mobile devices. This paper has analyzed the power saving efficiency of idle mode specified in IEEE 802.16e standard. From the results, it is evident that enabling of idle mode increases battery lifetime and thus the power performance. The study reveals that the energy consumed by the idle enabled MS is very less compared to that of idle disabled MS and hence idle mode offers higher power savings.

Acknowledgments. The authors would like to thank UGC for sanctioning the funds under major research project. Authors would also thank Nihon communication, Bangalore, for the simulation tool and support.

References

1. IEEE Standard 802.16e-2005 Amendment to IEEE Standard for Local and Metropolitan Area Networks–Part16: Air Interface for Fixed Broadband Wireless Access Systems-Physical and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands
2. IEEE Standard 802.16-2009 for Local and metropolitan area networks-Part 16: Air Interface for Broadband Wireless Access Systems
3. Nuaymi, L.: WiMAX: Technology for Broadband Wireless Access. John Wiley & Sons Ltd. (2007) ISBN: 0-470-02808-4
4. Mohanty, S., Venkatachalam, M., Yang, X.: A Novel Algorithm for Efficient Paging in Mobile WiMAX. In: Proceedings of IEEE Mobile WiMAX Symposium, Orlando (2007)
5. Kim, B., Park, J., Choi, Y.-H.: Power Saving Mechanisms of IEEE 802.16e: Sleep Mode vs. Idle Mode. In: Min, G., Di Martino, B., Yang, L.T., Guo, M., Rünger, G. (eds.) ISPA Workshops 2006. LNCS, vol. 4331, pp. 332–340. Springer, Heidelberg (2006)
6. Qualnet documentation, `http://www.scalablenetworks.com`

High Speed Programmable Digital Telemetry Filter for Flight Test

Navitha M.V.[1], M.Z. Kurian[2], G. Koteswara Rao[3], and Umashankar B.[3]

[1] Digital Electronics,
Sri Siddhartha University
Sri Siddhartha Institute of Technology,
Tumkur, Karnataka, India
navithamv@gmail.com
[2] Department of Electronics & Communication
Sri Siddhartha Institute of Technology,
Tumkur, Karnataka, India
mzkurianvc@yahoo.com
[3] Aeronautical Development Agency,
Bangalore, Karnataka, India
{gkrao,umashankar}@jetmail.ada.gov.in

Abstract. Digital telemetry filter is an essential subsystem of ground instrumentation which is a part of flight test instrumentation used for military aircraft flight testing. Flight test instrumentation includes on-board instrumentation and ground instrumentation. Typical waveform used in telemetry system for flight testing, which is hybrid in nature (PCM+FM+FM/FM), is multiplexed in frequency domain. Digital telemetry filter plays a critical role to separate data (PCM), hot mike (FM) and vibration (FM/FM) from the received baseband signal during flight test in real time. This paper describes a pipelined approach for the implementation of high speed digital telemetry filter for flight test on low power field programmable gate arrays (FPGA). This paper also provides a brief discourse on the effective application of VLSI design methodologies for efficient implementation of digital filter algorithms. The filter is designed in VHDL, simulated using ModelSim, synthesized using Quartus-II and the implemented on Cyclone-II FPGA.

Keywords: Flight test, Digital telemetry filter, Test bench, PLL, FPGA.

1 Introduction

Filtering is a linear operation. It is also used to remove the unwanted signals. Flight test instrumentation includes on-board instrumentation and ground instrumentation. The ground instrumentation consists of antennae, receivers, filters, demodulator, demultiplexer and further processing units. Digital filters can be programmed and thus can be used either as band pass or low pass or band reject or high pass filters depending on the user requirements. By the use of digital components the errors arising due to component drift can be eliminated.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 501–509.
springerlink.com

The advantage of the FPGA approach to the filter implementation include high sampling rate, superior performance than available with the traditional approaches, more flexible and low cost than an ASIC for moderate volume applications. Moreover recent advancements in field programmable gate arrays (FPGA) design technology has resulted in FPGA becoming the preferred platform for evaluating and implementing the digital filter algorithms.

The block diagram for the implementation of the digital telemetry filter is shown below. The multiplexed baseband signals (PCM+FM+FM/FM) are given to the analog to digital converter (ADC). The signed 2's complement 14-bit digital output is obtained and used as an input to the filter algorithm which is running on FPGA. The 14 bit data processed by filter algorithm is fed to the high speed DAC to reconstruct the filtered analog signal. This output obtained is converted back to the analog form by a digital to analog converter (DAC). The clock for the ADC and DAC is obtained from the PLL in the FPGA. Finally PCM, FM and FM/FM are obtained separately as filter outputs.

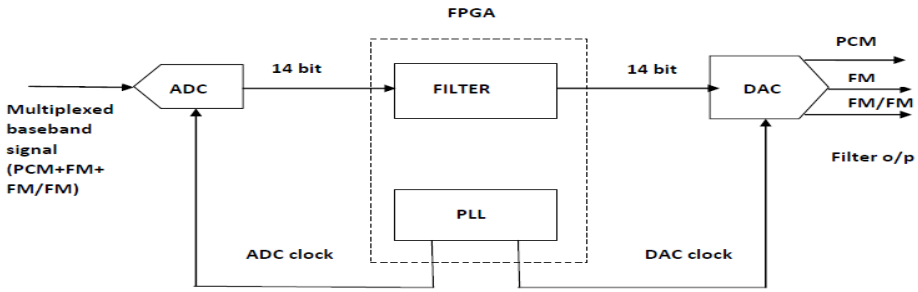


Fig. 1. Block Diagram of Filter Implementation

Initially the 4th order Butterworth filter is modeled and designed using FDA (Filter Design & Analysis) tool in MATLAB. VHSIC Hardware Description Language (VHDL) code for the design is obtained using Filter Design HDL Coder in FDA tool. Fixed point arithmetic is used to implement the IIR Butterworth filter algorithm. Simulation of the designed filter algorithm is carried using ModelSim SE simulator and then synthesis& timing analysis of the filter is carried out by using Quartus–II software. The hardware realized using the Quartus-II software has programmed in to Altera Cyclone-II FPGA for performance analysis during flight test in real time.

2 Background

A. Telemetry

Telemetry is the process by which an object's characteristics are measured (such as velocity of an aircraft), and the results transmitted to a distant station where they are displayed, recorded, and analyzed. The transmission media may be air and space for satellite applications, or copper wire and fiber cable for static ground environments like power generating plants. The purpose of a telemetry system is to collect data at a

place that is remote or inconvenient and to relay the data to a point where the data may be evaluated [11] [3]. The flight test telemetry system for which the filter has implemented is composed of On-board telemetry, Waveform or transmission channel and ground telemetry.

B. Ground Station Instrumentation

The ground station instrumentation consists of antennae, receivers, filters, demodulator, demultiplexer and further processing units. The multi channel filter described in this paper separates out the PCM, hot mike, vibration data from the incoming base band signal in real time. The filtered PCM data is fed to a bit synchronizer for further processing. The frequency modulated hot mike output signal is fed to an audio demodulator to get on-board audio. The FM/FM vibration signal is fed to a vibration digital frequency demultiplexer (DFD) to get the vibration data [3].

C. IIR Filter

2^{nd} order IIR filter is sometimes referred to as a 'bi-quad'. The biquad filter is an implementation of an infinite impulse response (IIR) filter with two poles and two zeros. The output signal from the filter can be non-zero infinitely after the input signal is changed from non-zero to zero. IIR filters have one or more nonzero feedback coefficients [12].

The biquad filter core can be used to implement low pass filters, band pass filters, high pass filters, or band reject filters. The design of a particular set of filter coefficients is generally done using analog filter design techniques. The poles and zeros of the resulting analog filters are then mapped over to the discrete time domain using the bilinear transformation. The difference equation of the biquad filter core is given below

$$a_0y[n] = b_0*x[n] + b_1*x[n-1] + b_2*x[n-2] + a_1*y[n-1] + a_2*y[n-2] \quad (1)$$

The above equation can be split into two equations. The second order difference equation is defined below.

$$w(n) = a_0x(n) -- a_1w(n-1) -- a_2w(n-2). \quad (2)$$

$$y(n) = b_0w(n) + b_1w(n-1) + b_2w(n-2). \quad (3)$$

where y(n) is the current filter output, the y(n-i)'s are previous filter outputs, the x(n-i)'s are current or previous filter inputs, the a_i's are the filter's feed forward coefficients corresponding to the zeros of the filter, the b_i's are the filter's feedback coefficients corresponding to the poles of the filter, and N is the filter's order.

The biquad filter core can be put in series with additional biquad filter cores to implement filters with more than two poles and zeros. The basic biquad can be extended so as to provide better attenuation. It involves having more than one biquad cascaded. For higher order filters, several biquad are cascaded. The cascade structure for 4^{th} order IIR is shown in the following figure.

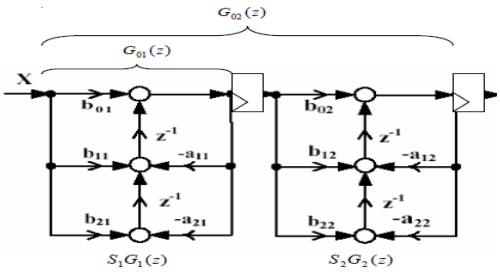


Fig. 2. Direct form-II for 4th order filter

D. Pipelined filter Sections

Pipelining and parallel processing can be combined for IIR filters to achieve a speedup in sample rate by a factor L×M, where L denotes the levels of block processing and M denotes stages of pipelining, or to achieve power reduction at the same speed. Pipelining leads to a reduction in the critical path by introducing pipelining latches along the critical data path and either increases the clock speed (or sampling speed) or reduces the power consumption at same speed in a DSP system. Parallel Processing increases the sampling rate by replicating hardware so that several inputs can be processed in parallel and several outputs can be produced at the same time.

Pipelined architecture is implemented in the proposed filter design to obtain above mentioned advantages. Direct from-II second order sections are made them work independently in this design. So each second order section will be executed independently. Pipeline registers are incorporated in the filter algorithm. Latency of this pipelined architecture is proportional to the number of second order sections. For 4^{th} order filter latency is 2 clock cycles.

E. PLL

A PLL (Phase Locked Loop) is a frequency-control system that generates an output clock by synchronizing itself to an input clock. The main blocks of the PLL are the phase frequency detector (PFD), charge pump, loop filter, voltage controlled oscillator (VCO), and counters, such as a feedback counter (M), a pre-scale counter (N), and post-scale counters(C).

The Quartus-II software provides the ALTPLL Mega Wizard interface to specify the PLL circuitry in the supported devices. The Mega Wizard Plug-In Manager configures the ALTPLL MegaWizard interface and builds ALTPLL mega functions efficiently. The ALTERA_PLL mega function can generate as many as 18 clock output signals. The generated clock output signals clock the core or external blocks outside the core. We can use the reset signal to reset the output clock value to 0 and disable the PLL output clocks. Each output clock has a set of requested settings where you can specify the value of output frequency, phase shift, and duty cycle.

3 System Design

Digital filtering is a numerical procedure or algorithm that transforms a given sequence of numbers into a second sequence that has some more desirable properties. The most straightforward way to implement a digital filter is by convolving the input signal with the digital filter's impulse response.

The IIR butter worth filter algorithm is modeled and designed using Filter Design and Analysis tool of MATLAB. Frequency response, phase response, impulse response and step response of the proposed algorithm is theoretically analyzed, and then the filter coefficients required for the suitable filter design are obtained using FDA tool.

The IIR filter algorithms for low pass, high pass, band pass and band reject are realized using VHDL. Simulation is done using Mentor Graphics ModelSim simulator to check impulse, step and ramp response of the realized filter algorithms. VHDL test benches are used to simulate the filter codes.

A. Digital Filter Modeling & Design

For design of digital filter for different configurations like low pass, high pass, band pass and band reject filter MATLAB is to be used. A fourth order butter worth band pass filter is modeled, designed and analyzed using FDA Tool. FDA Tool enables to design digital FIR or IIR filters by setting filter specifications, by importing filters from your MATLAB workspace etc. Tool also provides tools for analyzing filters, such as magnitude and phase response and pole-zero plots.

For designing the filter using FDA tool the following specifications are provided: Type of filter, Technique, Order specified, Sampling frequency, Cut off frequency. The frequency response and phase response is obtained by the tool along with the filter coefficients. The result obtained for the hot mike signal using band pass filter is shown below.

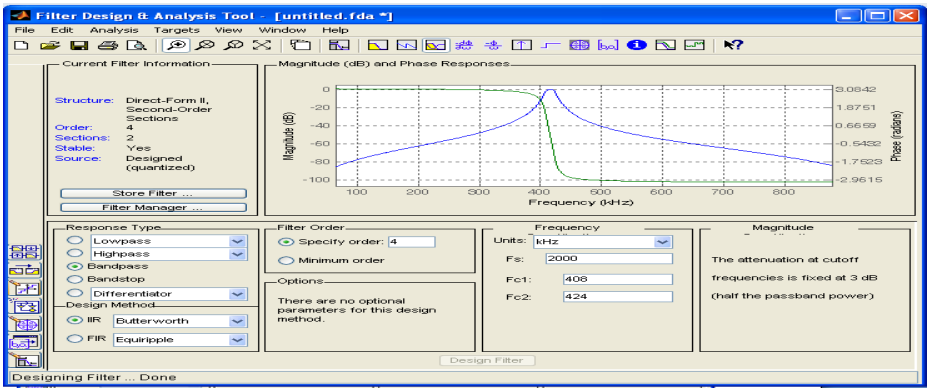


Fig. 3. Band Pass filter for hot mike using FDA Tool

B. Simulation

The 4th order IIR butter worth filter is designed using VHDL. ModelSim is used to simulate the designed filter algorithm. ModelSim is an IDE for hardware design which provides behavioral simulation of a number of languages, i.e., Verilog, VHDL, and SystemC. ModelSim requires the design under test which is filter code and the stimulus data which can be fed to the simulator using Test benches.

Separate Test benches are generated using VHDL to simulate the impulse, step, ramp and chirp response. Simulations are carried out using these test benches and compared the results of the filter with the expected results. The simulation results for impulse and step responses are given in fig 4 and fig 5 respectively. The impulse response the filter output must finally go to for the values near to zero. In case of step response the output should stabilize to the input values after few clock pulses.

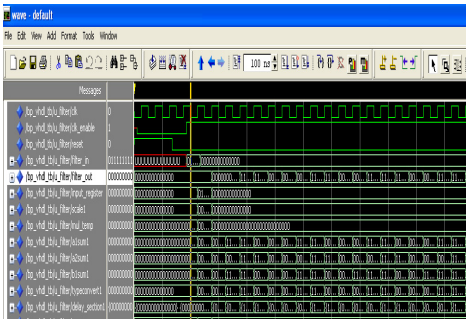


Fig. 4. Impulse Response output

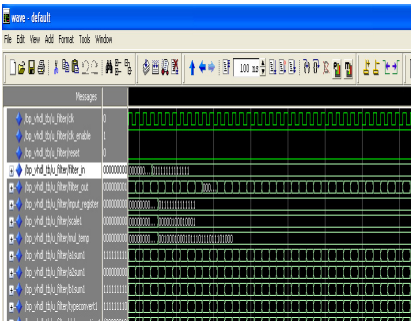


Fig. 5. Step Response output

Three processes are used to realize the biquad filter algorithm. In the first process input values are sampled from ADC and stored in process registers at rising edge of the system clock. The second process is used to store the delay elements. Third process statement is used to obtain the output of the filter algorithm and given it to the DAC.

4 System Implemetation

Synthesis of the fully simulated filter algorithm is carried out by using Altera Quartus-II software to realize the hardware which is compatible to implement on proposed FPGA core. Finally the hardware realized is programmed into the Altera Quartus-II FPGA for real time performance analysis.

The multiplexed baseband analog signal is sampled at required rate in real time and converted into 14 bit digital signed data by using Analog to Digital Converter AD9248. Sampled digital data is given to FPGA filter block to process, and then processed 14 bit data is fed to the Digital to Analog Converter AD9767 to get analog signal for further processing.

The clocks generated by using the ALTPLL Mega Wizard interface are used to clock the ADC and DAC blocks located outside the FPGA core. The input frequency to the PLL block is 24MHz which is fed from the oscillator of the FPGA board to derive the required clocks for ADC & DAC blocks. All derived clocks from PLL are synchronized to the input clock.

A. Synthesis & Timing Analysis

The Altera Quartus- II design software provides a complete, multiplatform design environment that easily adapts to our specific design needs. The synthesis of the filter algorithm is done using Quartus-II software. The desired circuit is specified either by using a hardware description language VHDL, or by means of a schematic diagram. Then synthesis is carried out that gives the logic elements (LE) required. Next stage is the functional simulation followed by fitting. Timing analysis is carried out in the next phase. Finally the designed circuit is implemented in a physical FPGA chip by programming the configuration switches that configure the LEs and establish the required wiring connections.

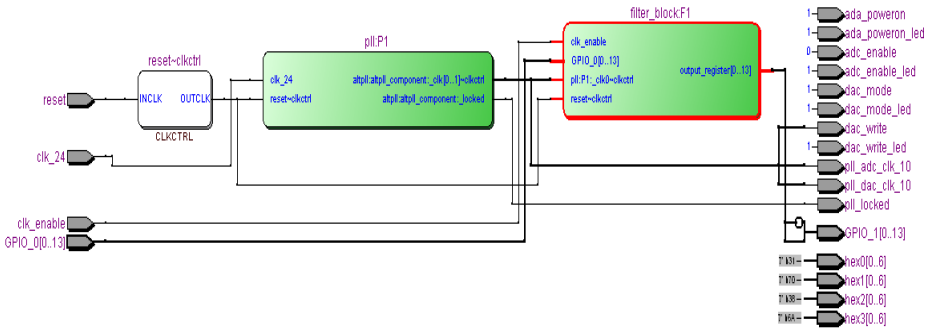


Fig. 6. Quartus-II output of filter algorithm

Place & Route of the realized hardware from synthesis is carried out on the proposed FPGA core. Timing analysis is performed by using Quartus-II Time-Quest Timing Analyzer. All timing constraints defined in SDC file are met. The programmer object file (.pof) is generated for the filter to configure the FPGA. The hardware realized by Quartus-II software is given above.

B. FPGA Implementation

The last stage in the implementation of digital programmable telemetry filter is to load Programmer Object File (.pof) into configuration device to configure the Cyclone-II FPGA. Altera® Cyclone II FPGAs extend the low-cost FPGA density range to 68,416 logic elements (LEs) and provide up to 622 usable I/O pins and up to 1.1 Mbits of embedded memory. Cyclone II FPGAs are manufactured on 300-mm wafers using TSMC's 90-nm low-k dielectric process to ensure rapid availability and low cost. Altera's latest generation of low-cost FPGAs—Cyclone II FPGA's offer 60% higher performances and half the power consumption of competing 90-nm FPGAs.

Finally the frequency response of the filter from MATLAB FDA Tool is compared with the filter output spectrum during Real Time and thus performance of the Digital Telemetry Filter implemented on FPGA is analyzed during flight of the Light Combat Aircraft (LCA). The frequency response of the band pass filter by FDA Tool is given in fig 7 and the real time spectrum of filter output is given in fig 8.

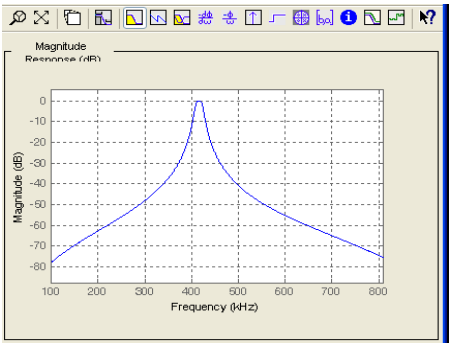


Fig. 7. Frequency response of BP filter by FDA Tool

Fig. 8. Real time spectrum of BP filter output during flight

5 Conclusions

Filtering is a critical aspect of the overall flight test instrumentation process. The design of filters is not a trivial job, particularly when the accuracy and reliability requirements are of a high level typical of those found in a flight test instrumentation system. In our paper we propose a high speed digital programmable telemetry filter which is evolved from low power VLSI technologies. Thus there would not be a need of separate filters for high pass, low pass, band pass and band reject filter. Thus a pipelined approach can be used for the implementation of high speed telemetry filter on low power FPGA. Thus this project will serve as a milestone in the field of aeronautics.

References

[1] Bruce Owens, D., Brandon, J.M., Croom, M.A., Fremaux, C.M., Heim, E.H., Vicroy, D.D.: Overview of Dynamic Test Techniques for Flight Dynamics Research at NASA LaRC. American Institute of Aeronautics and Astronautics, NASA Langley Research Center, Hampton, VA, 23681

[2] Borek, R.W., Pool, A.: Basic Principles of Flight Test Instrumentation Engineering. AGARDograph 160 Flight Test Instrumentation Series, vol. I(2)

[3] Carden, F., Jedlicka, R., Henry, R.: Telemetry Systems Engineering. Library of Congress Cataloging-in-Publication Data

[4] Pool, A., Bosman, D.: Basic Principles of Flight Test Instrumentation Engineering. AGARDograph 160

[5] Telemetry Tutorial, L-3 Communications Telemetry West, ML1800 Rev. A
[6] Punskaya, E.: Basics of Digital Filters (unpublished)
[7] Jackson, B.A.: Digital Filter Design And Synthesis Using High-Level Modeling Tools (in press)
[8] Chen, C., Li, B., Wang, C.: Chebyshev I Bandpass IIR Filter with 6th Order. Script of Digital Systems (2003) (unpublished)
[9] IIR_SOS IIR filter Second-Order-Section (2012), http://www.zipcores.com
[10] Pedroni, V.A.: Circuit Design Using VHDL (text book)
[11] Ghosh, M.: Design and Implementation of Different Multipliers Using Vhdl. National Institute of Technology, Rourkela (2007)
[12] An Introduction to Digital Filters, Application Note, Intersil (January 1999)
[13] Using the Analog Devices Active Filter Design Tool
[14] Infinite Impulse Response Filter Structures in Xilinx FPGAs, WP330 (v1.2), August 10 (2009)
[15] Using the Synopsys Design Constraints Format, Application Note Version 2003.6 (June 2003)
[16] Parhi, K.K.: Pipelined and Parallel Recursive and Adaptive Filters

Hierarchical Storage Technique for Maintaining Hop-Count to Prevent DDoS Attack in Cloud Computing

Vikas Chouhan and Sateesh Kumar Peddoju

Electronics & Computer Engineering Department,
IIT Roorkee, Roorkee, Uttarakhand, India
vikaschouhan.iitr@gmail.com
sateesh@ieee.org

Abstract. In cloud environment, cloud servers providing requested services to the client as per request, sometimes may crash due to denial of service (DoS) attack. It prevents the legitimate users from getting service. DoS attack is accompanied by IP Spoofing so as to hide the location of flooding and to make every request dissimilar. Hop-count value is helps in Preventing DOS attack in cloud environment. This value is determined from received IP Packet. So, there is essential requirement of storage for storing hop-count value of clients. In this paper, we present an approach for storing hop-count value which helps to prevent DDoS attacks in cloud environment. This new approach of hop Count Storage is save searching time at the time of hop-count Filtering & helps in preventing DoS attack in cloud environment. Also, this method decreases the unavailability of cloud services to legitimate users, increases accessibility and reduces memory requirement for storing hop-count value.

1 Introduction

Cloud computing can be defined as a new style of computing in which dynamically scalable and often virtualized resources are provided as a services over the Internet. Advantages of the cloud computing technology include cost savings, high availability, and easy scalability [1].

DoS attacks do not wish to modify data or gain illegal access, but instead they target to crash the servers and whole networks, disrupting legitimate users' communication. DoS attacks can be launched from either a single source or multiple sources. Multiple-source DoS attacks are called distributed denial-of-service (DDoS) attacks [2].

When the operating system notices the high workload on the flooded service, it will start to provide more computational power to cope with the additional workload. The attacker can flood a single, system based address in order to perform a full loss of availability on the intended service [4, 6].

These attacks are a type of Flooding Attack [2, 5], which basically consist of an attacker sending a large number of nonsense requests to a certain service, which is providing various services under cloud. As each of these requests has to be handled by

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 511–518.
springerlink.com

the service implementation in order to determine its invalidity, this causes a certain amount of workload per attack request, which in the case of a flood of requests usually would cause a Denial of Service to the server hardware [2].

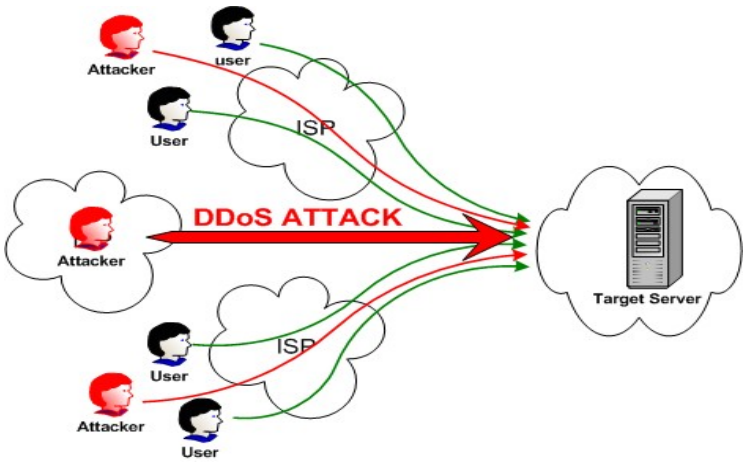


Fig. 1. DDoS attack [2]

2 Balanced Tree

A B-tree is a tree data structure that keeps data sorted and allows insertions and deletions that is logarithmically proportional to file size. Usually, sorting and searching algorithms have been characterized by the number of comparison operations that must be performed using order notation. A binary search of a sorted table with N records, for example, can be done in O(log2N) comparisons. In B-trees, internal nodes can have a variable number of child nodes within some pre-defined range. When data is inserted or removed from a node, its number of child nodes changes. In order to maintain the pre-defined range, internal nodes may be joined or split [7].

A data structure is a way of storing data in a computer so that it can be used efficiently. Often a carefully chosen data structure will allow the most efficient algorithm to be used. The choice of the data structure often begins from the choice of an abstract data structure.

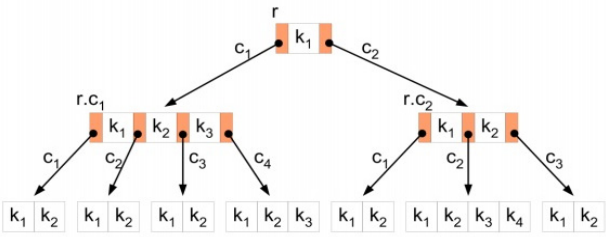


Fig. 2. Schema showing B-tree Data Structure [7]

A well-designed data structure allows a variety of critical operations to be performed, using as few resources, both execution time and memory space, as possible. Data structures are implemented using the data types, references and op-erations on them provided by a programming language. B-trees are particularly well-suited for implementation of databases, while routing tables rely on networks of machines to function [7].

3 Hop-Count Computation

Since hop-count information is not directly stored in the IP header, one has to compute it based on the Time-to-live (TTL) field. TTL is an 8-bit field in the IP header, originally introduced to specify the maximum lifetime of each packet in the Internet. Each intermediate router decrements the TTL value of an in-transit IP packet by one before forwarding it to the next-hop [3, 8].

3.1 Extract Final Value of TTL

When a Packet reaches its destination and extracting its TTL field value, this value is known as final TTL. The challenge in hop-count computation is that a destination only sees the final TTL value. It would have been simple had all operating systems (OSs) used the same initial TTL value, but in practice, there is no consensus on the initial TTL value. Furthermore, since the OS for a given IP address may change with time, we cannot assume a single static initial TTL value for each IP address [3].

3.2 Investigate the Initial Value of TTL

According to [3], most modern OSs uses only a few selected initial TTL values, 30, 32, 60, 64, 128, and 255. Only a few Internet hosts are apart by more than 30 hops, thus one can determine the initial TTL value of a packet by selecting the smallest initial value in the set that is larger than its final TTL. For example, if the final TTL value is 112, the initial TTL value is 128, the smallest of the two possible initial values, 128 and 255. Thus, given the final TTL value one can find the initial TTL value. Initial TTL values can be calculated as follows [9]:

Initial TTL=32 if final TTL <=32
Initial TTL =64 if 32<final TTL<=64
Initial TTL =128 if 64<final TTL <=128
Initial TTL =255 if 128<final TTL <=255

3.3 IP2HC Table

The inspection algorithm infers the initial TTL value and subtracts the final TTL value from it to obtain the hop-count. This value of hop count is stored into the IP2HC table. The IP2HC table [9] is a mapping between Source IP Address of a packets and stored hop count for that IP Address. It is a structure with Source IP address serving as index to match the hop count information.

4 Proposed Storage Techniques

The proposed storage technique uses the B-tree concept for hop count filtering mechanism, and provides a clear idea for storing hop count value of clients, so that it can be used in Cloud environment to prevent DoS attacks. This scheme inspects the hop count value of incoming packets to validate their legitimacy using only moderate amount of storage.

Continue monitoring of packets travelling over the cloud network, and thus, we extract information from monitored TCP/IP Packets for calculating the hop-count value. The analyses to investigate the limit of hop count (HC) values are

Initial TTL=32 if final TTL <=32, Then possible HC value range 0 to 32

Initial TTL =64 if 32<final TTL<=64, Then possible HC value range 0 to 31

Initial TTL =128 if 64<final TTL <=128, Then possible HC value range 0 to 63

Initial TTL =255 if 128<final TTL <=255, Then possible HC value range 0 to 126

So, we can say that calculated hop count value range between 0 to126.

First we create a root node and it's divided into 126 parts. Each part of root node index by number (represent hop count value) and linked with source IP address. Initially each value points to null. After calculating hop count value from the received TCP/IP packet at the cloud server need to insert client IP address corresponding to hop count value. For example, if value of hop count is p then this node point to the node containing client IP address. We are discussing two techniques for storage of hop count.

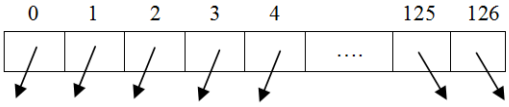


Fig. 3. Structure of root node

4.1 Linked Storage Scheme

In this scheme, Clients IP address information is linked by node containing hop count value. In fig. 4 IP address are represented by A: B: C: D.

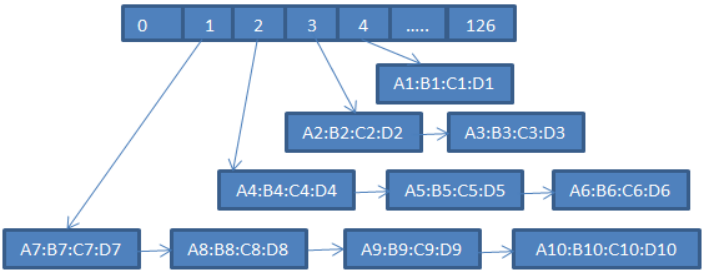


Fig. 4. Linked Storage Scheme

If multiple client having same hop count value then create link list of each client IP addresses. Node containing value 1 points to the link list of IP addresses (A7:B7:C7:D7, A8:B8:C8:D8, A9:B9:C9:D9, A10:B10:C10:D10) because of these clients having hop count value 1.

4.2 Hierarchical Hop Count Storage Scheme

In this scheme, we are using concept related to B tree data structure for storing IP addresses of clients on the basis of their hop count values. In this data structure a child node can have variable number of child nodes.

In fig. 5, IP address are divided into four part (A: B: C: D) and following IP Addresses (A1:B1:C1:D1, A2:B1:C1:D1, A3:B1:C1:D1, A4:B1:C1:D1, A11:B2:C1:D1, A12:B2:C1:D1, A13:B2:C1:D1, A21:B12:C1:D1, A22:B12:C1:D1, A23:B12:C1:D1, A31:B23:C1:D1, A32:B23:C1:D1, A33:B23:C1:D1 etc.) having hop count value 3 which is shows in figure.

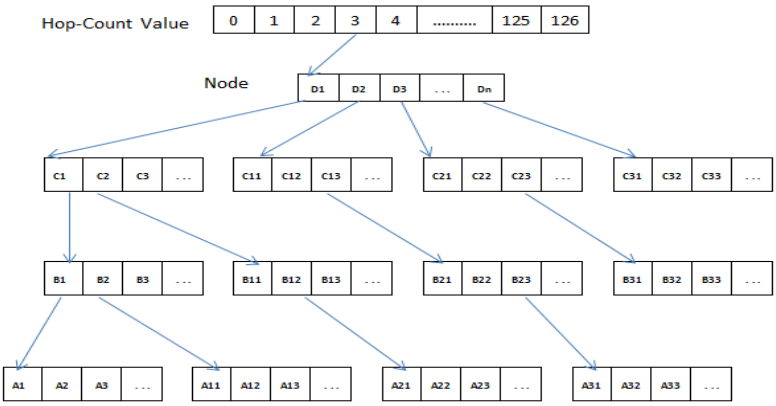


Fig. 5. Hierarchical Hop Count Storage Scheme

A hierarchical storage scheme capable of improving hit rate and reduces the complexity for extracting stored hop count value from storage. For example, cloud server receives an IP A32: B23: C13: D2 and its computed hop count value is 3 then refer root node which contain Hop count value equal to 3. This scheme works on three mode Insertion, Detection and Deletion mode. In Detection mode, computing the hop count value and recognize client IP exist or not. And in Insertion mode, insert client IP address if received connection request. When validity of client is expired then enter into Deletion mode to remove client IP address information from the storage.

Algorithm 1. Consider the following notations: HC = Hop Count value, Tf= final value of TTL and Ti=initial value of TTL.

For each packet,

Extract the final ttl (Time to Live) value and Source IP address.
Compute Hop count value
Set Ti = Investigate_Initial_TTL(Tf)
HC = Ti-Tf

If connection request then enter into Insertion mode.
else Enter into Detection Mode

A. *Detection Mode*

If child node D_n exist then check
If child node C_n exist then check
If child node B_n exist then check
If child node A_n exist then
'Allow the packet'
If validity of client is expired then
Switch to the Deletion Mode
Otherwise 'packet is spoofed'

B. *Insertion Mode*

Steps:
1.If child node D_n exist then goto Step 2 else goto Step 5
2. If child node C_n exist then goto Step 3 else goto Step 6
3. If child node B_n exist then goto Step 4 else goto Step 7
4. If child node A_n exist then goto Step 9
5. Create a nodes & insert value of D_n
6. Create a node & insert value of C_n
7. Create a node & insert value of B_n
8. Create a node & insert value of A_n
9. Exit

C. *Deletion Mode*

// Deletion perform from a leaf node
Delete node A_n and check if node A is empty then
Delete node B_n and check if node B is empty then
Delete node C_n and check if node C is empty then
Delete node D_n

5 Important Features of Proposed Approach

These scheme increase accessibility of HC value and reduces the memory requirement for storing HC values. Easy to perform time efficient operations like searching, insertion and deletion searching.

We proposed two scheme, linked storage scheme save searching time as compare to IP2HC table. Because of in IP2HC table needs to search in sequential order (The source IP address serves as the index into the table to retrieve the correct hop-count for this IP address). But in this scheme create a link list of client's IP addresses then again needs lots of time in searching HC value. So, we introducing Hierarchical hop count storage scheme which is save lots of searching time of IP address and help in preventing DDoS attack in cloud environment and also this scheme save memory and decreases unavailability of cloud services.

6 Conclusions

Cloud Computing is gaining popularity, but with the widespread usage of cloud the issue of cloud security is also surfacing. One of the major threats to Cloud security is Distributed Denial of Service Attack (DDoS) or simply Denial of service attack (DoS). This attack will damaging system and decreases resources availability without any previous information. To improve resource availability of resources, it is essential to provide a time efficient mechanism to prevent DDoS attacks. This paper presented a time efficient approach by using hierarchical data structure, which is not only store the hop count but also detects spoofed packets.

References

1. Furht, B., Escalante, A.: Handbook of Cloud Computing, pp. 3–11. Springer (2010)
2. Garg, D.: DDOS Mitigation Techniques-A Survey. In: International Conference on Advanced Computing, Communication and Networks, pp. 1302–1309 (2011)
3. Haining, W., Cheng, J., et al.: Defense Against Spoofed IP Traffic Using Hop-Count Filtering. IEEE/ACM Transactions on Networking 15(1), 40–53 (2007)
4. Kumar, P.A.R., Selvakumar, S.: Distributed Denial-of-Service (DDoS) Threat in Collaborative Environment - A Survey on DDoS Attack Tools and Traceback Mechanisms. In: IEEE International on Advance Computing Conference (IACC), pp. 1275–1280 (2009)
5. Templeton, S., Levitt, K.: Detecting spoofed packets. In: Proceedings of The Third DARPA Information Survivability Conference and Exposition (DISCEX III), Washington, D.C (2003)
6. Mann, P.S., Kumar, D.: A Reactive Defense Mechanism based on an Analytical Approach to Mitigate DDoS Attacks and Improve Network Performance. International Journal of Computer Applications 12(12), 43–46 (2009)
7. Marécha, B.: B-tree introduction (2007), `http://computing.unn.ac.uk/openDBproject/content/system/developments/FileStacks/B-TreeFileStack/btree.pdf`

8. Mopari, I.B., Pukale, S.G., et al.: Detection and defense against DDoS attack with IP spoofing. In: International Conference on Computing, Communication and Networking (ICCCN), pp. 1–5 (2008)
9. Venkatesu, N., Chakravarthy, V.D., et al.: An Effective Defense Against Distributed Denial of Service in GRID. In: First International Conference on Emerging Trends in Engineering and Technology (ICETET), pp. 373–378 (2008)

VHDL Synthesis and Simulation of an Efficient Genetic Algorithm Based on FPGA

N. Rajeswaran[1], T. Madhu[2], and M. Suryakalavathi[3]

[1] JNTUH/SNS College of Technology,
Coimbatore, Tamilnadu, India
rajeswarann@gmail.com
[2] Swarnandhra Institute of Engg., and Technology,
Narasapur, Andhrapradesh, India
tennetimadhu@yahoo.com
[3] Jawaharlal Nehru Technological University,
Hyderabad, Andhrapradesh, India
munagala12@yahoo.co.in

Abstract. Genetic Algorithm (GA) is an artificial intelligence procedure and one of the probabilistic heuristic search algorithms based on the mechanism of natural selection and evaluation. The GA is used to select the characteristic parameters of the classifiers, the input features and find the optimum solution for a variety of complex problems like Very Large Scale Integrated (VLSI) design, layout and test automation. Field Programmable Gate Array (FPGA) is an integrated circuit designed to be configured by the customer or designer after manufacturing and is very widely used in VLSI Circuits. The GA architecture is simulated and verified by using VHDL (Very High Speed Integrated Circuit Hardware Description Language).

1 Introduction

Digital systems, which are extremely intricate and increasing in complexity, are used in wide range of domestic and industrial applications so as to ensure the reliability. It is necessary to test their performance to identify any defects prior to using them in a fully operational environment. The cost of testing VLSI chips is a significant function of the overall manufacturing cost. The time required to test a chip should be minimized. The tremendous growth of the recent techniques to increase the transistor density of FPGA has yielded good results and is more powerful over the years [9][4]. General purpose GA requires that the fitness function be easily changed, the hardware implementation must exploit the reprogrammability of certain types of FPGAs, which are programmed via a bit pattern stored in the static RAM and are thus easily reconfigured [6]. FPGA is a semiconductor device from Programmable Read Only Memory (PROM) and Programmable Logic Device (PLD). It contains programmable logic components and programmable interconnects [7]. FPGAs are pre-fabricated silicon devices that can be electrically programmed to become almost any kind of digital circuit or system [1][8]. The FPGA contains CLB (Configurable Logic Blocks),

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 519–525.
springerlink.com

Horizontal and Vertical lines. In contrast to Application Specific Integrated Circuits (ASIC), FPGAs are configured after fabrication and they can also be reconfigured. The CLB's are calculated the user defined functions in FPGA. The Input/Output Blocks (IOB) are used to connect the FPGA to other elements. Interconnect is important for writing data between CLB and from IOBs to CLBs.

2 Genetic Algorithm

GA is a probabilistic search algorithm based on the mechanism of natural selection and evaluation [3].In GA the term "chromosomes" encode a group of linked features and the "Genes" encode the activation or deactivation of a feature. GA is started with a set of solutions called population. A solution is represented by a chromosome. The population size is preserved throughout each generation. Some of the selected chromosomes randomly mate to produce new offspring [5]. Chromosomes with high fitness value have high probability of being selected. The new generation may have higher average fitness value than those of old generation. The process of evaluation is repeated until the end condition is satisfied. Genetic algorithm described by Goldberg is specially suited to solve large scale combination optimization problems [2]. The fig.1 shows the flow chart of GA. The initialization of strings is represented by the population size and after the initialization it will move in to the genetic process that is Selection, Cross over and Mutation. Finally, it will check for the output of the genetic process to reach the optimum and expected solution and if so then the execution will be stopped else it starts again from the initial steps.

2.1 GA Process

Step 1: Initialization of data strings – Population Size (n chromosomes).
Step 2: Calculate the fitness f(x) of each chromosome x in the population.
Step 3: Create a new population by repeating the following steps [4-9] until the new population is completed.
Step 4: Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected).
Step 5: With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
Step 6: With a mutation probability mutate new offspring at each locus (position in chromosome).
Step 7: [Accepting] Place new offspring in a new population.
Step 8: [Replace] Use new generated population for a further run of algorithm.
Step 9: [Test] If the end condition is satisfied, stop, and return the best solution in current population.
Step 10: Repeat from Step 2 and continue the process until the optimum solution reached.

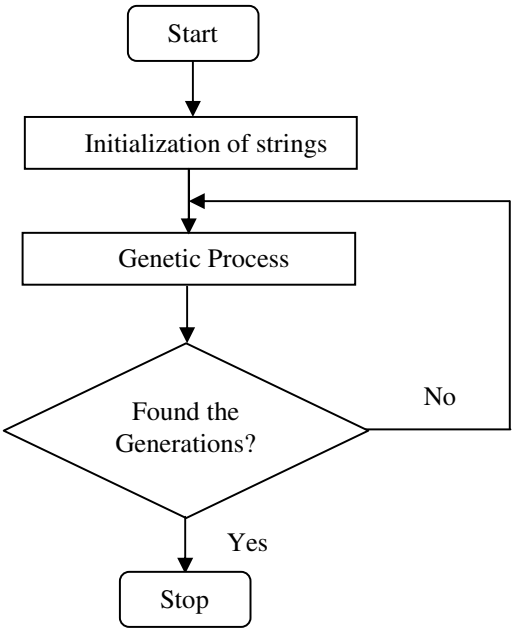


Fig. 1. Flow chart of GA

3 Simulation Results

The simulations parameters are tabulated in Table 1.The simulated outputs are shown in fig. 2, 3 and 4. The detailed synthesis reports of utilization of hardware components in FPGA are tabulated in Table 2. The Rand for the Random value and Wheel here we used the Roulette Wheel for the Random value selection. RUN is a global startup and running signal, RESET is a global reset signal and CLK is a global synchronizing clock signal.

3.1 Pseudocode for GA

```
architecture rtl of GA is
-- Build an enumerated type for the state machine-----------------------------------
type state_type is (idle, birth, selection, crossover, mutation, store);
……
-- Register to hold the current state-----------------------------------------------
signal o0,o1,o2,o3,o4 : std_logic;
signal Child_input1, Child_input2, Child_input1_b, Child_input2_b, GA1,GA2:
……
--Selection state-----------------------------------------------------------------------
          if(state_out="000100") then
 ……
 --Here general Roulette wheel-----------------------------------------------------
            for i in 0 to 3 loop
             if(Rand(i)<wheel(0)) then
……
```

```
--Crossover state-------------------------------------------------------------------------
          if(state_out="001000") then
           temp2_cros:=Child_input2_s;
 swp:=temp1_cros(1);
          temp2_cros(1):=swp;
……….
--Mutation state------------
                  if(state_out="010000") then
                           temp1_mu:=Child_input1_c;
                           temp2_mu:=Child_input2_c;
……….
-------------------------------------------------------------------------------------------
```

Table 1. Simulation parameters of GA

Parameters	Bit Value
Rand	1011
Wheel	1111
Input	1111000011110110
Output	0010100101110001
Rst	0
Run	1
Ce	1

Table 2. Synthesis report of GA

Device	Utilization
Number of slices	8 out of 4656 0.171%
Number of Slice Flip Flops	3 out of 9312 0.032%
Number of 4 input LUTs	14 out of 9312 0.15%
Number of IOs	14
Number of bonded IOBs	14 out of 158 8%
Number of GCLKs	1 out of 24 4%
Minimum period	3.668ns (Maximum Frequency: 272.628MHz)
Minimum input arrival time before clock	4.024ns
Maximum output required time after clock	6.095ns
Total memory usage	174372 kilobytes

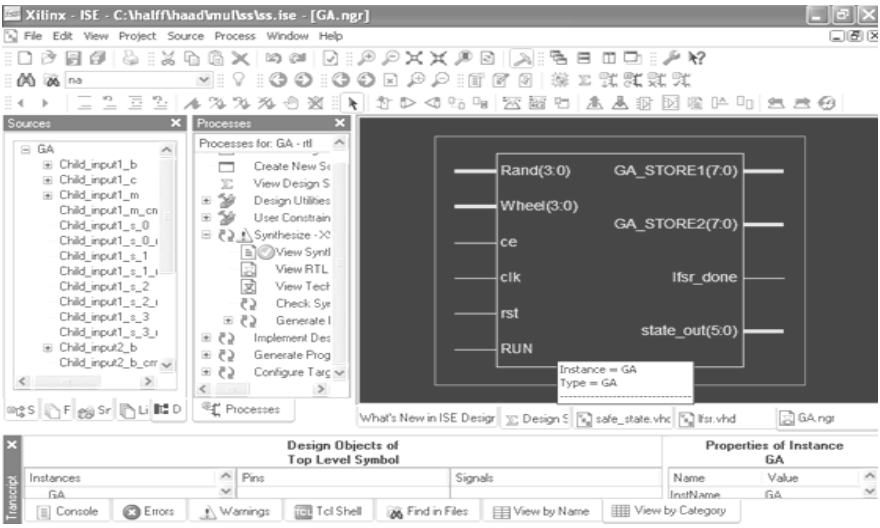


Fig. 2. RTL schematic view of GA

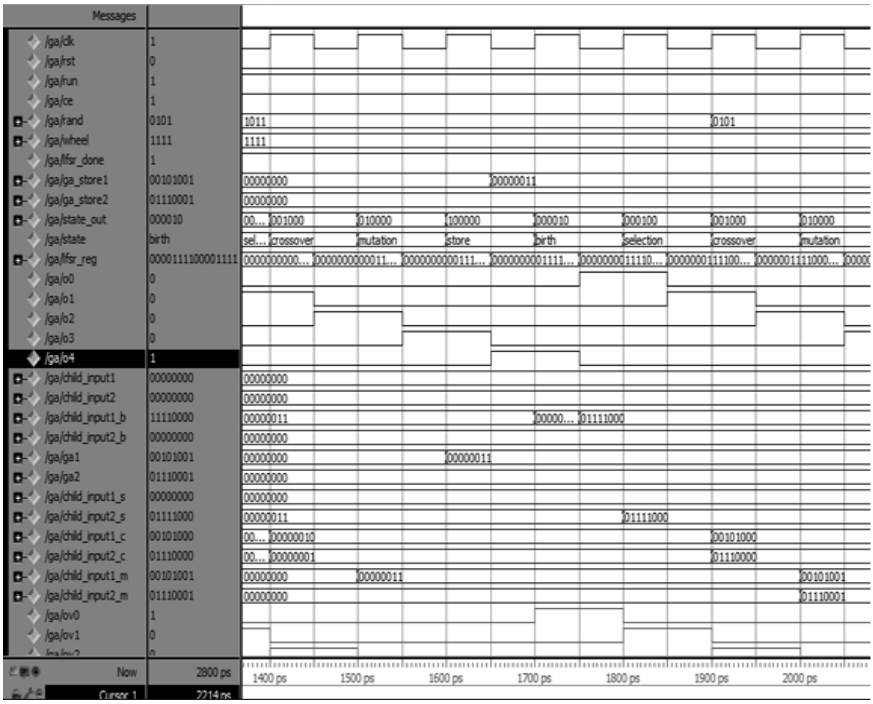


Fig. 3. Simulated output of GA

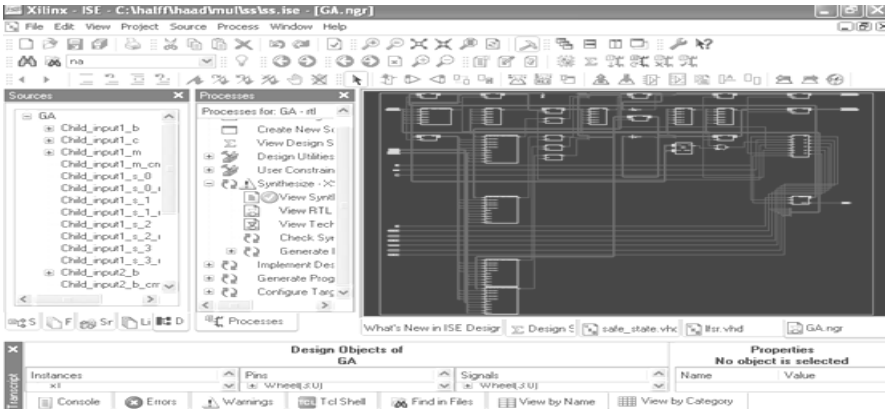


Fig. 4. Technological schematic view of GA

4 Conclusion

Now days, the reliability of the digital systems is being tested by Artificial Intelligence (AI) based approaches, the robust and efficient way to detect the faults. It is somewhat cumbersome to develop an ASIC or FPGA for search and optimization problems. Developing an ASIC takes much time and is expensive compared with FPGA. The GA is a promising method for solving such difficult technological problems. The GA architecture is synthesized and simulated by using VHDL programming language presented here. The GA is easy to drop into local optimal solution and to increase the convergence speed. It can also be used for complex and loosely defined problems in Very Large Scale Integrated circuits.

References

1. Aporntrwan, C., Changstitvatana, P.: A Hardware implementation of the compact genetic algorithm. In: Proceedings of the IEEE Congress on Evolutionary Computation, Seoul, Korea, pp. 624–629 (2001)
2. Goldberg, D.E.: Genetic Algorithms in Search, Optimization & Machine Learning. Addison Wesley (1989)
3. Whitley, D.: A genetic algorithm tutorial (2001), `http://samizdat.mines.edu/gatutorial`
4. Miihlenbein, H.: Evolutionary Theory and Applications. In: Aarts, E.H.L., Lenstra, J.K. (eds.) Local Search in Combinatorial Optimization. Wiley, New York (1993)
5. Frounchi, J., Zarifi, M.H., Far, S.A., Taghipour, H.: Design and Analysis of Random Number Generator for Implementation of Genetic Algorithms using FPGA. In: Proceeding of 5th International Conference on Electrical and Electronics Engineering (ELECO 2007), Bursa, Turkey, pp. 401–404 (December 2007)
6. Chambers, L.D.: Practical Handbook of Genetic Algorithms: Complex Coding Systems, vol. III

7. Scott, S.D., Seth, S., Samal, A.: A Synthesizable VHDL Coding of a Genetic Algorithm. Technical Report UNL-CSE-97-009,University of Nebraska-Lincoln, November 19 (1997)
8. Tachibana, T., Murata, Y., Shibata, N., Yasumoto, K., Ito, M.: General Architecture for Hardware Implementation of Genetic Algorithm. In: Proceedings of 14th IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM 2006), pp. 291–292 (April 2006)
9. Lei, T., Cheng, Z.M., Wang, J.-X.: The Hardware Implementation of a Genetic Algorithm Model with FPGA. In: Proceedings of IEEE International Conference on Field Programmable Technology (FPT), pp. 374–377 (2002)

Forensic Sketch Matching Using SURF

Dileep Kumar Kotha and Santanu Rath

National Institute of Technology, Rourkela
Orissa, India – 769008
{dileep98490,rath.santanu}@gmail.com

Abstract. This paper deals with the problem of forensic sketch matching. Research in past decade offered solutions for matching sketches that were drawn while looking at the subject (viewed sketches). In this paper, we emphasize on matching the forensic sketches, which are drawn by specially trained artists in police department based on the description of subject by an eyewitness. Recently, a method for forensic sketch matching using LFDA (Local Feature based Discriminant Analysis) was published. Here, the same problem is addressed using a novel preprocessing technique combined with a local feature descriptor called SURF (Speeded Up Robust Features). In our method, we first preprocess the images using a special preprocessing technique suitable for forensic sketch matching. After the preprocessing, SURF is used to extract features in the form of 64-variable vectors for each image. Then all these vectors of one image are combined to form the SURF descriptor vector for that image. These descriptor vectors are then used for matching. This method was applied to match a dataset of 64 Forensic Sketches against a gallery of 1058 photos. From our experiments, it was observed that our approach of image preprocessing combined with SURF had shown promising results with a good accuracy.

1 Introduction

Today, advances in biometric technology have provided law enforcement agencies additional tools in the identification of criminals. In addition to the incidental evidence, if a dormant fingerprint is found at the scene of crime or a surveillance camera captures an image of a suspect's face, then these clues are used in determining the suspect using biometric identification techniques. However, many crimes occur where none of the above discussed information is present. Also, the lack of technology to effectively capture the biometric data like finger prints within a short span after the scene of crime, is a routine problem in remote areas. Despite these repercussions, many a times, an eyewitness account of the crime is available who had seen the criminal. The Police department deploys a forensic artist to work with the witness in order to draw a sketch that limns the facial appearance of the culprit. These sketches are known as forensic sketches. Once the sketch is ready, it is sent to the law enforcement officers and media outlets with the hope of catching the suspect. Here, two different scenarios may arise for the culprit: 1) The person may have already been convicted once or 2) The person has not been convicted even once or this is the first time,

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 527–537.
springerlink.com

he may be committing felony. This paper deals with the first type scenario. If the criminal has been convicted at-least once, a mug shot photo (photo taken, while the person is being sent to jail) is available. Using an efficient forensic sketch matching system, the police can narrow down the potential suspects which will reduce the future crimes by the same criminal drastically.

There are two different types of face sketches that are discussed in this paper. They are: *viewed sketches* and *forensic sketches*. *Viewed sketches* (Fig. 1) are the sketches drawn while looking at the photograph of the person or the person himself. *Forensic sketches*(Fig. 2) are the sketches drawn by interviewing a witness to gain description of the suspect. There are a lot of problems in face sketch recognition when compared to normal face recognition (in which both probe and gallery images are photographs). The textures of sketches, whether they may be viewed or forensic are quite different from that of the gallery of photographs that were being matched against. Previous work in the matching is done only on Viewed Sketches [3],[4],[5], [6], [7], even though most real world scenarios involve forensic sketches only. Forensic sketches have additional problems compared to viewed sketches. Due to the petulant nature of memory, the exact appearance of the criminal cannot be remembered by the witness. This leads to an incomplete and inaccurate depiction of the sketches which reduces the recognition performance substantially.

2 Related Work

Even though the research in sketch matching started a decade ago, it is mostly confined to viewed sketches. Most of the early work on viewed sketches is done by Tang *et al.* [3],[5],[6]. In these studies, a synthetic photograph is generated from the sketch and then matching is performed with established face recognition algorithms.

In the recent years, research on sketches is done with the aid of feature based descriptors. Klare and Jain [4] published a SIFT based approach for the sketch to photo matching. Other methods similar to this such as Coupled Spectral Regression [9], Local Binary Patterns [8], [10], [11] are used for matching near-infrared images (NIR) to visible light images (VIS).

Fig. 1. An example of viewed sketch (Left) and it's corresponding photograph (Right)

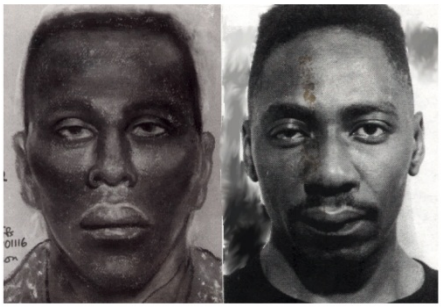


Fig. 2. An example of forensic sketch (Left) and it's corresponding photograph (Right)

Klare and Jain[1] published an LFDA based approach for matching forensic sketches to mug shot photos. It is the first large scale experiment conducted on forensic sketch matching. We compare our results with the LFDA; since it is reported to be the one with the highest accuracy till date in forensic sketch matching. We show experimentally that with a novel preprocessing technique, combined with the detector and descriptor powers of SURF [2], a better accuracy than LFDA can be achieved.

3 Preprocessing

A novel preprocessing technique is discussed in this section. This preprocessing is different from the conventional face recognition preprocessing techniques where the face is preprocessed so that the region from forehead to chin and cheek to cheek is visible. Here, we preprocess the images, so that the hairline and neck region along with the ears are also visible (as shown in Fig. 3). This is due to two reasons:

1. Experiments conducted by Frowd *et al.* [12] showed that human beings remember the familiar and unfamiliar faces with the help of internal and external features of the face respectively. Since culprits are essentially unfamiliar, the external features of the face region are more salient and hence need not be removed.
2. Forensic Sketch artists not only draw the internal features of the face, but also the external features. Also, Jain *et al.* [13] reported that when matching forensic sketches, using only the external features (chin,hairline,ears) gave better accuracies when compared to using only the internal features (eyes, nose, mouth etc.)

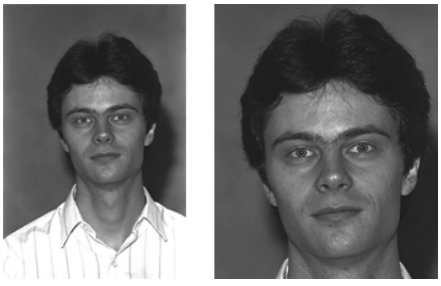


Fig. 3. An example of the image preprocessing. Note that external features of the face are not lost in the preprocessed image.

Since SURF is both rotation and scale invariant, we did not preprocess the images further. The use of haar wavelet responses makes SURF invariant, to a bias in illumination [2].

4 Speeded Up Robust Features

SURF stands for Speeded Up Robust Features. It is an approach which is generally used to construct a robust image detector and descriptor that can be used in computer vision tasks like object recognition and 3D reconstruction. Recent experiments by Du *et al.* [14] proved SURF to be the most robust detector and descriptor available for face recognition. Also, using SURF feature descriptors, the differences in image modalities between a sketch and a photo are mostly diminished.

The features calculated with SURF, are both rotation and scale invariant. In a typical face recognition experiment, there is always a need to scale up/down the images and also to rotate the subject's face so that both eye levels fall on a straight line. This overhead is completely removed with SURF.

As a detector, SURF locates the interest points in the image that produce major variation while the descriptor constructs feature vectors around each of these interest points. In the next few sections we describe how SURF can actually be used for recognition purposes.

4.1 Interest Point Detection

To detect the interest points, SURF uses the determinant of the approximate Hessian matrix. Blob like structures are detected in the image, where the local determinant is maximum (see Fig. 4a). In the Hessian matrix approximation, we use integral images instead of the original ones reducing the time required for calculations.The Hessian matrix $H(X,\sigma)$ for a given point $X = (x, y)$ of an image at a scale σ

$$H(X,\sigma) = \begin{bmatrix} L_{xx}(X,\sigma) & L_{xy}(X,\sigma) \\ L_{xy}(X,\sigma) & L_{yy}(X,\sigma) \end{bmatrix}, \tag{1}$$

Where $L_{xx}(X,\sigma)$, $L_{xy}(X,\sigma)$ and $L_{yy}(X,\sigma)$ are the convolutions of the Gaussian second order partial derivatives of the image I at point X.

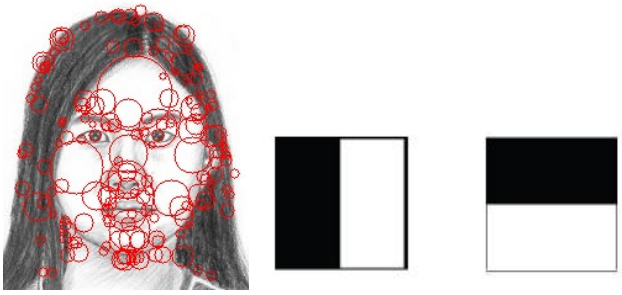
A set of 9 × 9 box filters are used as approximations of Gaussian second order derivatives with $\sigma = 1.2$, to reduce the computation time. These filters represent the lowest scale(i.e. highest spatial resolution) for computing blob response maps and are denoted as $D_{xx}(X,\sigma)$, $D_{xy}(X,\sigma)$ and $D_{yy}(X,\sigma)$. The weights applied to the rectangular region are kept simple for computational efficiency. These yield:

$$\det\left(H_{approx}\right) = D_{xx}D_{yy} - \left(\omega D_{xy}\right)^2, \tag{2}$$

Where ω is a weight for the energy conservation between the Gaussian kernels and the approximated Gaussian kernels. To be scale invariant SURF implements scale spaces as image pyramids. In a general scenario, these images are repeatedly smoothed with a Gaussian and subsequently sub-sampled in order to achieve a higher level of the pyramid. But in SURF, since we use box filters and integral images, we can directly apply the filters of any size at exactly same speed, directlyon the original image.

4.2 Interest Point Description

SURF uses the sum of Haar wavelet responses to describe the features of an interest point, which make it invariant to rotation. Fig. 4b shows the Haar wavelet filters, that are used to compute the responses in x and y directions. To extract the descriptors, we first construct a square region centered at the interest point and oriented along the orientation decided by a special selection method as described in [2].



(a) (b)

Fig. 4. (a) The interest points detected with SURF for a sketch, (b) The Haar Wavelet types used for SURF

Now the square region is split up equally into smaller 4 × 4 square sub-regions (as shown in Fig. 5). This preserves important spatial information. For each sub-region, we compute Haar-wavelet responses at 5 × 5 equally spaced sample points. We denote d_x as the Haar Wavelet response in horizontal direction and d_y as the Haar wavelet response in vertical direction. For each sub-region, d_x and d_y are calculated and these are weighted with a Gaussian centered at the interest point to increase the robustness towards geometric deformations and localization errors.

The wavelet responses d_x and d_y are summed up over each sub-region and these form a first set of entries to the feature vector. In order to bring in information about the polarity of the intensity changes, we also extract the sum of the absolute values of the responses, $|d_x|$ and $|d_y|$. Now each sub-region has a four-dimensional descriptor vector v for it's underlying intensity structure,

where. $v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right)$. Concatenating these vectors of all the 4 × 4 sub-regions, we get a descriptor vector of length 64. The wavelet responses are invariant to a bias in illumination (offset).

4.3 Speed Up the Matching

To speed up the matching, we used the sign of the Laplacian (i.e. the trace of the Hessian matrix) for the interest point. If two point pairs are of different sign, their features are not matched. Fig. 6 gives the example blobs of the sign where they are different and hence are not matched. More detailed description of SURF matching can be found in [2].

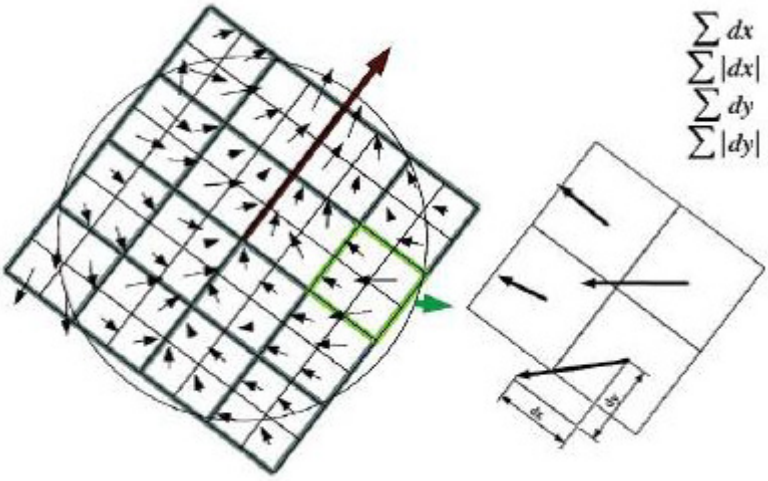


Fig. 5. To build the descriptor, an oriented quadratic grid with 4 × 4 square sub-regions is laid over the interest point

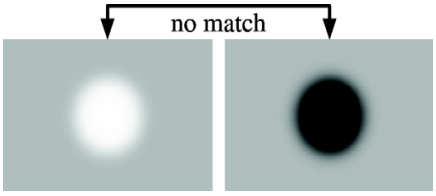


Fig. 6. Figure showing the sign of the Laplacian

5 Viewed Sketch Matching Results

We first performed experiments on viewed sketches to test our SURF based system. 188 pairs of viewed sketches were collected from CUHK face data set [3]. Taking a

random sample of 100 pairs, we conducted the recognition experiment. There is no training required since we are using the detector capabilities of SURF. The Cumulative match curve that was generated is as shown in Fig. 7.

At rank-10, we achieved an accuracy rate of 77%. Although this result lags behind the viewed sketch matching results of [3] and [4], we have shown that without any training or higher level preprocessing a good accuracy can be achieved with SURF. The reason for lower accuracy may be attributed to the fact that we preprocessed the images keeping in mind the cognitive research on forensic sketches, but not on viewed sketches.

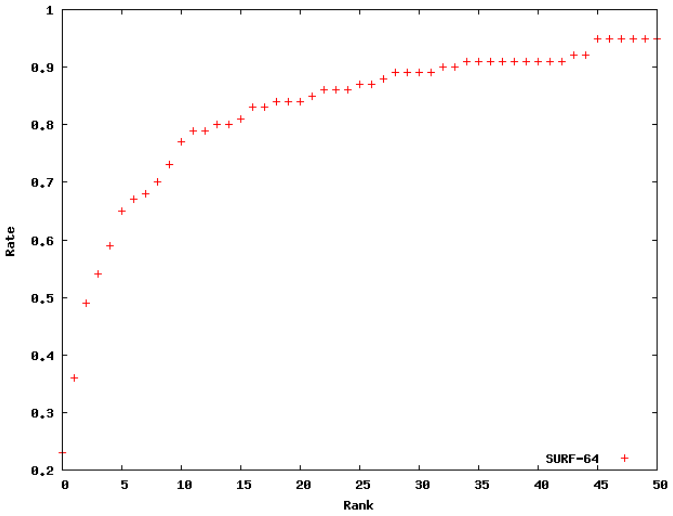


Fig. 7. Rank curve (CMC) showing the matching performance with Viewed Sketch Images

6 Forensic Sketch Matching Results

Before the experiment, we made a database of 64 forensic sketches along with their corresponding mug shot photographs. These sketches were collected from two different sources:

1. 54 sketches from the forensic sketch artist Lois Gibson [15]
2. 10 sketches from the forensic sketch artist Karen Taylor [16]

Collection of mug shot database is a huge problem, since there is no publicly available database for them. Only 64 mug shot photos which were the pairs of the sketches we have collected were available. We could not collect any additional mug shot photos. So, in order to populate the gallery, we used the *fa type* images(994 in number) of FERET color database [17]. As a result, we have a gallery 1058 photographs in the end.

Also, when we are doing the matching of forensic sketches, we are concerned with the accuracy at rank-50 i.e. whether or not the true subject is within the top-50 retrieved images. This is because forensic sketch matching differs a lot from the conventional face recognition scenarios. In normal face recognition, human interaction is limited only to the ambiguous cases. Since in forensic sketch matching, we are matching a sketch to a photo, and that sketch too is drawn just based on the verbal description of the witness, there are a lot of chances for ambiguity. Hence, law enforcement officers are generally concerned with the top R retrieved results. Here, we consider R to be 50. We first used all the 64 forensic sketches we have, for matching. We achieved a very good accuracy rate of 46.87%. We believe, that this is the best recognition rate achieved so far in the area of forensic sketch matching. The rank curve (Cumulative Match Curve) that was generated is as shown in the Fig. 8.

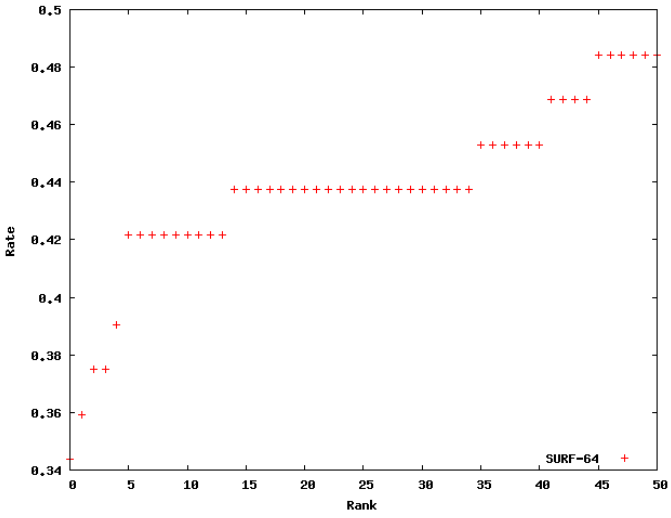


Fig. 8. Figure showing the CMC that was generated when 64 forensic sketches are matched against 1058 photographs

In order to compare our result with the existing systems LFDA [1] and FaceVACS [18], we tried to mimic the experiment in [1]. Klare and Jain [1] took 49 good quality forensic sketches, from a lot of 159 forensic sketches available to them and showed their results. From the lot of 64 forensic sketches we have, we separated 49 sketches as good quality and performed matching on them. A note tothe reader is that the 64 sketches we have are a subset of the 159 sketches in [1].The results are shown in Tab. 1. The results clearly show the SURF based method, along with the novel preprocessing technique we proposed as a winner. The accuracy can be further improved if race, gender and ancillary information are included.

Table 1. Comparison of rank-50 accuracies of our method (SURF) with LFDA and FaceVACS

Method	Rank-50 accuracy
SURF with novel preprocessing	61.23%
LFDA	52%
FaceVACS	23%

Fig. 9 shows the examples of matching with our proposed system. Sometimes the top retrieval may look visually more similar to the sketch rather than the true subject (see Fig. 10). This gives us another dough to explain why we consider top-50 retrieved images rather than one single image that appears at rank-1.

(a) (b)

Fig. 9. Example matches when 49 good quality forensic sketches are matched against 1058 photographs (a) Three of the best matches which were discovered at rank-1 and (b) Three of the worst matches discovered at ranks 320,217 and 287 (from left to right) successively

(a) Forensic Sketch (b) True subject (c) Top retrieval

Fig. 10. An example showing the failed retrieval for a good quality sketch. Even though the top retrieval (c) is not true subject (b), it visually looks more similar to the forensic sketch (a).

7 Conclusions and Future Work

We presented a novel approach for forensic sketch matching with the help of SURF. Matching forensic sketches is a very difficult problem for two main reasons:

1. The sketch quality is directly proportional to the victim's memory
2. We need to match across image modalities

We solved the latter problem with the help of SURF and with a special preprocessing technique we tried to solve the former one. Also, by removing the needs of training and other higher level preprocessing techniques (scaling and rotating of images), we reduced the time required for preprocessing drastically. Further, we provided an optimal method for forensic sketch matching and proved experimentally it's superiority compared to the existing systems. Future work can be extended by making improvements to SURF.

There is a need for continual research on forensic sketch matching to assist the law enforcement agencies to apprehend criminals quickly, before they commit another crime. We have sent requests to the various universities doing the research on forensic sketch matching to make their databases publicly available. A bigger database of forensic sketches is needed to further dive into the complexity of the problem.

References

1. Klare, B., Li, Z., Jain, A.K.: Matching Forensic Sketches to Mug Shot Photos. IEEE Transaction Pattern Analysis and Machine Intelligence, 639–646 (March 2011)
2. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)
3. Wang, X., Tang, X.: Face Photo-Sketch Synthesis and Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 31 (2009)
4. Klare, B., Jain, A.: Sketch to Photo Matching: A Feature-Based Approach. In: Proc. SPIE Conf. Biometric Technology for Human Identification VII (2010)
5. Tang, X., Wang, X.: Face Sketch Recognition. IEEE Trans. Circuits and Systems for Video Technology 14(1), 50–57 (2004)
6. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A Nonlinear Approach for Face Sketch Synthesis and Recognition. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1005–1010 (2005)
7. Zhong, J., Gao, X., Tian, C.: Face Sketch Synthesis Using E-HMM and Selective Ensemble. In: Proc. IEEE Conf. Acoustics, Speech and Signal Processing (2007)
8. Liao, S., Yi, D., Lei, Z., Qin, R., Li, S.: Heterogeneous Face Recognition from Local Structures of Normalized Appearance. In: Proc. Third Int'l Conf. Biometrics (2009)
9. Lei, Z., Li, S.: Coupled Spectral Regression for Matching Heterogeneous Faces. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1123–1128 (June 2009)
10. Klare, B., Jain, A.: Heterogeneous Face Recognition: Matching NIR to Visible Light Images. In: Proc. Int'l Conf. Pattern Recognition (2010)
11. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)

12. Frowd, C., Bruce, V., McIntyr, A., Hancock, P.: The relative importance of external and internal features of facial composites. British Journal of Psychology 98(1), 61–77 (2007)
13. Klare, B., Zhifeng, L., Jain, A.: On matching forensic sketches to mug shot photos. MSU Technical report (2010)
14. Du, G., Su, F., Cai, A.: Face recognition using SURF features. In: Proc. of SPIE, vol. 7496, pages 749628–1 (2009)
15. Gibson, L.: Limited page sample E-book copy of Forensic Art Essentials. Elsevier (2008)
16. Taylor, K.: Limited page sample E-book copy of Forensic Art and Illustration. CRC Press (2001)
17. FERET Database. NIST (2001), http://www.itl.nist.gov/iad/humanid/feret/
18. FaceVACS Software Developer Kit, Cognitec Systems GmbH (2010), http://www.cognitec-systems.de

Comparison of Configurations of Data Path Architecture Developed Using Template

B. Bala Tripura Sundari and Varsha Krishnan

Amrita Vishwa Vidyapeetham, Coimbatore-641112, India
b_ bala@cb.amrita.edu,balasrikanth2003@yahoo.com
varshkrish@gmail.com

Abstract. Data path circuits play a vital role in today's processors. Data path, which defines the structural model for interconnection of resources, needs to be flexible. It consists of a computing architecture modeled by a set of virtual templates. The focus of this work is to obtain the efficient configuration of the templates in order to map data intensive applications. The resource constrained scheduling process is employed to schedule the operations pertaining to different configurations. To achieve this scheduling, the data flow graph is modeled by grouping the resources for different configurations as a vertex cover model. Data intensive applications involve large amounts of data reuse and the number of registers and ports required is determined for each configuration. The number of registers, ports and latency are measures of efficiency of the configuration and using these measures the most efficient configuration of the templates to map an application is determined.

Keywords: Configuration, Data path, Template, Scheduling.

1 Introduction

Applications developed today require to be functional at high speed and they are computationally intensive also. This has led to reconfigurable computing gaining significance in which the configurability of software solutions with the high performance of specialized hardware architectures is aimed at [1]. High performance and flexibility are the significant requirements of any computing architecture [3]. The architectural template in a data path is required to be flexible in order to map various embedded and computationally intensive Digital Signal Processing (DSP) applications. An improved speed of execution in terms of latency measure, and reduced area in terms of lesser resources used, are aimed at in a data path in order to obtain a highly efficient architecture. This will consequently lead to a lesser time to be marketed to the embedded system developers, for whom, these features for an architecture are of high interest. Reconfigurable computing is a good compromise between flexibility and high performance. The reconfigurable devices are a good compromise between Application Specific Integrated Circuit-'s (ASIC) and the processors in terms of flexibility and high performance [9]. ASIC-'s, as the name suggests are suited for one particular application only and thus consist of the right selective functional units required for that specific application. ASIC masks are highly expensive and it takes a long time to

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 539–548.
springerlink.com

develop a custom Integrated Circuit(IC).But ASIC's consume lesser power than reconfigurable devices. Whereas processors consume much more power than reconfigurable devices and also are not very efficient in terms of execution speed, but processors have the advantage that they are flexible after fabrication and fixed set of functional units can be chosen pertaining to an application.

The proposed work involves an extensive experimentation on the different configurations of data intensive DSP applications, that are mapped onto a particular template of a data path architecture and determining an optimum configuration that best suits the application. The efficiency of each configuration is measured in terms of the registers utilized and also the execution latency.

2 Related Work

Several reconfigurable architectures with the benefits of flexibility and high performance are available. But all the architectures do not discuss various configurations involved in an application that can be mapped onto an available template. They focus majorly on how the advantages of operation level parallelism, pipelining and operation chaining are combined in order to achieve more efficiency in terms of time and area utilization which is also one of the main objective of this work using different configurations.

In Ebeling *et al* [4] model, the architecture allows pipelining and operational level parallelism in an application but pipelining is only at the inter-functional unit level. The components themselves which are pipelined are not taken care of which leads to time slacks between functional units that execute faster and slower units. But this work, which is a template based method involves provision for pipelining and, operation level parallelism, and also pipelined components can be supported on the virtual template developed. The architecture based on aggressive operation chaining [5], is a template based architecture where data path is composed of a flexible computational component (FCC), which is a pure combinational circuit and it can implement any 2×2 template (cluster) of primitive resources. Thus, the benefit of the intracomponent chaining of operations is available but the template is fixed and does not involve considering different configurations of the same application mapped onto it .In [10], although all the architectural optimization techniques like operation level parallelism, pipelining and intra operation chaining are mapped onto a single flexible architecture, only one configuration is considered for an application and efficiency of the architecture based on area and speed are compared with different architectures on the whole. This work involves considering various configurations possible on a virtual template for one particular application itself, and comparing the efficiencies among them in terms of latency and register utilization.

3 Methodology

The focus of our method is to develop a virtual architecture, which acts as the computational unit of a data path. In addition to this, the interconnection, computational unit and storage units are considered. This virtual architecture is developed with the use of templates involving adders and multipliers to perform operations in the algorithm

under consideration The application algorithms that are considered by us for mapping on to the virtual architecture are FIR filtering of different orders, IIR filtering and discrete wavelet transform. Scheduling and register allocation are then performed for each configuration of the application involved. The number of registers required and the latency involved are measured for each computation.

3.1 The Virtual Template

To map the applications involving intensive computations a virtual template is developed first using C. This template includes an array of nodes which form the resource of the architectural template. Each node also called a unified cell comprises of an adder unit and a multiplier unit together, as shown in Fig. 1.The nodes have accumulator registers in order to store temporary results for each time slot. Depending on the application involved, the desired number of nodes can be created to form the template leading to the advantage that resources are not wasted by creating more number of nodes. Each of these nodes is scheduled to perform operations based on the resource constrained scheduling process.

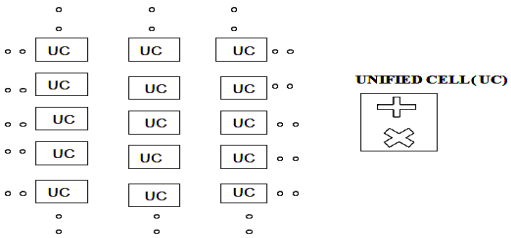


Fig. 1. Virtual architectural template with the unified cell

3.2 Scheduling and Register Allocation

Scheduling [2], [7] involves the process of assigning time slots to the operations in a particular application. The requirement for scheduling arises from the need for the processors to perform multitasking and thus the operations to be scheduled have to be chosen based on a certain priority. Scheduling operations in an application based on priority is the list scheduling. Scheduling can either be resource constrained or time constrained [6]. Here a resource constrained scheduling is employed where the number of available resources for a particular application is fixed, and then the scheduling process is carried out. Here the data flow graph for a task is first modeled as a vertex cover and time slots are assigned to this cover based on the resources available. The cover is taken in different possible ways in order to account for different configurations based on the constraint imposed on the number of available resources. The operations, after being scheduled, are allocated registers to store the temporary results which are accumulator registers. In addition, pipeline registers may also be needed in cases where the temporary results have to be utilized in another time slot grouping. The total number of registers including both the accumulator registers and pipelined registers, and the execution latency are then computed.

4 Experiment

Data intensive DSP benchmarks are considered. The template structure is generated using C and the data flow graph pertaining to the benchmark is used to give interconnections of nodes considering data input and data reuse. The execution time of one adder is 1 time unit and that of multiplier is 2 time units for the examples. The scheduling for the different configurations of the data flow graph using the vertex cover model on the template developed is done using C. The results for the number of registers obtained and the latency is computed in C. The High Level synthesis tool SPARK is used to generate the RTL code, which is synthesized using QuartusII and the number of registers is computed here also and RTL is generated. A general conclusion is drawn based on the results obtained from C and QuartusII softwares.

4.1 Finite Impulse Response (FIR) Filter

The FIR filter is a renowned digital filter having a characteristic feature that the impulse response is finite due to the absence of feedback. The FIR filters employed in DSP applications are computationally intensive with lot of data reuse and hence it is necessary to find an efficient configuration with efficiency measured in terms of registers utilized and latency of the filter to be mapped onto the virtual architecture. The FIR filter with an output y(n) with inputs x(n), x(n-1)…x(0) and coefficients w(n),w(n-1)…w(0) can be expressed as :

$$y(n) = (x(n)*w(n)) + (x(n-1)*w(n-1)\ )+\ldots \quad (1)$$

As an example, the vertex cover groups the computations of 3-tap FIR filters to be mapped onto the template developed with the nodes numbered 1 through 3 concurrently, followed by nodes 4 to 6 and then through nodes 7 to 9. The three outputs obtained are y(2), y(3) and y(4) for the inputs x(0), x(1), x(2) and x(3).These inputs are reused from one computation to the other represented by arrows in the figure. The coefficients used here are w(0), w(1) and w(2).The two different configurations of the filter are represented in Fig.2 and Fig.3, in which the resources are mapped along the computation and horizontally mapped respectively.

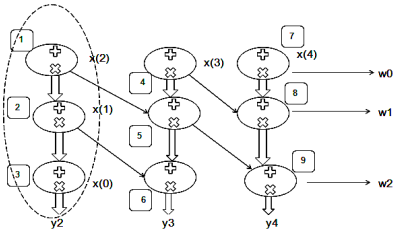


Fig. 2. Mapping along computation line

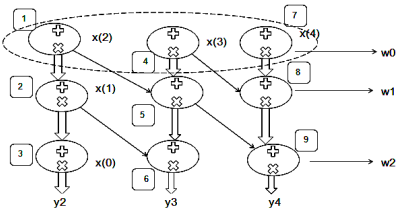


Fig. 3. Horizontal mapping

In this example it is assumed that the resources available are three adders and three multipliers. In the case of mapping along the computation line as shown in Fig.2, the vertex cover is taken along the flow of computation to determine a single output. Hence this cover, consisting of nodes 1, 2 and 3 will be considered for execution in the first time slot. Thus the rest of the computations are scheduled accordingly. There is no requirement for any pipelined register here, since the partially computed results are not carried on to the next slot and hence are independent. Whereas in the horizontal mapping as shown in Fig. 3, the vertex cover is taken in the horizontal direction consisting of nodes 1,4 and 7 leading to the necessity for pipelined registers which arises since the temporary results of one time slot are to be carried over to the next scheduled time slot.

4.1.1 Results for FIR

Both the configurations were implemented for various orders of filters and the results obtained are presented in Table 1 which is implemented using C, and Table 2 which is implemented using SPARK and QuartusII.

Table 1. FIR implemented using C

		Along Computation Line - **1** Input Port				Horizontal Mapping - **3** Input Ports			
Filter Order	Output Taken	No. of Accumulator Registers	No. of Reuse Registers	No. of Pipeline Registers	Latency	No. of Accumulator Registers	No. of Reuse Registers	No. of Pipeline Registers	Latency
3	y2,y3,y4	15	4	0	8	15	4	12	22
4	y3,y4,y5	21	6	0	11	21	6	18	31
6	y5,y6,y7	33	10	0	17	33	10	30	49

Table 2. FIR implemented using SPARK and QuartusII

		Along Computation Line			Horizontal Mapping		
Filter Order	Output Taken	No. of Registers	No. of Pins	No. of Input Ports	No. of Registers	No. of Pins	No. of Input Ports
3	y2,y3,y4	463	483	1	639	579	3
4	y3,y4,y5	656	659	1	728	659	3
6	y5,y6,y7	1427	1379	1	1717	1379	3

Table 1 shows the increase in number of input ports and latency (to obtain y2 in filter of order 3) in case of horizontal mapping and also pipeline registers are inserted. From Table 2 it can be seen that after synthesis, the number of registers required is more in case of horizontal grouping than that required in mapping along the direction of computation.

4.2 Infinite Impulse Response (IIR) Filter

The IIR filters are digital filters having a feedback from the output and their impulse response is infinite. The IIR filter is implemented in a single configuration only since it involves a feedback, other configurations are not possible due to intra iteration precedence constraints. Here, the IIR filter employed is shown in Fig. 4.The final output y4 is to be obtained.

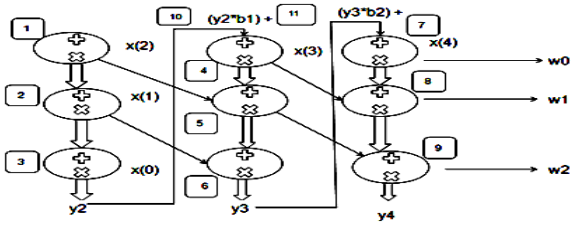


Fig. 4. IIR filter implementation

4.2.1 Results for IIR

IIR filters were implemented for various orders and the results obtained are presented in Table 3.Implementation is done using C and also SPARK and QuartusII.

Table 3. IIR implemented using SPARK, QuartusII and C

	Implementation in C				Implementation in SPARK and QuartusII		
Filter Order	No. of Accumulator Registers	No. of Reuse Registers	Latency	No. of Input Ports	No. of Registers	No. of Pins	No. of Input Ports
3	19	6	30	1	690	643	1
4	25	8	39	1	905	835	1
8	49	16	75	1	1751	1619	1

From Table 1 and Table 3 it can be observed that due to the presence of feedback, the number of registers for the same filter orders and latency are more in case of IIR than FIR.

4.3 Discrete Wavelet Transform (DWT)

The DWT is a discrete-scale ,discrete-time decomposition of finite energy sequences used to represent ferquency content as it evolves in time [8].The DWT is calculated recursively as a series of convolutions and decimations.At each octave j , an input

sequence $S_{j-1}(n)$ is fed into a low pass and a high pass filter with coefficients $G(n)$ and $H(n)$, respectively. Here n is the sample index and j is the octave index.The computation in octave j is expressed as :

$$S_j(n) = \sum_k g_k \, S_{j-1}(2n-k) \tag{2}$$

$$W_j(n) = \sum_k h_k \, S_{j-1}(2n-k) \tag{3}$$

The Fig. 5 shows the block diagram of DWT for 2 octaves. The two different configurations of the DWT are represented in Fig.6 and Fig.7 in which the resources are mapped along the computation and horizontally mapped respectively.

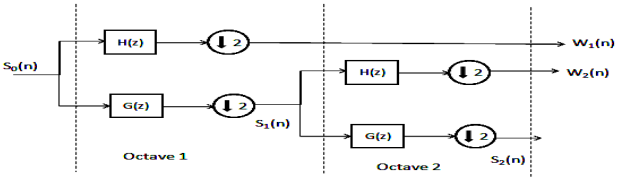


Fig. 5. Block Diagram of DWT

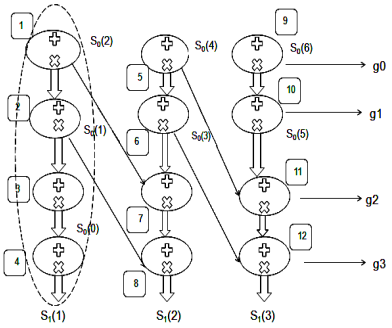


Fig. 6. Mapping along computation line

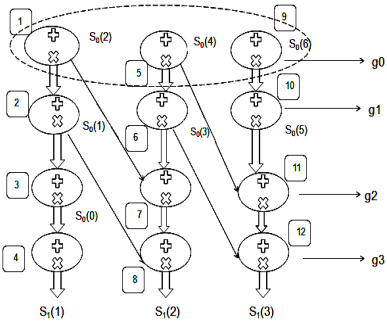


Fig. 7. Horizontal Mapping

4.3.1 Results for DWT

The Table 4 and Table 5 show the results for implementation of DWT with octave 1implemented with j=1, k=0 to 3 and varying n values.

Table 4. DWT implemented using C

	Along Computation Line - **1** Input Port				Horizontal Mapping				
n (sample) Values	No. of Accumulator Registers	No. of Reuse Registers	No. of Pipeline Registers	Latency	No. of Accumu-lator Registers	No. of Reuse Registers	No. of Pipeline Registers	Latency	No. of Input Ports
1,2,3	21	4	0	11	21	4	18	31	3
1,2,3,4	28	6	0	11	28	6	24	41	4
1,2,3,4,5	35	8	0	11	35	8	30	51	5

Table 5. DWT implemented using SPARK and QuartusII

	Along Computation Line			Horizontal Mapping		
n (sample) Values	No. of Registers	No. of Pins	No. of Input Ports	No. of Registers	No. of Pins	No. of Input Ports
1,2,3	653	675	1	725	675	3
1,2,3,4	654	691	1	726	691	4
1,2,3,4,5	754	803	1	809	803	5

Table 4 shows the latency increase in horizontal mapping and pipeline registers are also made use of. From Table 5 it is observed that in case of horizontal mapping, the number of registers used as well as the number of ports required are more as against mapping along computation line.

4.4 Comparison of Results

The results obtained for FIR,IIR and DWT are compared in terms of the number of registers utilised and the latency.

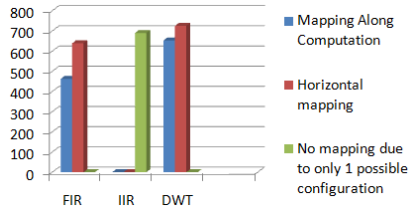


Fig. 8. Comparison of number of registers used

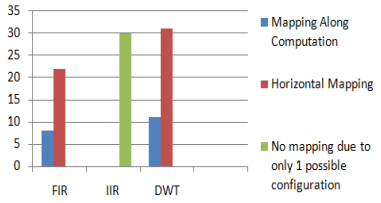


Fig. 9. Comparison of Latencies

Fig.8 shows the comparison of registers used for FIR and IIR filters of order 3 obtained from Table 2 and Table 3 respectively, and registers used for n = 1,2,3 samples for DWT from Table 5. Fig.9 shows the comparison of the latencies for FIR and IIR filters of order 3 obtained from Table 1 and Table 3 respectively, and latency for n=1, 2, 3 samples for DWT from Table 4.

5 Conclusion and Future Work

Development of a data path architecture for any computationally intensive application requires it to have a high execution speed and also the resources used must be the least possible. The computational part of the data path focused here and is developed with the help of templates. Other methods discussed earlier did not take into consideration the potential configurations for computations. The proposed work intends to include the attribute of computing the applications using different configurations. Through extensive experimentation, results explicitly show that computations performed along the direction of computation have lesser execution latency and the numbers of registers used are also lesser compared to horizontal mapping. Thus among the configurations compared, execution along the direction of computation proved to be the best.

The proposed future works is the use of the binding algorithm and extend it to bench marks of higher level nested loop algorithms. The re-configurability of the templates to run various applications can be implemented.

References

1. Compton, K., Hauck, S.: Reconfigurable computing: A survey of systems and software. ACM Comp. Surveys 34(2), 171–210 (2002)
2. Darte, A., Quinson, C.: Scheduling Register-Allocated Codes in User-Guided High-Level Synthesis. In: Proc. IEEE Int. Conf. Application-Specific Systems, Architectures and Processors, July 9-11, pp. 140–147 (2007)
3. DeHon, A., Wawrzynek, J.: Reconfigurable Computing: What, Why and Implications for Design Automation. In: Proc. 36th Design Automation Conference, pp. 610–615 (1999)
4. Ebeling, C., Cronquist, D.C., Franklin, P., Secosky, J., Berg, S.G.: Mapping Applications to the RaPiD Configurable Architecture. In: Proc. 5th Annual IEEE Symp. on FPGA for Custom Computing Machines, April 16-18, pp. 106–115 (1997)
5. Galanis, M., Theodoridis, G., Tragoudas, S., Goutis, C.: A high performance data-path for synthesizing DSP kernels. IEEE Trans. Computer Aided Des. Integr. Circuits Syst. 25(6), 1154–1163 (2006)
6. Hwang, C.-T., Lee, J.-H., Hsu, Y.-C.: A Formal Approach to the Scheduling Problem in High Level Synthesis. IEEE Transactions on Computer Aided Design 10(4), 464–475 (1991)
7. Jain, R., Somalwar, K., Werth, J., Browne, J.C.: Heuristics for scheduling I/O operations. IEEETrans. Parallel and Distributed Syst. 8(3), 310–320 (2002)
8. Parhi, K.K.: VLSI Digital Signal Processing Systems: Design and Implementation. John Wiley & Sons (1999)

9. Todman, T.J., Constantinides, G.A., Wilton, S.J.E., Mencer, O., Luk, W., Cheung, P.Y.K.: Reconfigurable computing: architectures and design methods. IEEE Proc.-Comput. Digit. Tech. 152(2), 193–207 (2005)
10. Xydis, S., Economakos, G., Soudris, D., Pekmestzi, K.: High Performance and Area Efficient Flexible DSP Datapath Synthesis. IEEE Trans. VLSI Systems 19(3), 429–442 (2011)

LDPC and SHA Based Iris Recognition for Smart Card Security

K. Seetharaman and R. Ragupathy

Department of Computer Science and Engineering,
Annamalai University, Annamalai Nagar,
Chidambaram, Tamil Nadu-608002, India
{kseethadde,cse_ragu}@yahoo.com

Abstract. We introduce a novel way to use Low Density Parity Check (LDPC) error correction code to reduce the variability and noise in iris code, which is generated from Iris Recognition System (IRS) and Secure Hash Algorithm (SHA-512) to transform codewords into hash string to make them as a Cancellable Biometric. In the enrolment process, by using majority voting scheme a unique iris code is generated from n-iris codes of IRS from n-eye samples of same person in different time interval. Then the codewords from LDPC encoding, parity check matrix of LDPC and hash string of codewords from SHA-512 are stored into smart card for high security access environment. In the verification process, iris code from live person is generated by using IRS then LDPC decoding is applied with the help of stored codewords and parity check matrix in the smart card. For authentication, new hash string, produced by employing SHA-512 on corrected iris code is compared with hash string stored in smart card. The LDPC code reduces the Hamming distance for genuine comparisons by a larger amount than for the impostor comparisons. This results in better separation between genuine and impostor users which improves the verification performance. Security of this scheme is very high due to the security complexity of SHA-512, which is 2^{256} under birthday attack. Experimental results show that this approach can assure a higher security with a low false rejection or false acceptance rate.

Keywords: Error Correction Code, Low Density Parity Check, Smart card, Secure Hash Algorithm, Iris Recognition System.

1 Introduction

In recent years, the use of iris for human identification has significantly grown due to the outstanding advantages with respect to traditional authentication methods based on personal identification numbers (PINs) or passwords. In fact, since iris is intrinsically and uniquely associated with an individual, they cannot be forgotten, easily stolen or reproduced. However, the use of iris may also have some drawbacks related to possible security breaches. Since iris characteristics are limited and immutable, if an attacker has access to the database where they are stored, the system security may be irreparably compromised. To deal with this problem, iris systems with secure

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 549–559.
springerlink.com

template storage were introduced. In these systems, irreversible cryptographic transformations, such as hash functions, are used to produce secure templates before storing them. Unfortunately, slight differences in the acquired iris data due to acquisition noise, result in a large difference in the cryptographic functions output. In these conditions, even comparisons between templates acquired from the same user will fail. To deal with this acquisition noise, an Error Correction Codes (ECC) can be used. Since the application of the ECCs has a great influence on the FRR and FAR values of the system, the choice of the code must be done carefully. In this paper, the ECCs properties which influence the performance of the system are analysed. To illustrate how these properties influence the performance of the system, LDPC codes and RS codes are used, which are two of the most commonly used ECCs in iris systems with secure template storage.

In [1], Vetro *et al.* used LDPC codes in combination with a hash function to provide secure iris template storage. Santos *et al.* [2] considered the same system architecture of [1] and proposed a universal mask which selects only the 5142 most reliable bit positions of the 9600 bits in the iris templates to enhance the security of the system. Sutcu *et al.* [3] and Nagar *et al.* [4] developed secure biometric systems based on LDPC codes and fingerprints. In [5], Argyropoulos *et al.* proposed a biometric recognition approach formulated as a channel coding problem, with LDPC codes employed for user verification. Biometric cryptosystems using the iris and RS codes are proposed in [6], [7] and [8]. In [9], Feng *et al.* developed a method to protect biometric face data on smart cards employing RS codes. Wu *et al.* [10] used RS codes and the logical XOR operation to encrypt biometric palmprint data. Kanade *et al.* [11] concatenated Hadamard and RS codes for iris template secure storage on smart cards. We use LDPC code for correcting the errors in the iris templates.

In the field of Pattern Recognition, Daugman [12] proposed an algorithm for iris recognition. Subsequently many researchers used that algorithm as a benchmark. It finds the iris in a live video image of a person's eye, defines a circular pupillary boundary between the iris and the pupil portions of the eye, and defines another circular boundary between the iris and the sclera portions of the eye. The algorithm fits the circular contours via Integro-differential operator, and normalizes the iris ring to a rectangular block of a fixed size. After that it finds a 2,048-bit iris code according to the real and imaginary parts of 2D Gabor filters outputs. By using the hamming distance, the algorithm compares the code with stored iris codes. Tisse *et al.* [13] implemented a combination of the gradient decomposed Hough transform/Integrodifferential operators for iris localization and the analytic image (a combination of the original image and its 2D Hilbert transform) to extract pertinent information from iris texture. Similar to the algorithm by Daugman, they sampled binary emergent frequency images to form a feature vector and used Hamming distance for matching. Ma *et al.* [14] adopted multi-channel Gabor filters for feature extraction and weighted Euclidean distance for matching. Ma *et al.* [15] adopted circular symmetric filters for feature extraction and a modified nearest feature line method for matching. In this paper we use the IRS proposed by K. Seetharaman and R. Ragupathy [16], which presents a novel approach on iris recognition. We use CASIAIrisV3 [17] iris database for conducting experimental tests.

Recent idea of using message digest algorithm to make cancellable biometrics enable us to use Secure Hash Algorithm (SHA). The SHA is a series of cryptographic

hash functions published by the National Institute of Standards and Technology (NIST). NIST proposed the SHA-0 as Federal Information Processing Standard Publication (FIPS PUB) 180 in 1993 [18] and announced a revised version, the SHA-1 (also called SHA-160) in FIPS PUB 180-1 as a standard instead of the SHA-0 in 1995 [19]. In 2001, the NIST published SHA as FIPS PUB 180-2 [20] consisting of four algorithms, namely SHA-160, SHA-256, SHA-384 and SHA-512. For transforming the error corrected iris code into cancellable iris code, SHA-512 is used. In this paper, SHA-512 hash is employed for authentication due to its security and uniqueness.

Conventional smart card invented in 1974 [21] has gone several development phases during the years. Today it is credit-card-sized card equipped with microprocessor, memory and input/output handler. It is a portable, low cost, intelligent device capable of manipulating and storing data. Adding individuals' unique characteristics into smart card chip, smart card becomes more secure medium, suitable for use in a wide range of applications that support biometric methods of identification. There are numerous ID systems implemented worldwide based on biometric smart card and biometric technology. One such example is UK's Asylum Seekers Card – contain photo for visual recognition and fingerprint template stored on smart card chip for biometric identification [22]. In biometric identification process we can distinguish between three types of smart card regarding their typical technical features and the type of authentication they support. The three types of smart card [23] are Template-on-card (TOC), Match-on-card (MOC), and System-on-card (SOC). In this paper, TOC type of smart card is used to store hash of SHA-512 and error corrected iris code of LDPC.

In this paper, we propose a new approach for Smart card Security built on LDPC and SHA Based Iris Recognition. The rest of the paper is organized as follows. Section 2 exhibits the Biometric Smart card technology used in this paper. The idea of enrolment process of proposed system is exhibited in Section 3. How verification process works in the proposed system is discussed in Section 4. Some experimental results and performance analysis are given in Section 5. The conclusion is drawn in Section 6.

2 Biometric Smart Card

A well-known type of smart cards is the Fun Card. The Fun card belongs to microprocessor-contact smart card. It consists of the AT90S8515 microcontroller which is a low-power CMOS 8-bit microcontroller and the AT24C64 EEPROM which provides 65,536 bits (8KB) of serial electrically erasable and programmable read only memory. The smart card programmer has been designed to enable read/write from/to the smart card. The programmer is connected to the PC using the parallel port, due to its higher speed compared with serial port and the ability to generate multiple signals at the same time. The block diagram shown in Fig. 1 consists of four parts which are signal selection circuit, voltage interfacing circuit, connection pins to the parallel port, and connection pins to the smart card. Where C1-C8 is the pins of the smart card and S0-S2 are the selecting signals. Table 1 shows the function of each pin in the used smart card. Time taken for reading and writing is very minimal for this type of card i.e. Smart card reading and writing time for 380 bits out of 8kb are 3sec and 6 sec respectively.

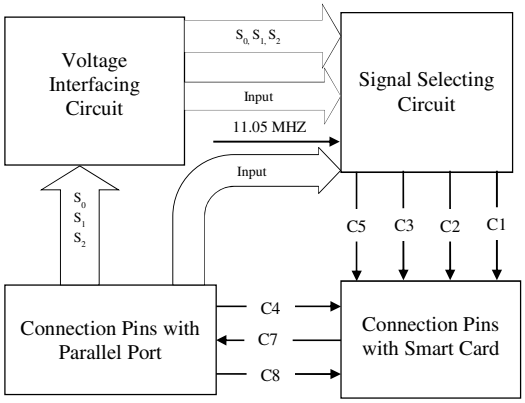


Fig. 1. Block diagram of designed programmer

Table 1. Description of each pin used in smart card

Pin No.	Name	Function	Direction
C1	Vcc	Power supply 5 VDC	In
C2	Reset	CPU Reset line	In
C3	XTAL	Main clock up to 11MHz	In
C4	MOSI	SPI master input	In
C5	Vss	Power Ground	In
C6	Nc	Not Connected	-
C7	MISO	SPI Master output	Out
C8	SCK	SPI serial clock	In

3 Enrolment Process of the Proposed System

The IRS proposed by K. Seetharaman and R. Ragupathy [16] is used for generating n number of iris codes from n number of eye samples collected from same person on different time interval. From the n number of iris code a unique iris code x is constructed by using majority voting scheme. LDPC encoding scheme operates on x and produces codewords, also called as Error Corrected Iris Code (ECIC). Parity matrix *H* of LDPC and ECIC from LDPC encoding together forms code s. Simultaneously, SHA-512 produces hash h from code x. Finally, code s and hash h from SHA-512 are stored in smart card. The entire process is depicted in Fig. 2. The novelty of IRS includes improving the speed and accuracy of the iris segmentation process, fetching the iris image so as to reduce the recognition error, producing a feature vector with discriminating texture features and a proper dimensionality so as to improve the recognition accuracy and computational efficiency. The Canny edge detection and

circular Hough transforms are used for the segmentation process. The segmented iris is normalized using Daugman's rubber sheet model from $[-32^0, 32^0]$ and $[148^0, 212^0]$. The phase data from 1D Log-Gabor filter is extracted and encoded efficiently to produce a proper feature vector.

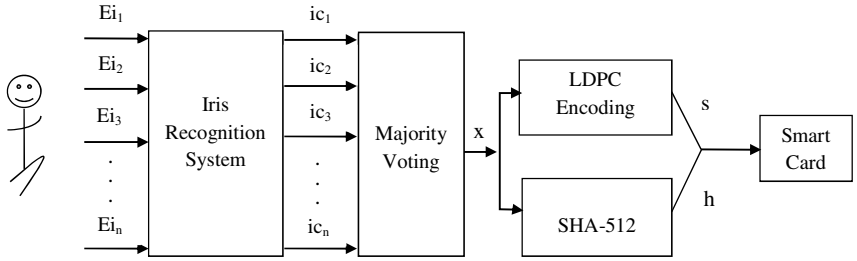


Fig. 2. Block diagram of enrolment process of the proposed system

Construction of unique iris code x from n number of iris code is done in a simple method called majority voting. Fabrication of such unique iris code x from three sample iris codes is explained in Fig. 3. From the unique code x, ECIC is formed by LDPC and hash version h is transformed by SHA-512. First, we discuss LDPC encoding and then second we deliberate SHA-512.

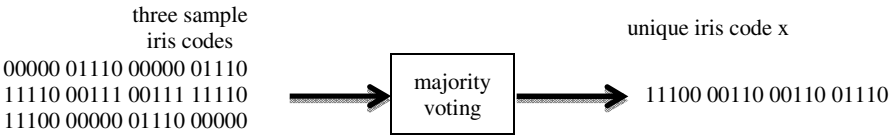


Fig. 3. Construction of unique iris code

In general, LDPC codes are defined by a sparse parity-check matrix. This sparse matrix is often randomly generated, subject to the sparsity constraints. Fig. 4 is a graph fragment of an example LDPC code using Forney's factor graph notation.

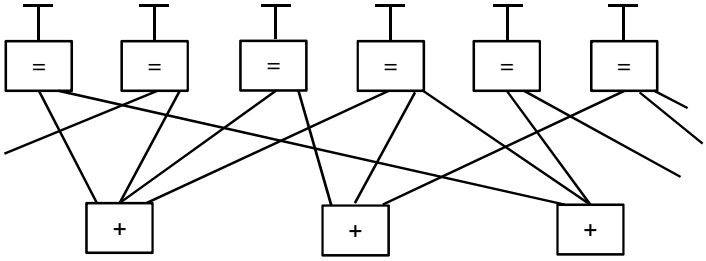


Fig. 4. Graph fragment of an example LDPC encoding

In this graph, n variable nodes in the top of the graph are connected to (n–k) constraint nodes in the bottom of the graph. This is a popular way of graphically representing an (n, k) LDPC code. The bits of a valid message, when placed on the T's at the top of the graph, satisfy the graphical constraints. Specifically, all lines connecting to a variable node (box with an '=' sign) have the same value, and all values connecting to a factor node (box with a '+' sign) must sum, modulo two, to zero (in other words, they must sum to an even number).After construction of unique iris code x, each column in the iris code x is considered as message in LDPC encoding and encoded with the help of generator matrix *G*. After forming codewords or ECIC by multiplying all columns with *G*, append the parity check matrix *H* to form code s. Following example illustrates the method of LDPC encoding. Ignoring any lines going out of the picture, there are 8 possible 6-bit strings corresponding to valid codewords (i.e., 000000, 011001, 110010, 101011, 111100, 100101, 001110, 010111). This LDPC code fragment represents a 3-bit message encoded as six bits. Redundancy is used, here, to increase the chance of recovering from channel errors. This is a (6, 3) linear code, with n = 6 and k = 3. Once again ignoring lines going out of the picture, the parity-check matrix representing this graph fragment is

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

In this matrix, each row represents one of the three parity-check constraints, while each column represents one of the six bits in the received codeword. In this example, the eight codewords can be obtained by putting the parity-check matrix *H* into this form $\left[-P^T \mid I_{n-k}\right]$ through basic row operations

$$H = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{pmatrix}$$

From this, the generator matrix *G* can be obtained as $\left[I_k \mid P\right]$ (noting that in the special case of this being a binary code $P = -P$), or specifically

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Finally, by multiplying all eight possible 3-bit strings by *G*, all eight valid codewords are obtained. For example, the codeword for the bit-string '101' is obtained by

$$\begin{pmatrix} 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Simultaneously, the unique iris code x is transformed to hash h, also called as cancellable iris code by SHA-512 message digest algorithm. SHA-512 found in FIPS PUB 180-2 documentation is adapted for this system. Sample hash h from SHA-512 for a unique iris code x is given in Fig. 5.

SHA-512 hash length: 64
SHA-512 hash value: -36 92 34 -113 105 56 -58 -1 -32 -64 86 -66 19 87 -85 -67 36 29 59 108 -91 -22 102 82 53 103 116 -1 -23 -126 -99 9 -113 -14 25 -38 -109 113 -86 -75 114 110 -28 71 109 -40 -11 70 -13 77 -94 -35 117 -86 29 62 -80 -119 -36 37 102 15 74 96
SHA-512 hash string length: 128
SHA-512 hash string:
dc5c228f6938c6ffe0c056be1357abbd241d3b6ca5ea6652356774ffe9829d098f f219da9371aab5726ee4476dd8f546f34da2dd75aa1d3eb089dc25660f4a60

Fig. 5. Sample hash h of SHA-512

4 Verification Process of the Proposed System

From eye sample collected from live person iris code $\hat{x}$ is generated by IRS. LDPC decoding scheme operates on $\hat{x}$ and produces $\tilde{x}$ with the help of parity matrix H and parity information, which are kept in code s stored in smart card. Like in enrolment process SHA-512 produces hash $\tilde{h}$ from code $\tilde{x}$. Finally, hash $\tilde{h}$ from SHA-512 and hash h from smart card is compared for authentication. This verification process is illustrated in Fig. 6.

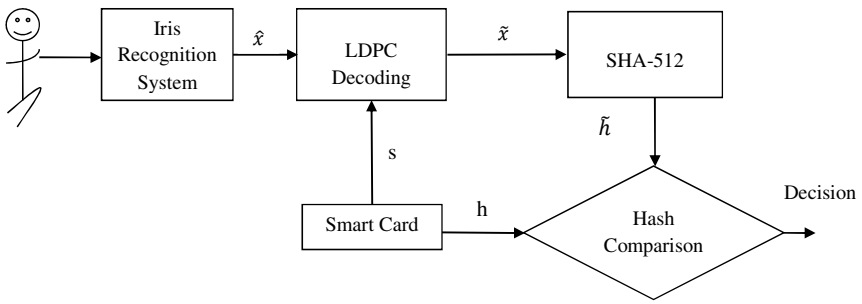


Fig. 6. Block diagram of verification process of the proposed system

To illustrate LDPC decoding, consider that the valid codeword 101011, from the example discussed in section 3. If the first bit of iris code from live person is changed then we get 001011 after appending parity. Since the iris code must have satisfied the code constraints, the iris code can be represented by writing them on the top of the factor graph. The result can be validated by multiplying the corrected codeword r by the parity-check matrix H

$$z = Hr = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

Because the outcome z (the syndrome) of this operation is the 3×1 non-zero vector, look at column 1of H which is the only equivalent to the outcome z. So flip the first bit as 1 and continue the validation. Thus, the iris code can be decoded iteratively

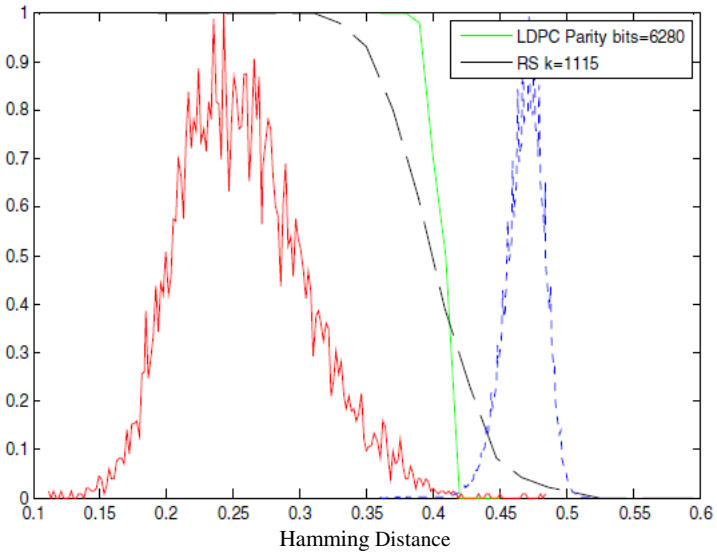
$$z = Hr = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Now, the outcome z (the syndrome) of this operation is the 3×1 zero vector, the resulting codeword r is successfully validated.

5 Experimental Results

For the public database CASIAIrisV3 [17], we choose 200 classes (eyes) and 1500 images in the subset labelled as CASIA-IrisV3-Interval. For each iris class, we choose four samples for enrolment process. In verification process, rest of the iris image in the database is compared with the other entire iris. The total number of comparisons is (1500 X 1499)/2 = 1, 124, 250, where the total number of intra-class comparisons is 7648 and that of inter-class comparisons is 1,116,602. Almost, One hundred percent correct recognition rates are obtained on CASIA-IrisV3 data sets.

To show the error correction capability of LDPC, we consider Reed Solomon (RS) code from the family of ECC. Fig. 7 shows selected curves of the RS (with k=1115) and LDPC code (with 6280 parity bits) codes overlaid on top of the genuine and impostor normalized Hamming Distance (HD) distributions. As can be easily observed, the RS correction curves are significantly less steep than the LDPC curves. Moreover, the RS code is also less granular than the LDPC. This leads to performance degradation, with False Rejection Rate (FRR) and False Acceptance Rate (FAR) values varying from 0.08% to 21.293% and from 0.014% to 57.36%, respectively. The corresponding EER value is 2.44%. But for LDPC, the resulting FRR and FAR values range from 0.754% to 1.87% and from 0.036 to 0.365%, respectively. For this situation, the estimated Equal Error Rate (EER) would be 0.41%.



Fig, 7. RS (with k=1115) and LDPC code (with 6280 parity bits) codes overlaid on top of the genuine and impostor normalized HD distributions

6 Conclusion

In IRS iris is segmented by a simple and fast technique, which is based on Canny edge detector and Hough transform and introduced the 320 normalisation method to eliminate Regions 1 type of noise i.e. obstructions due to eyelids and eyelashes in the lower and upper iris regions. Consequently, the detection time of upper and lower eyelids and 64.4% cost of the polar transformation are saved. Compared with Daugman's method, a significant decrement of the error rates is observed. LDPC codes have shown to lead to better recognition performance results than RS codes, due to the better steepness and granularity properties. Low FRR and FAR is achieved by using LDPC codes in this system. MD5 algorithm could produce identical hashes for two different messages if the initialization vector could be chosen, so we cannot adapt MD5 for authentication. As we use SHA-512, Security of this scheme is very high due to the security complexity of SHA-512 is 2256 under birthday attack. In smart card, we store cancellable iris code in the form of SHA-512 hash. As a conclusion remarks, it can be stated that, the proposed system has superior performance in terms of security, accuracy and consistency compared with other published technology.

Acknowledgments. Portions of the research in this paper used the CASIA iris image database collected by Institute of Automation, Chinese Academy of Sciences.

References

1. Vetro, A., Rane, S., Yedidia, J.: Distributed Source Coding: Theory, Algorithms and Applications. Securing Biometric Data. Elsevier (2009)
2. Santos, T., Soares, L.D., Correia, P.L.: Iris Verification System with Secure Template Storage. In: European Signal Processing Conference (EUSIPCO), Aalborg, Denmark (2010)
3. Sutcu, Y., Rane, S., Yedidia, J.S., Draper, S.C., Vetro, A.: Feature Extraction for a Slepian-Wolf Biometric System using LDPC Codes. In: IEEE International Symposium on Information Theory (ISIT), pp. 2297–2301 (2008)
4. Nagar, A., Rane, S., Vetro, A.: Privacy and Security of Features Extracted from Minutiae Aggregates. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 1826–1829 (2010)
5. Argyropoulos, S., Tzovaras, D., Ioannidis, D., Strintzis, M.G.: A Channel Coding Approach for Human Authentication From Gait Sequences. IEEE Transactions on Information Forensics and Security 4(3), 428–440 (2009)
6. Wu, X., Qi, N., Wang, K., Zhang, D.: A Novel Cryptosystem Based on Iris Key Generation. 4th International Conference on Natural Computation (ICNC) 4, 53–56 (2008)
7. Moi, S.H., Saad, P., Rahim, N.A., Ibrahim, S.: Error Correction on IRIS Biometric Template Using Reed Solomon Codes. In: 4th Asia International Conference on Mathematical/Analytical Modeling and Computer Simulation (AMS), pp. 209–214 (2010)
8. Fhloinn, E.N., Purser, M.: Iris Matching Using Error-Correcting Codes. In: IET Irish Signals and Systems Conference (ISSC), pp. 179–184 (2006)
9. Feng, Y.C., Yuen, P.C.: Protecting Face Biometric Data on Smart card with Reed-Solomon Code. In: Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 29–29 (2006)
10. Wu, X., Wang, K., Zhang, D.: A Cryptosystem Based on Palmprint Feature. In: 19th International Conference on Pattern Recognition (ICPR), pp. 1–4 (2008)
11. Kanade, S., Camara, D., Krichen, E., Petrovska-Delacretaz, D., Dorizzi, B.: Three Factor Scheme for Biometric-based Cryptographic Key Regeneration using Iris. In: Biometrics Symposium (BSYM), pp. 59–64 (2008)
12. Daugman, J.: Biometric personal identification system based on iris analysis Patent no. 5291560 (1994)
13. Tisse, C., Martin, L., Torres, L., Robert, M.: Person identification technique using human iris recognition. In: Proc. Vis. Interface, pp. 294–299 (2002)
14. Ma, L., Wang, Y., Tan, T.: Iris recognition based on multichannel Gabor filtering. In: Proceedings of ACCV 2002, the 5th Asian Conference on Computer Vision, pp. 23–25 (2002)
15. Ma, L., Wang, Y., Tan, T.: Iris recognition using circular symmetric filters. In: Proc. Int. Conf. Pattern Recognition, pp. 414–417 (2002)
16. Seetharaman, K., Ragupathy, R.: Iris Recognition for Personal Identification System. Procedia Engineering. Elsevier (accepted, 2012)
17. Chinese Academy of Sciences' Institute of Automation (CASIA): Iris image database CASIA-Iris-V3, `http://www.cbsr.ia.ac.cn/IrisDatabase.htm`
18. National Institute of Standards and Technology: Secure hash standard. Federal Information Processing Standards Publications FIPS PUB 180 (1993)
19. National Institute of Standards and Technology: Secure hash standard. Federal Information Processing Standards Publications FIPS PUB 180-1 (1995)
20. National Institute of Standards and Technology: Secure hash standard. Federal Information Processing Standards Publications FIPS PUB 180-2 (2001)

21. Rankl, W., Effing, W.: Smart card–Hand Book. Wiley &Sons, New York (1999)
22. The Industry Journal for Security & Business Professionals: Hi-Tech Security Solutions, http://www.securitysa.com
23. Yun, Y.W., Pang, C.T.: An Introduction to Biometric Match-On-Card, http://www.itsc.org.sg

A Broker Based Architecture for Adaptive Load Balancing and Elastic Resource Provisioning and Deprovisioning in Multi-tenant Based Cloud Environments

Thamarai Selvi Somasundaram[1], Kannan Govindarajan[1], M.R. Rajagopalan[2], and S. Madhusudhana Rao[2]

[1] Madras Institute of Technology, Anna University, Chennai
stselvi@annauniv.edu, kannan.gridlab@gmail.com
www.annauniv.edu/care
[2] CDAC, Chennai
mrr@cdac.in, rmadhu@cdac.in
http://www.cdac.in/

Abstract. Cloud computing is the promising technology that provides computational, storage, network and database resources by employing the virtualization technology in the infrastructure layer. Nowadays, most of the web applications are hosted in the multi-tenant based virtualized cloud environment and resource management becomes the serious and challenging task in this environment. The major goals of resource management are scalability, availability, effective utilization of the resources and increase the profit of the Cloud Service Providers (CSPs). To achieve the above objectives there is a need for a common entity that acts as a mediator between the users and CSPs. It should be capable of handling the user application requests, selection of resource, managing the life cycle of virtual instances such as creation, monitoring and deletion, balancing the load across the virtual instances and etc. in the cloud environment. In this paper, we have proposed a Cloud Resource Broker (CRB) that is facilitated with Adaptive Load Balancing (ALB) and Elastic Resource Provisioning and Deprovisioning (ERPD) mechanism. It handles the user application requests, balancing the load across the virtual instances and provisioning/deprovisioning the virtual instances in an elastic manner. The proposed work is simulated as well as tested using real-world application in Eucalyptus based private Cloud infrastructure. It increases the performance measures such as scalability and availability of the application which is running in the cloud infrastructure. The performance metrics are measured in terms of number of users access the application successfully, improved response time and etc.

Keywords: Cloud Computing, Multi-tenant, Load balancing, Scalability, Eucalyptus, Cloud Resource Broker.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 561–573.
springerlink.com

1 Introduction

Cloud Computing [1] is the combination or type of parallel and distributing computing paradigms, and it has the characteristics of on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. Cloud Computing service model is categorized into Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) and it is widely employed in the business, scientific and industrial applications. The recent trend of business, scientific and industry is deploying their web applications as virtual instances that are running in the CSP's cloud infrastructure. In this infrastructure the two extreme cases will occur, in the first case the number of users accessing the application increases that leads to increase the load and traffic of the web applications deployed in the virtual instances. In the second case, the number of users access the application drastically goes down that leads underutilization of the virtual instances. So it is essential to incorporate the mechanism to balance the load across the virtual instances and handle these two extreme cases to make sure that web application is highly available and stable. The dynamic variation in demand is satisfied by elastic resource provisioning and deprovisioning of virtual instances in an on-demand manner.

In recent years, the companies and enterprises are try to achieve the scalability in terms of application, platform, database and infrastructure level. Multitenancy [2] is the concept introduced in cloud computing in every level to solve the scalability issues. In SaaS multi-tenancy is defined as "a single application instance shared by multiple customers". In PaaS multi-tenancy is defined as "a single platform/container instance capable of handling or deploying different type of applications". In DaaS multi-tenancy is defined as "a single database and single schema that is shared by multiple organizations/tenants". In IaaS multi-tenancy is defined as "a single hardware shared by multiple users using the concept of virtualization". We have defined the evolution of multi-tenancy in IaaS as four maturity levels similar to SaaS maturity [3] level and it is shown in Figure 1.

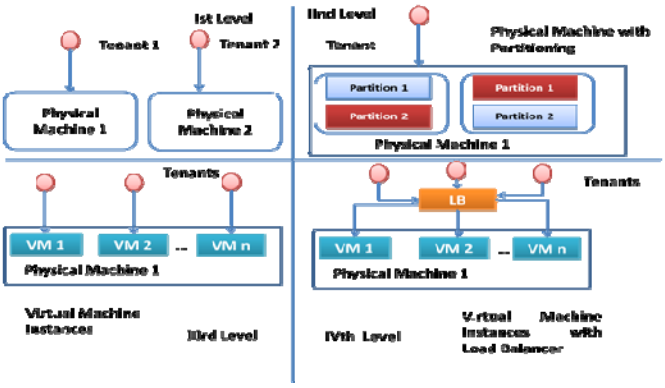


Fig. 1. Maturity Levels of IaaS

The first model provides the individual physical machine to individual tenant only single tenant can use the system. At a time only a single tenant can use one system at a time. The second model enables a physical machine to be used as two or more different systems but that cannot be concurrently used by tenants. It is possible to partition the system and each partition may be used to have different execution environments. The logical partitioning of the system is equivalent to the second configurable level of SaaS maturity level. In the third level virtualization technology plays a major role and a single hardware infrastructure is shared by multiple users. The separate virtual machine instances are created using the same physical machine and each user is running their applications in an isolated manner. In the fourth level a single physical hardware is shared by multiple users. The load balancer is introduced to balance the load and provision/deprovision the virtual machine instances based on the demand. The fourth level enables multi-tenant, scalability and efficient.

The scalability issues are classified into two types such as Horizontal scalability and Vertical Scalability. The Horizontal scalability is the process of adding or replicating the virtual instances to handle the traffic or load of a web application. Vertical scalability is the process of increasing the capacity of RAM, memory and etc. to handle the load as well as to maintain the performance levels of the application. In comparison to Horizontal Scalability is very difficult achieve dynamically. These two scalability issues should be handled by an efficient mechanism. Load Balancing (LB) is the mechanism or technique that is mainly used to balance the load across the web application servers which are running as virtual instances in the cloud infrastructure. It provides a cost-effective and easy-to-use for automatically distribute the incoming application traffic across multiple virtual instances. The main objective of the LB is to increase the scalability and availability of the web application, effective utilization of resources and improved response time.

The load balancing can be performed in two ways such as software and hardware load balancing [4]. Hardware load balancing consists of multilayer switch, Software load balancing is provided as a part of an operating system or as an application. The load balancing algorithms [5] are classified into three types such as sender-initiated, receiver-initiated and symmetric and it is classified based on the information of who has initiated the process. The load balancing policies [6] are classified into two types such as static and dynamic based on system's current state. Conventionally, when the user is accessing a web application using http request which is deployed in the application or web server running in the cloud infrastructure the request is forwarded to the Load Balancing Service (LBS). LBS finds out the IP address of the application or web server and forward the request to the respected server. But the major drawback of http request is it is employed with stateless protocol and it will not maintain the session about the clients. The existing popular open source load balancers such as Pound [7], HAProxy [8], and Nginx [9] also face the same problems and also it does not have any mechanism to maintain the state. Backspace's [10] Load Balancer and Amazon's [11] Elastic Load Balancing are employed with session affinity or session sticky feature to maintain the session about the client requests but it is made for commercial use.

Based on the above objectives we have started to develop our own open source cloud resource broker incorporated with adaptive load balancing mechanism to

provision/deprovision the virtual instances based on the traffic and load. In brief, the contributions of the research work are summarized below:

- ✓ To design and develop a cloud resource broker to handle the application request and Elastic Resource Provisioning and Deprovisioning (ERPD) of virtual instances based on demand. (A)
- ✓ To develop an Adaptive Load Balancing (ALB) mechanism to balance the across the virtual instances and distribute the application requests in the virtual instances in a balanced manner. (B)
- ✓ To explore the feasibility of incorporating the session affinity feature in the proposed architecture. (C)
- ✓ Integration of (A) & (B) to effectively manage the user application requests and increase the scalability and availability of web applications running in the cloud infrastructure.

The rest of the paper is organized as follows: Section two describes the proposed architecture and its components in detail; section three describes the proposed model, description and the proposed algorithm. Section four discusses the implementation details, real time experimental setup and execution of real-world application to test the proposed work, section five discusses the simulation results and its inferences. Section six highlights the related works closely related to our proposed work. Finally, section 7 concludes the research work and explores the possibilities of future work.

2 Proposed Architecture

The proposed cloud resource broker architecture for adaptive load balancing and elastic resource provisioning/deprovisioning of virtual instances in the multi-tenant based cloud infrastructure is represented in Figure 2.

A. Application Request Handler
The Application Request Handler component handles the user application requirements. The requirements are mainly in terms of software, hardware and QoS requirements. The software requirements are required libraries and operating system for example Red Hat Enterprise Linux 5.0 with Mysql 5.x, Jdk1.6 and Apache Tomcat 6.x server. The hardware requirements are processor speed, hard disk space, ram speed and etc. The QoS requirements are deadline, response time and etc. It matches the user application requirements with the available cloud resources and filters the potential resources that are capable of running the user application request.

B. Controller
The Controller is the core component and it invokes the appropriate components in the Cloud Resource Broker. The Controller handles the incoming/outgoing requests from/to the following components such as Application Request Handler (ARH), Cloud Scheduler (CS), Cloud Resource Provisioner (CRP), Adaptive Load Balancer (ALB), Cloud Load and Resource Information (CLRI) Aggregator.

C. Cloud Scheduler

The Cloud Scheduler selects the best resource from the matched resources that are capable of satisfying the user application requirements. The Cloud Scheduler has employed best fit algorithm and this algorithm selects the cloud resource which is having more computing power and fewer resource load.

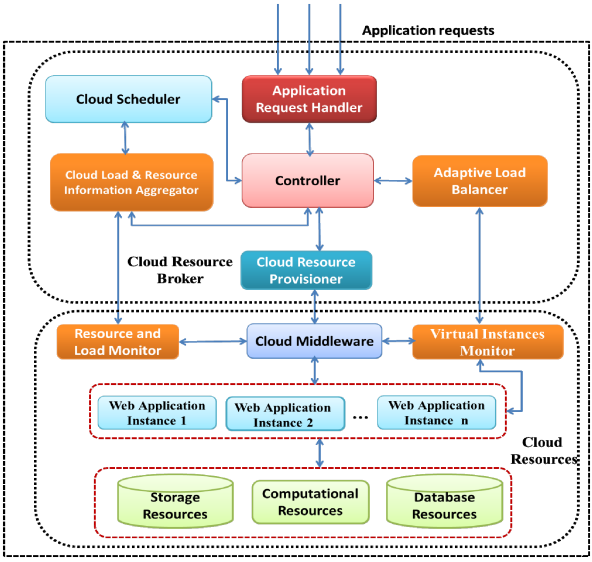


Fig. 2. High-level Architecture of Cloud Resource Broker

D. Cloud Load Balancer

The Cloud Load Balancer collects the load information of each virtual instance which is running the web application or website in the cloud infrastructure. It routes the incoming application request based on the traffic and load. It computes the average load of the virtual instances for every application requests. It maintains the threshold value for every virtual instance for each application request. If the load of the virtual instance increases above the threshold value it invokes the CRP for creating the new virtual instances. If the load of the virtual instances decreases below the threshold value it invokes the CRP for deleting the virtual instances. The pseudo code is represented in Algorithm 1.

E. Cloud Resource and Load Information Aggregator

This component is responsible for aggregating the three types of resource information such as processor, network, load and etc. from the registered CSP's. It periodically collects the information and the same has been updated in the cloud XML information repository. It interacts with Cloud Monitoring and Discovery Service (CMDS) [12] to retrieve the cloud resource information.

F. Resource and Load Monitor
This component is extended version of our earlier work CMDS to monitor the Eucalyptus based private cloud resources. It makes use of the external information providers such as Ganglia [13] to fetch the processor speed, ram memory, hard disk space and etc. NWS [14] to fetch the bandwidth, latency and etc. and our own user-defined script to retrieve the hypervisor and load information.

G. Cloud Resource Provisioner
The Cloud Resource Provisioner (CRP) is responsible for the creation and deletion of virtual instances. It interfaces with Eucalyptus Cloud Middleware using the Typica API [15] to provision/deprovision the virtual instances based on the application requirements.

H. Virtual Instance Monitor
This component is residing in the Cloud Middleware level and it gets the virtual machine ID from the cloud middleware. It monitors the traffic and load of the virtual instances that is deployed with web application

3 Proposed Model and Its Description

A cloud infrastructure is connected by group of 'm' physical servers and each server is capable of hosting 'n' Virtual Machines (VMs). Each VM requires a set of resources that includes processor speed, RAM memory, harddisk space and etc. The VMs are created on the cloud resources based on the demand and the sample application resource requests are shown in Table 1. The requests are arrived in a Poisson distribution rate of AR_i and the requests are processed in a Service rare of SR_i. The broker has the queue capacity of BQ_C. Based on the application requirements, VI_n number of virtual instances are created and it is bundled with web application, web server and database server and it is shown in Figure 3.

Table 1. Sample Application Requirements

Fields	**Requirements**
Username	tester
Number of Users/ Hr.	60 – 80
Peak Hours	9 Am – 5 Pm
Average Peak Users	1000

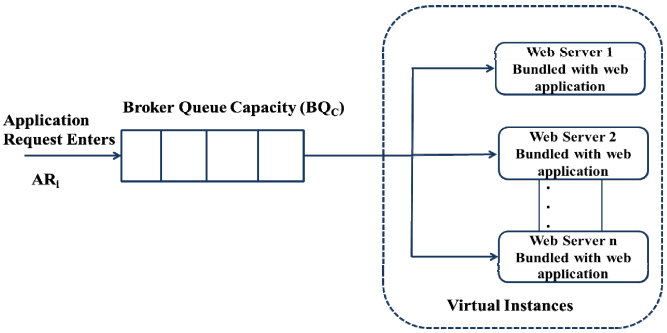


Fig. 3. Queuing System in ALB

Algorithm 1. Pseudo Code for ALB and ERPD

Input: Set of application requests to host and run the application.
Output: Create the virtual instance and run the application.
Pre-requisite: Start the resource broker and fetch the registered cloud resource information by interfacing with cloud middleware such as Eucalyptus.
Step 1 Submit the application requests with requirements, parse the requirements and store it in the Broker Queue (BQ).
Step 2 Match the application requirements with available cloud resources and select the resource that has more computing power and fewer loads.
Step 3 Create the virtual instances in the selected resource.
Step 4 Get the load information of all the created virtual instances in a periodic interval and calculate the load cost of the virtual instances using the Equation (3) and set the Threshold Value (TV) for all the instances.
Step 5 For Each Application Request

```
{
  For (I = 1 to N virtual instance)
  {
    Compute the Load (L) of each instance
Compute the Average Load (AL) of all the instances
  }
  For (I = 1 to N virtual instance)
  {
If (Load > Threshold Value (TV) ||
    Utilization of RAM > 75%)
    Create New Virtual Instance
Else If (Load < Threshold Value (TV) ||
          Utilization of RAM < 10%)
     Delete the Virtual Instance
Else (Load == Threshold Value (TV))
Forward the requests to currently running virtual Instance
  }
}
```

Step 6 End

4 Simulated Results and Discussions

In this paper first we have simulated and compared the proposed load balancing with conventional load balancing algorithms such as round robin, least loaded balancing and etc. The experiment model topologically maps into one cloud resource broker with multiple Cloud Service Providers (CSPs). The experiment has carried out with nearly 1000 cloud resources with different capabilities in the aspects of number of processors, processor speed, ram speed, hard disk memory, load, type of hypervisor and etc. The application request is generated randomly using the random access model (Feitelson) [16] model. This model generates the application parameters such as length of application (L_A), application arrival rate (A_A) and number of application requests (N_A). The application requests are generated in a random manner that generates 100 to 1000 application requests in random fashion. The performance measures of response time and throughput is shown in Figure 4 and Figure 5. The response time of the graph in Figure 4 shows that the created virtual instances handle up to 600 requests after that the response time of the requests shows some varies and it is depicted in Figure 4.

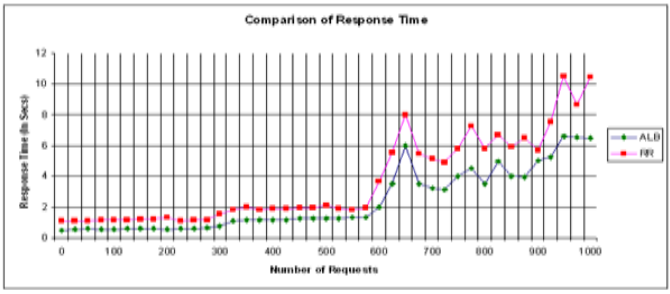


Fig. 4. Comparison of response time

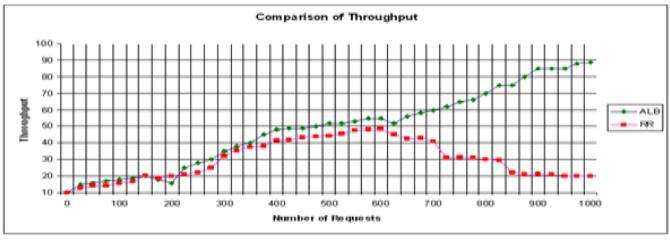


Fig. 5. Comparison of Throughput

5 Implementation Details, Experimental Setup and Real-World Application Execution

5.1 Implementation Details

The proposed work is jointly carried out and implemented by the Centre for Advanced Computing Research and Education (CARE), Anna University, and CDAC, Chennai

India. The architectural framework is implemented as REST (REpresentational State Transfer) [17] based web service and it is deployed in Sun Glassfish Server 3.0.1 and mysql 5.0. 77 used as the database. The main advantage of the RESTful web services are light-weight, stateless, and basically it works on the principle of http methods such as GET, PUT, POST and DELETE. The proposed architecture is implemented using the following services in broker level and middleware level.

- ✓ Application Request Handler (ARH) Service (Broker Level Service)
- ✓ Controller Service (Broker Level Service)
- ✓ Cloud Scheduler (CS) Service (Broker Level Service)
- ✓ Cloud Load Balancer (CLB) Service (Broker Level Service)
- ✓ Cloud Resource and Load Information (CLRI) Aggregator Service (Broker Level Service)
- ✓ Cloud Resource Provisioner (CRP) Service (Broker Level Service)
- ✓ Cloud Resource and Load Monitor Service (Middleware Level Service)
- ✓ Virtual Instance Monitor Service (Middleware Level Service)

5.2 Experimental Setup

We evaluate the performance our proposed work using real time application execution in the following experimental setup is shown in Figure 6. Our experimental setup consists of three types of cloud resources such as xeneucaserver1.care.mit.in, kvmeucaserver1.care.mit.in and xeneucaserver2.care.mit.in. The Xen [18] hypervisor based cloud resources consists of one cloud controller, two cluster controllers and each cluster controller has 10 node controllers. The cluster and node controllers are installed with Cent OS 5.2 as operating system, xen-3.2 as hypervisor with 2 GB RAM, 160 GB harddisk, 3200 MHz processor speed. The KVM [19] cluster consists of consists of one cloud controller, two cluster controllers and each have 4 node controllers. The cluster controller and node controller are installed with Fedora 12 as Operating system, KVM as its hypervisor with 2 GB RAM, 160 GB harddisk, 3.2 GHz processor speed. The CELB is installed in server hardware with 4 CPU, each CPU with quad core processors, 2000 MHz per processor, 16 GB RAM with Cent OS 5.5. The Ganglia version of 3.2.1 and NWS version of 2.13 are installed in the cloud resources and it acts as information providers to retrieve the cloud resource information.

5.3 Real-world Application Execution

As a Proof of Concept (PoC) we have tested the real-world application of web based online editor named as NebulusWain developed by our students. It is developed using HTML 5 and JavaScript. The editor is incorporated with menu bar and the tool bar, for performing operations such as save, compile, run, find and replace, cut, copy and etc. The online editor is helpful for C and C++ programmers and students to compile and run their programs. It is supported by the browsers such as Mozilla, Firefox, Safari, Chrome, Opera, Netscape navigator and etc. The online IDE has been bundled with Cent OS 5.2 operating system image, jdk 1.6.0_26 and apache tomcat-6.0.24 and it is uploaded in the eucalyptus based private cloud resources. The proposed work is tested in a particular day of 8 hours in the distributed computing laboratory.

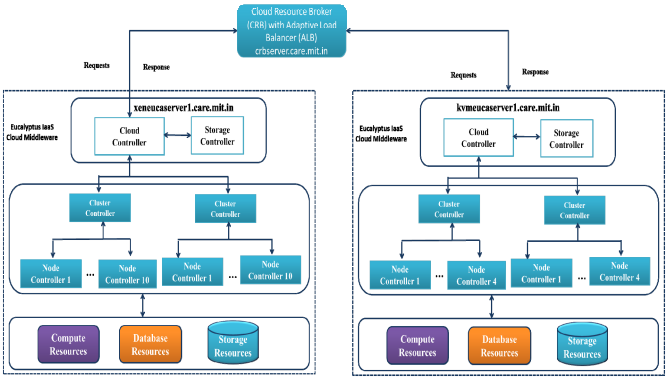


Fig. 6. Experimental Setup

Test Case Evaluation I

This test case is carried out by varying the number of users such as from 10 to 100. The number of virtual instances created for each request in the range of 1-5. The generated results are represented as graphs and it is shown in Figure 7 and 8. The Figure 7 represents the response time of the user application requests for the created virtual instances and evidently prove that the increasing number of virtual instance with web server gives better performance measures of scalability and availability. The performance metrics are measured in terms number of requests successfully handled that is throughput and the response time for each request.

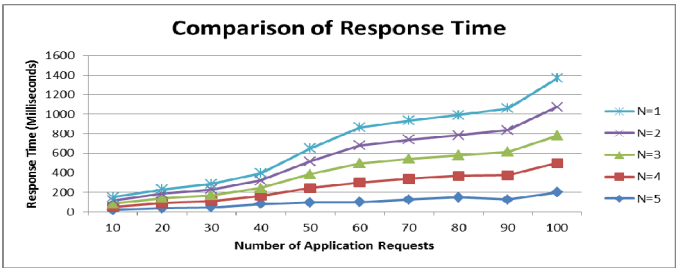


Fig. 7. Comparison of Response Time

The Figure 8 represents the throughput of the user application requests submitted to the created virtual instances.

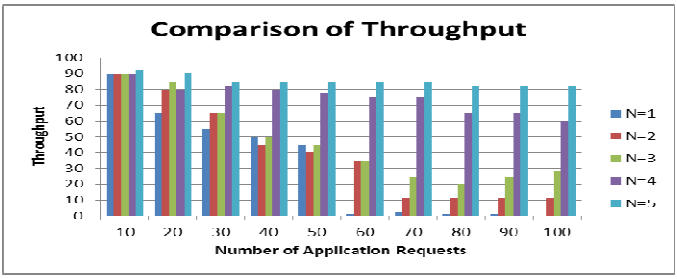


Fig. 8. Comparison of Throughput

6 Related Works

Pound [7] is a reverse proxy, load balancer and HTTPS front-end for Web server(s). It was developed to distribute the load among several web servers and to allow a convenient SSL wrapper among the web servers. HAProxy [8] is an open-source load balancer and it has the features of high availability, load balancing, and proxy for TCP and HTTP-based applications. It is particularly suited for applications require high load that needs persistence or layer 7 processing. Nginx [9] is an HTTP and reverse proxy server, as well as a mail proxy server. The popular open source load balancers [7], [8] & [9] has the support for load balancing in round-robin or DNS based techniques and it is not well suited for handling the elasticity property in cloud computing arena. But the main drawback of this technique is it distributes the requests in a round robin manner without knowing the current capacity of the server. Server Load Balancing (SLB) [20] is other popular load balancing technique and it is mainly useful in the unpredictable nature of the number of requests. SLB provides the scalability and availability to provide the needs of ever increasing server load, conventionally multiple web server instances are employed to host a website so that the load can be distributed evenly. Rackspace [7] and Amazon [8] are the major commercial providers and they provide the load balancing mechanism of replicate the virtual machines. They distribute the incoming application requests based on the demand across multiple Amazon EC2 instances through Elastic Load Balancing (ELB) capabilities. Amazon Auto Scaling [21] allows consumers to scale up or down based on the criteria of average CPU utilization across a group of compute instances. Load Balancers (LBs) should have the support for the addition of new servers in order to distribute load among several servers. In our proposed work we have incorporated the threshold based load balancing approach in the broker to distribute the load across the virtual instances.

Eucalyptus [22] is the open source cloud middleware and it consists of cloud controller, cluster controller, node controller and storage controller. It is incorporated with Greedy, Round Robin, and Power Save scheduling algorithm. It does not aware of load balancing and auto scaling property. OpenNebula (2005) [23] is an open source cloud middleware that creates virtual machines in a physical cluster and its main focus is virtual resource management in the infrastructure. It does not have the provision of load balancing capability and it is working in the infrastructure layer.

Thamarai Selvi et. al [24] has proposed and implemented a Java based architectural framework to schedule and support the virtual resource management in the Grid environment. It handles the various scheduling scenarios of Physical, Coalloc, Virtual Cluster and etc. Vauero et. al [25] has discussed the dynamically scaling in the cloud environment in terms of server, network and platform for applications. The server level scalability is achieved using elasticity controller and expresses the rules and policies for scaling the virtual instances. The platform level scalability is achieved using the concepts of replicating the container and database. The network level is achieved using the concepts of network slicing. The server level scalability mechanism helps and motivated us to develop the server level load balancing and scalability in the cloud resource broker.

7 Conclusion and Future Work

The proposed work mainly focused on developing a Cloud Resource Broker (CRB) with Adaptive Load Balancing (ALB) and Elastic Resource Provisioning Deprovisioning (ERPD) mechanism. It is developed to achieve the objectives of scalability and availability in multi-tenant based cloud environment. The proposed work is helpful for making intelligent scheduling decisions for scale-in and scale-out the virtual instances based on the traffic and load. It is simulated and the same work is implemented and tested using the real-world application of Online editor in Eucalyptus based private cloud infrastructure to provision/deprovision the virtual instances for application requests. The results have proven that the scalability and availability is improved and it is measured in terms of number of requests handled and minimization of the response time of the user requests. The main drawback of the proposed work is the load balancer does not have the session affinity feature to maintain the session between multiple virtual instances of the requests from the same user. This work is under progress.

Currently, the proposed work is implemented in a centralized mode as a future work it can be migrated to a decentralized mode that will enhance the scalability further and achieve the distributed load balancing mechanism. In addition, the future work will explore the possibilities to incorporate the other type of private clouds such as OpenNebula and etc.

Acknowledgment. The authors sincerely thank the Ministry of communication and Information Technology, Government of India, for financially supporting the Centre for Advanced Computing Research and Education of Anna University Chennai, India in this project. Also, we thank our beloved third year students namely Mr. Kiran, Ms. Aparna, Mr. Vengateshwaran, Ms. Radhika, Ms. Shrima Lakshmi, Ms. Janani, and Mr. Venkatesh for the development of web based online editor.

References

[1] NIST, National Institute of Standards and Technology (2011), http://csrc.nist.gov/publications/drafts/800-145/Draft-SP-800-145_cloud-definition.pdf

[2] Guo, C.J., Sun, W., Huang, Y.: A Framework fr Native Multi-Tenancy Application Development and Management. In: The 9th IEEE International Conference on E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and Eservices, CEC/EE 2007, July 23-26 (2007)
[3] Chong, F., Carraro, G.: Architecure Strategies for Catching the Long Tail, http://msdn.microsoft.com/en-us/library/aa479069.aspx
[4] Coulouris, G., Dolimore, J., Kindberg, T.: Distributed Systems: Concepts and Design. Addison Wesley Longman (1994)
[5] Yagoubi, B., Lila, H.T., Moussa, H.S.: Load Balancing in Grid Computing. Asian Journals of Information Technology (2006)
[6] Hardware and Software Load Balancing (2008), http://aws.amazon.com/articles/1639
[7] Pound (2010), http://www.apsis.ch/pound
[8] HAProxy. HaProxy load balancer (2011), http://haproxy.1wt.eu/
[9] Nginx. Nginx web server and load balancer (2010), http://nginx.org/en/
[10] Rackspace's Load Balancer (2012), http://www.rackspace.com/cloud/cloud_hosting_products/loadbalancers/
[11] Amazon Elastic Load Balancing (2012), http://aws.amazon.com/elasticloadbalancing/
[12] Somasundaram, T.S., Govindarajan, K.: Cloud Monitoring and Discovery Service (CMDS) for IaaS resources. Has been accepted in ICoAC (2011)
[13] Ganglia (2012), http://ganglia.sourceforge.net/
[14] NWS (2011), http://nws.cs.ucsb.edu/
[15] Typica API (2011), http://code.google.com/p/typica/
[16] Feitelson, D.G., Rudolph, L.: Metrics and Benchmarking for Parallel Job Scheduling. In: Feitelson, D.G., Rudolph, L. (eds.) IPPS-WS 1998, SPDP-WS 1998, and JSSPP 1998. LNCS, vol. 1459, pp. 1–24. Springer, Heidelberg (1998)
[17] REST (2000), http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style.htm
[18] Xen (2010), http://xen.org/
[19] KVM (2011), http://www.linux-kvm.org/
[20] Bourke, T.: Server Load Balancing. O'Reilly, ISBN 0-596-00050-2
[21] Azeez, A.: Auto-scaling web services on amazon ec2 (2008), http://people.apache.org/~azeez/autoscaling-web-services-azeez.pdf
[22] Nurmi, D., Wolski, R., Grzegorczyk, C., Obertelli, G., Soman, S., Youseff, L., Zagorodnov, D.: The Eucalyptus Open source Cloud-computing System. In: Proceedings of Cloud Computing and Its Applications (October 2008)
[23] OpenNebula: The Open Source Toolkit for Cloud Computing (2011), http://opennebula.org/start
[24] Vauero, L., Rodero-Merino, L., Buyya, R.: Dynamically Scaling Applications in the Cloud. ACM SIGCOMM Computer Communication Review 41(1) (2011)

Variation in Active Site Amino Residues of H1N1 Swine Flu Neuraminidase

G. Nageswara Rao, P. Srinivasarao, A. Apparao, and T.K. Rama Krishna Rao

Aditya Institute of Technology & Management
{gnraoaitam,peri.srinivasarao}@yahoo.com,
{apparaoallam,ramakrishnatk}@gmail.com

Abstract. In this paper, we report the variations of amino acid residues between H5N1 and H1N1 swine flu neuraminidase sequences at protein level. Random search in NCBI Flu database resulted in Canadian viral gene and analysis using blast technique revealed sites that are variant among sequences for which 3-dimensional structures were known. PDB summary database and multiple alignments were employed for validation of the results. Based on the mutations observed within active site region, homology derived model was constructed using swiss-pdb viewer. The residue variation observed was with respect to Tyr347 in H5N1 versus Asn344 in H1N1 neuraminidase sequence, which resulted in geometrical modification of ligand binding domain.

1 Introduction

Swine influenza was first proposed to be a disease related to human influenza during the 1918 flu pandemic. The H1N1 form of swine flu is one of the descendants of the strain that caused the 1918 flu pandemic [Jeffery K. Taubenberger, David M. Morens. 1918 Influenza: The mother of all pandemics. Rev Biomed 2006; 17:69-79]. The human influenza a virus continues to thrive among populations and continues to be a major cause of morbidity and mortality [Frost WH. Statistics of influenza morbidity. Public Health Rep. 1920; 35:584–97]. The virus showed various mutations [Glaser L, Stevens J, Zamarin D, Wilson IA, Garcia-Sastre A, Tumpey TM, et al. A single amino acid substitution in the 1918 influenza virus hemagglutinin changes the receptor binding specificity. J Virol. 2005; 79:11533–6] since it first originated thereby making the existing vaccines ineffective on a regular basis [Elodie Ghedin, Naomi A. Sengamalay, Martin Shumway et. al. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature 2005; 437, 1162-1166].

Influenza, commonly referred to as the flu, is an infectious disease caused by RNA viruses of the family Orthomyxviridae (the influenza viruses), that affects birds and mammals. The most common symptoms of the disease are chills, fever, sore throat, muscle pains, severe headache, coughing, weakness and general discomfort. Typically, influenza is transmitted through the air by coughs or sneezes, creating aerosols containing the virus. Influenza can also be transmitted by bird droppings, saliva, nasal secretions, faeces and blood. An avian strain named H5N1 raised the concern of a

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 575–583.
springerlink.com

new influenza pandemic, after it emerged in Asia in the 1990s. In April 2009 a novel flu strain evolved that combined genes from human, pig, and bird flu, referred as 'swine flu' [Yasushi Itoh, Kyoko Shinya, Maki Kiso et. al. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature 2009; 460, 1021-1025].

2 Materials and Methods

The viral gene sequences were accessed and extracted from NCBI (National Centre for Biotechnology Information) Flu database [www.ncbi.nlm.nih.gov]. From the H1N1 sequences deposited in NCBI, the Canadian origin neuraminidase gene (Figure 1) was selected randomly to perform sequence comparisons.

The fasta format of the sequence selected for analysis is given below.

>gi|255734960|gb|ACU31180.1| neuraminidase [Influenza A virus (A/Canada-NS/RV1554/2009(H1N1))]

MNPNQKIITIGSVCMTIGMANLILQIGNIISIWISHSIQLGNQNQIETCNQSVIT
YENNTWVNQTYVNISNTNFAAGQSVVSVKLAGNSSLCPVSGWAIYSKDNSV
RIGSKGDVFVIREPFISCSPLECRTFFLTQGALLNDKHSNGTIKDRSPYRTLMSC
PIGEVPSPYNSRFESVAWSASACHDGINWLTIGISGPDNGAVAVLKYNGIITDT
IKSWRNNILRTQESECACVNGSCFTVMTDGPSNGQASYKIFRIEKGKIVKSVE
MNAPNYHYEECSCYPDSSEITCVCRDNWHGSNRPWVSFNQNLEYQIGYICSG
IFGDNPRPNDKTGSCGPVSSNGANGVKGFSFKYGNGVWIGRTKSISSRNGFE
MIWDPNGWTGTDNNFSIKQDIVGINEWSGYSGSFVQHPELTGLDCIRPCFWV
ELIRGRPKENTIWTSGSSISFCGVNSDTVGWSWPDGAELPFTIDK

ClustalW program [www.ebi.ac.uk/clustalw] was utilized to perform multiple sequence alignments. The template 3D structures were downloaded from Protein Data Bank [www.recsb.org/pdb]. PDB summary database [www.ebi.ac.uk/pdbsum] was employed to study active site residue region. Viral neuraminidase structure was built using Swiss-PdbViewer. Initially the sequence (H1N1 neuraminidase) to be modelled is loaded from Swiss model menu and then the option move raw sequence into the structure followed by move structure into raw sequence is performed. Then the reference sequence (3CL2) is loaded from open pdb file option of the file menu and performed iterative magic fit of the fit menu by which the target sequence and the template structure fits into each other.

GenBank: GQ465699.1

Influenza A virus (A/Canada-NS/RV1554/2009(H1N1)) segment 6 neuraminidase (NA) gene, complete cds

Comment Features Sequence

```
LOCUS       GQ465699              1422 bp    cRNA    linear   VRL 11-AUG-2009
DEFINITION  Influenza A virus (A/Canada-NS/RV1554/2009(H1N1)) segment 6
            neuraminidase (NA) gene, complete cds.
ACCESSION   GQ465699
VERSION     GQ465699.1  GI:255734959
DBLINK      Project:37813
KEYWORDS    .
SOURCE      Influenza A virus (A/Canada-NS/RV1554/2009(H1N1))
  ORGANISM  Influenza A virus (A/Canada-NS/RV1554/2009(H1N1))
            Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
            Influenzavirus A.
REFERENCE   1  (bases 1 to 1422)
  AUTHORS   Bastien,N., Graham,M., Tyler,S., Van Domselaar,G., Drebot,M.,
            Plummer,F., Aranda,C.A., Zavala,E.P., Eshaghi,A., Gubbay,J.,
            Guyard,C., Guthrie,J., Duncan,C., Elgngihy,N., Tijet,N.,
            Farrell,D., Drews,S.J., Hatchette,T., Davidson,R., Sarwal,S.,
            Watson-Creed,G., Preiksaitis,J., Pabbaraju,K., Wong,S. and Li,Y.
  CONSRTM   Unknown Pathogen Investigation Collaborative Team (UPICT) and
            Instituto de Diangnostico y Referencia Epidemiologicos (inDRE)
```

Fig. 1. H1N1 Neuraminidase gene selected for analysis

3 Results and Discussion

Initially BLAST analysis was employed to evaluate the percent identities, similarities and number of gaps. Apart from this, based on PAM and BLOSUM matrices, considering Score and E-value, alignments are chosen for structure predictions.

The neuraminidase belongs to sialidase superfamily and the data from NCBI suggests that Sialidases or neuraminidases function to bind and hydrolyze terminal sialic acid residues from various glycoconjugates as well as playing roles in pathogenesis, bacterial nutrition and cellular interactions. They have a six-bladed, beta-propeller fold with the non-viral sialidases containing 2-5 Asp-box motifs (most commonly

Ser/Thr-X-Asp-[X]-Gly-X-Thr- Trp/Phe). This Conserved Domain also includes eubacterial, eukaryotic, and viral sialidases.

BLAST analysis was carried out with default parameters and the scores, top alignments are given in Figures 2 and 3.

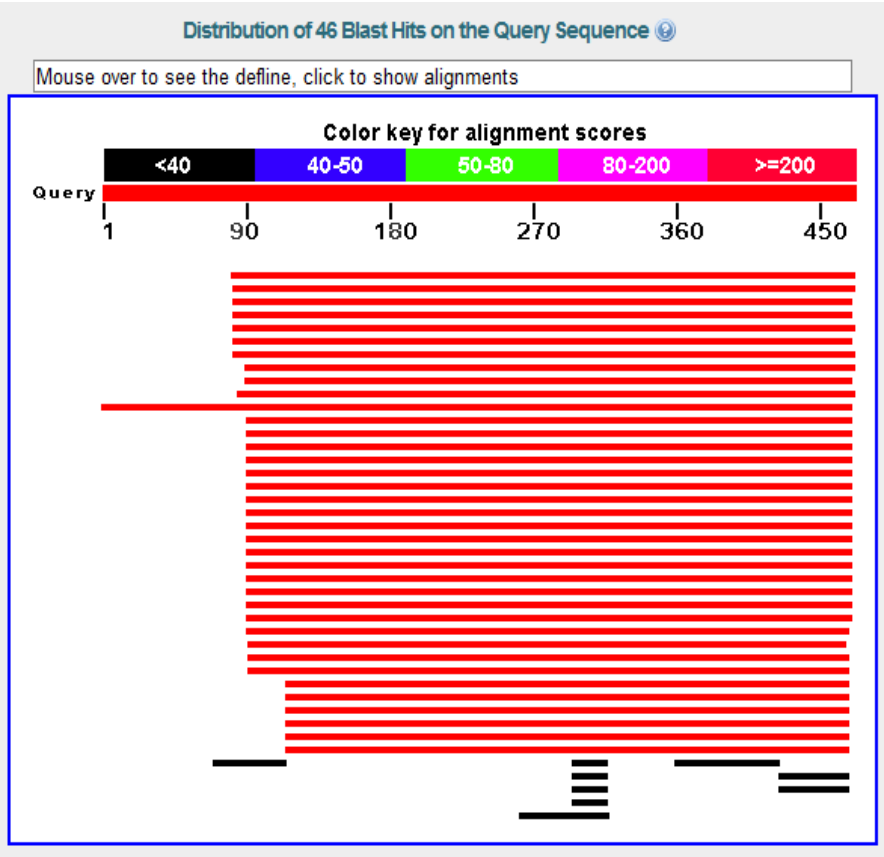


Fig. 2. BLAST analysis graphical representation

From the above top two and below alignments, although 3CL2, a H5N1 neuraminidase was considered for further work because 3CL2 was bound with oseltamivir. Therefore, structural and sequential differences between 3CL2 and H1N1 sequences were performed (Figure 4).

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
3NSS_A	Chain A, The 2009 Pandemic H1n1 Neuraminidase N1 Lacks The 150-	797	797	82%	0.0	S
2HTY_A	Chain A, N1 Neuraminidase >pdb|2HTY|B Chain B, N1 Neuraminidase	750	750	82%	0.0	S
3CL2_A	Chain A, N1 Neuraminidase N294s + Oseltamivir >pdb|3CL2|B Chain E	744	744	82%	0.0	S
3CKZ_A	Chain A, N1 Neuraminidase H274y + Zanamivir >pdb|3CL0|A Chain A,	743	743	82%	0.0	S
3CYE_A	Chain A, Cyrstal Structure Of The Native 1918 H1n1 Neuraminidase F	737	737	82%	0.0	S
3BEQ_A	Chain A, Neuraminidase Of ABREVIG MISSION11918 H1N1 STRAIN >p	733	733	82%	0.0	S
2HTV_A	Chain A, N4 Neuraminidase >pdb|2HTV|B Chain B, N4 Neuraminidase	558	558	82%	0.0	S
2HT5_A	Chain A, N8 Neuraminidase >pdb|2HT7|A Chain A, N8 Neuraminidase	468	468	81%	6e-164	S
3O9J_A	Chain A, Influenza Na In Complex With Compound 5 >pdb|3O9K|A Ch	466	466	80%	4e-163	S
3SAL_A	Chain A, Crystal Structure Of Influenza A Virus Neuraminidase N5 >p	453	453	82%	4e-158	S
1NMB_N	Chain N, The Structure Of A Complex Between The Nc10 Antibody A	382	382	99%	3e-129	S
1NNA_A	Chain A, Three-Dimensional Structure Of Influenza A N9 Neuraminida	363	363	80%	7e-123	S
1XOE_A	Chain A, N9 Tern Influenza Neuraminidase Complexed With (2r,4r,5r)-	363	363	80%	8e-123	S
1NCA_N	Chain N, Refined Crystal Structure Of The Influenza Virus N9 Neuram	363	363	80%	9e-123	S
1A14_N	Chain N, Complex Between Nc10 Anti-Influenza Virus Neuraminidase	363	363	80%	1e-122	S
5NN9_A	Chain A, Refined Atomic Structures Of N9 Subtype Influenza Virus Ne	362	362	80%	2e-122	S
1NCB_N	Chain N, Crystal Structures Of Two Mutant Neuraminidase-Antibody	362	362	80%	2e-122	S
1L7H_A	Chain A, Crystal Structure Of R292k Mutant Influenza Virus Neuramir	361	361	80%	3e-122	S
3NN9_A	Chain A, Refined Atomic Structures Of N9 Subtype Influenza Virus Ne	361	361	80%	4e-122	S
1INY_A	Chain A, A Sialic Acid Derived Phosphonate Analog Inhibits Different	361	361	80%	5e-122	S
6NN9_A	Chain A, Refined Atomic Structures Of N9 Subtype Influenza Virus Ne	361	361	80%	5e-122	S
1NCC_N	Chain N, Crystal Structures Of Two Mutant Neuraminidase-Antibody	360	360	80%	7e-122	S
4NN9_A	Chain A, Refined Atomic Structures Of N9 Subtype Influenza Virus Ne	360	360	80%	8e-122	S
1NMA_N	Chain N, N9 Neuraminidase Complexes With Antibodies Nc41 And Nc1	360	360	80%	9e-122	S
1NCD_N	Chain N, Refined Crystal Structure Of The Influenza Virus N9 Neuram	360	360	80%	1e-121	S
1L7G_A	Chain A, Crystal Structure Of E119g Mutant Influenza Virus Neuramir	360	360	80%	2e-121	S
2B8H_A	Chain A, ANWSWHALEMAINE184 (H1N9) REASSORTANT INFLUENZA V	360	360	80%	2e-121	S
1V0Z_A	Chain A, Structure Of Neuraminidase From English Duck Subtype N6	355	355	79%	1e-119	S

Fig. 3. BLAST analysis result

However, a careful observation of active site lining residues resulted in residue mutation in human H1N1 sequence. In other words, the residue variation was observed with respect to Tyr347 in H5N1 versus Asn344 in H1N1 neuraminidase sequence. Owing to the active site residue mutation, the protein model was built using SPDBV software.

An active site residue mutation was identified on comparison with H5N1 avian flu Neuraminidase enzyme. Hence, the 3D structure of H1N1 neuraminidase was built which can aid in detecting more potent binding inhibitor using computer-aided drug binding and screening studies (Figures 4-7).

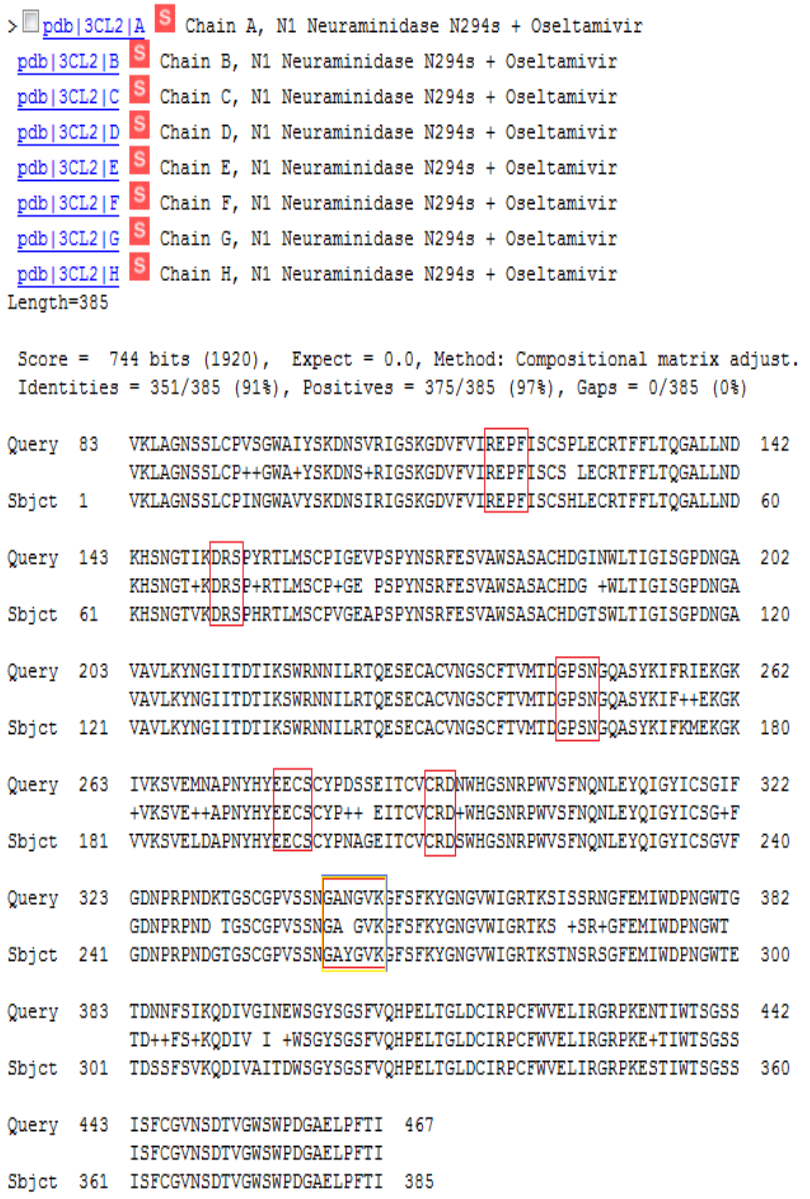


Fig. 4. Variations of amino acid residues between H5N1 and H1N1 neuraminidase sequences

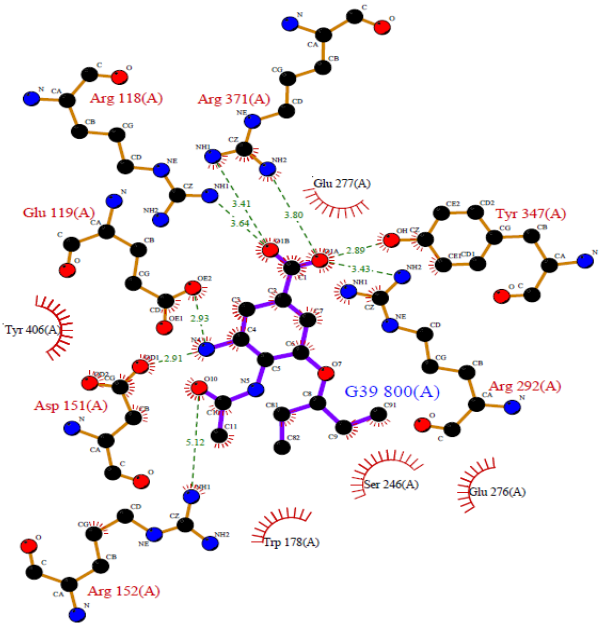


Fig. 5. Active site region of 3CL2 bound to oseltamivir

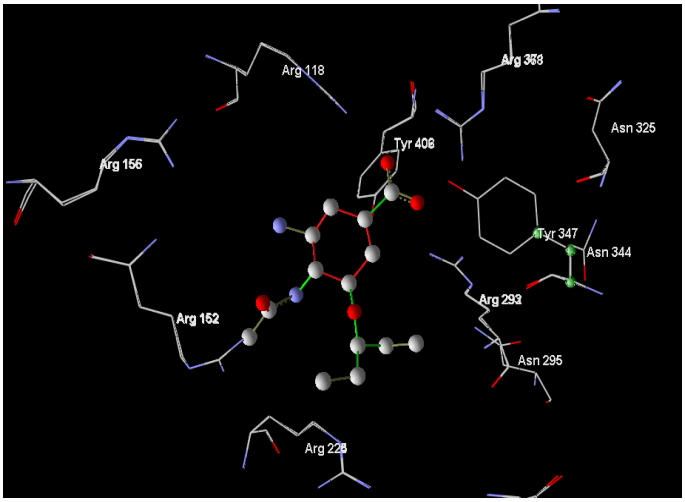


Fig. 6. Comparison of docked images of superimposed H5N1 and H1N1 acitve site regions showing Tyr347 of H5N1 replaced by Asn344 in H1N1

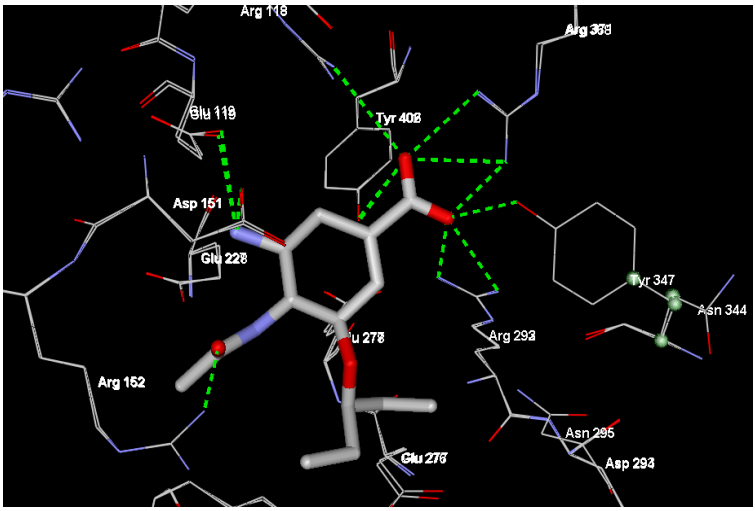


Fig. 7. Modelled protein with interacting H-bonds

Conclusion

Literature reports suggest the importance of computational tools in finding few features that are relevant and important in understanding the structure and function of various mutational events in genes or proteins. One such study reported in this paper suggested the fact that with few computational efforts, variations in amino acid residue regions within the protein sequence can be known and accurate homology models can be built within short period of time. In this work, an active site residue mutation was identified in H1N1 neuraminidase upon comparison with H5N1 avian flu Neuraminidase enzyme. Hence, the 3D structure of H1N1 neuraminidase was built which can aid in detecting more potent binding inhibitor using computer-aided drug binding and screening studies.

References

1. Olsen, C.W., Brown, I.H., Easterday, B.C., Reeth, K.V.: Diseases of swine By Straw, B.E., Taylor, D.J., Swine Influenza. ch. 28 , pp. 469–470
2. http://www.who.int
3. AlKhawaja, S.: Consultant, Infectious Disease Physician. Ministry of Health Kingdom of Bahrain Medical Bulletin 31(2), 1–4 (2009)
4. Brown, D.: System set up after SARS epidemic was slow to alert global authorities (2009), http://www.washingtonpost.com/wp-dyn/content/article/2009/04/29/AR2009042904911.html
5. Chiu, S.S., Lau, Y.L., Chan, K.H., Wong, W.H.S., Peiris, J.S.M.: Influenza-related hospitalizations among children in HongKong. N. Engl. J. Med. 347, 2097–2103 (2002)

6. Monto, A.S.: The role of antivirals in the control of influenza. Vaccine 21, 1796–1800 (2003)
7. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V., Djoumbou, Y., Eis-ner, R., Guo, A.C., Wishart, D.S.: DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. Nucleic Acids Res. 39(Database issue), D1035–D1041 (2011)
8. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E.: The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000)
9. Jones, G., Willett, P., Glen, R.C.: Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J. Mol. Biol. 245, 43–53 (1995)
10. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., Wolfson, H.J.: PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res. 33, W363–W367 (2005)
11. Chang, D.T.-H., Oyang, Y.-J., Lin, J.-H.: MEDock: a web server for efficient prediction of ligand binding sites based on a novel optimization algorithm. Nucleic Acids Research 33(suppl. 2), W233–W238
12. Schames, J.R., Henchman, R.H., Siegel, J.S., Sotriffer, C.A., Ni, H., McCammon, J.A.: Discovery of a novel binding trench in HIV integrase. J. Med. Chem. 47(8), 1879–1881 (2004)
13. Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J.: Nat. Rev. Drug. Discov. 3, 935–949 (2004)
14. Wang, R., Lu, Y., Wang, S.: J. Med. Chem. 46, 2287–2303 (2003)
15. Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P.: J. Med. Chem. 42, 5100–5109 (1999)

A Semantic Search Engine to Discover and Select Sensor Web Services for Wireless Sensor Network

Chinmohan Nayak and Manoranjan Parhi

Dept. of Computer Science & Engineering,
ITER, Siksha 'O' Anusandhan University
nayakchinmohan@gmail.com,
manoranjanparhi@iter.ac.in

Abstract. Sensor Web Services are the most emerging distributed applications and have potential usage in a wide range of application domain. The semantic based sensor service discovery is proposed to enhance the discovery of sensor services. sensor web service generates a large number of heterogeneous raw data, so it's a big challenge now-a-days to organize these raw data using various techniques so as to make the discovery and the selection easy and efficient. This paper extends the functionality of UDDI by introducing semantic description which is stored in the semantic repository at the same time the service gets registered. To provide the requested services a match maker is usually required. The match making algorithm in this paper is a generic semantic discovery algorithm which is not restricted only to the keyword based search rather is used to find the best possible services and the selection of the right service for the right user.

1 Introduction

Sensor Web Services are modular, self-describing, self-contained applications that are accessible over the Internet. This is an emerging trend which has been identified as the technology for business process execution and application integration. There are also increasing number of both publicly (external) available sensor services and sensor services only exposed internally within an organization. It is becoming a kind of mainstream middleware technology of interoperation and integration between heterogeneous applications and resource sharing in Internet environment. While considering all of these factors, it becomes a challenge for the external users or a systems to discover and invoke the sensor derived data. The current discovery mechanism supported by UDDI is not powerful enough for automated discovery. The main inhibitor is the lack of semantics in the discovery process and the fact that UDDI does not use information in the service descriptions during discovery. This makes UDDI less effective, even though it provides an interface for keyword and taxonomy based searching. The key to semantic discovery of Web services is having semantics in the description itself and then using semantic match making algorithms to find the required services.

In this paper, we develop a framework for semantically sensor web service discovery where we incorporate the semantics and integrate it with UDDI registries.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 585–591.
springerlink.com

Our aim is the discovery of Web services on a semantic comparison between a client query and available sensor services. This architecture supports both service publishing and service discovery. The discovery contribution of this paper lies in four fold. First is the direct discovery by exact matching. If this step fails, automatically the requested query for the service is matched with the semantics. Third, we use a dictionary based approach to capture real world knowledge if the second step is not successful and it will also function ate automatically if a failure occurs in the third step. The fourth and the final fold in our model will be an advanced search having its separate searching interface which will be used by a requester when he wants a particular service according to his non-functional (QoS) requirements.

Fig. 1. Overall Architecture Of Service Registry and Service Discovery

The remaining part of the paper is organized as follows. An overview of the related work is described in Section 2. Section 3 presents the proposed Framework, an effective searching algorithm for sensor web service discovery based on functional and QoS requirements, the information flow between various layers of the proposed architecture, and various parameters used in our model. Section 4 gives the implementation details of the proposed technique and finally section 5 presents the conclusion and future work.

2 Related Works

The authors in [8] have proposed an algorithm for an efficient search but it is limited to only keywords and also they have not implemented the algorithm. The authors in [2] proposed a registry for sensor network discovery and registration called Sensor Registry Service. The Sensor Registry Service is too abstract in the service oriented Sensor web because too little attention has been given to the detail functionality of the sensor registry service. In [5] a mechanism to discover sensor web registry services based on functional requirements is proposed. However nonfunctional requirements (QoS Parameters) of the services are not considered at all. In [7] a unique SOA approach is presented to design a sensor web registry that can be hosted on a special server called Sensor Name Server that cooperates and collaborates in searching a sensor network. However the author has given more emphasis on design of sensor web

registry rather than sensor discovery process. In [4] a sensor network registry is proposed and the query parameters for sensor network discovery are analyzed by 5W1H method. Here the authors have mentioned that the sensor network registry receives the discovery query using XML (XQuery). However XQuery and XPath are the advanced XML based technology which is very difficult for the novice requesters to understand. In a similar effort, the authors in [9] proposed WOOGLE, a search engine which focus on retrieving WSDN operations. Woogle (which discontinued its service in 2006), collected services from accessible services registries and provided clients with capabilities to perform keyword-based search. However, the main underlying concept behind the method implemented in woogle was based on the assumption that web services belong to the same domain of interest and are equal in terms of their behavior in accomplishing the required functionality. Other approaches focused on the semantic support for web services as presented in [6], the authors proposed a novel approach to integrate services considering only their availability, the functionalities they provide, and their non-functional QoS properties rather than considering the users direct request. In [3] the authors proposed a solution for this problem and introduced the Web Service Relevancy Function (WSRF) that is used for measuring the relevancy ranking of a particular Web service based on QoS metrics and client preferences. However one of the challenges in this work is the client's ability to control the discovery process across accessible service registries for finding services of interest, yet semantic matching of services has not been considered. The authors in [1] proposed an important concept of Static Discovery and Dynamic Selection(SDDS) approach to web service discovery but their approach is very tedious which will adversely affect the execution time resulting in a poor performance. Also they have not mentioned anything about the implementation of their system.

3 Proposed Work

Current Web service standards focus on technical Conventions. Though they solve many problems on the technical level, the semantics of Web services and Web service descriptions as a whole are not addressed by them [8]. Motivated by the increasing number of Sensor services we are here to propose a brand new approach for the service registry and discovery of these sensor services so that each and every consumers expecting some services will get the desired result always. Our approach is very similar to that of the author's in [1], we have designed an easy and efficient algorithm along with the implementation to overcome the limitations found in [1]. We propose addition of semantic data in the present WSDL by Using a separate Semantics Repository where all the semantic information of the services will be stored. Once the service is registered and the semantics are stored, the next important issue is how to find semantic similarity between the semantic annotations and different domains. Since the UDDI approach suffers from some serious bottlenecks, our proposed work provides a series of algorithms to avoid all the difficulties faced by WSDL and UDDI.

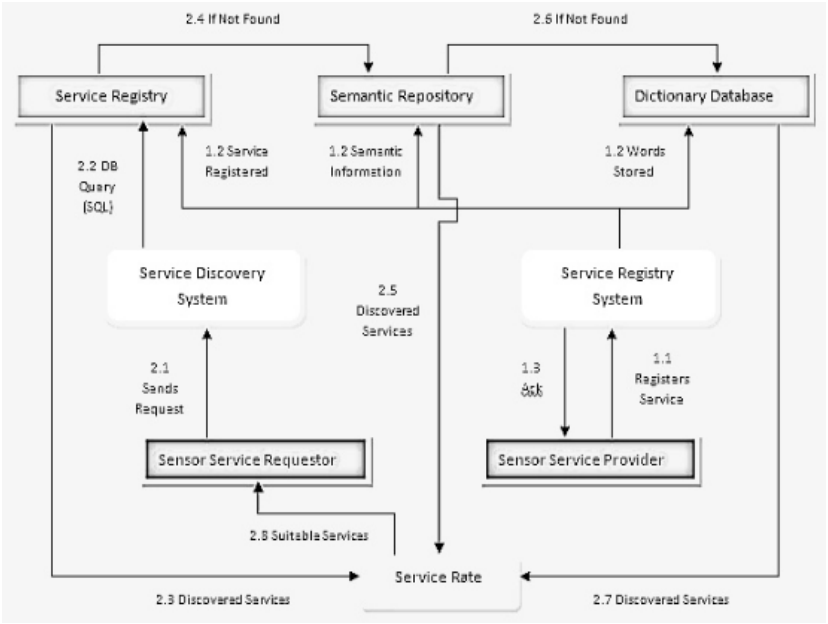


Fig. 2. Proposed Architecture for Sensor Service Discovery

The main focus of our framework is the intermediate layer called Service Discovery System (SDS) which provides interoperability between the service requester and Sensor Storage System. The Semantic Repository and Sensor Service Registry combinedly represent Sensor Storage System. Two more important operations added to the SDS of our proposed framework are Service Rate System (SRS) for finding the most suitable sensor service based on rate of service and we have also proposed an Advanced Search to find the most relevant services based on the functional and non-functional (QoS) requirements of the user. This will be the last and independent operation of the Service Discovery System. In Fig. 2, the sensor service registration process is the process no 1 which starts from the sensor service provider, so the flow is represented as 1.1, 1.2 and so on. The Sensor Service discovery is the process no 2 starts with Sensor service requester and is annotated in a similar fashion. Fig. 3 represents the sequence diagram of our proposed model.

3.1 Proposed Framework

The proposed architectural framework of our model will flow as below. A typical usage scenario is described here by considering an example in which a Wireless Sensor service provider registers his service and a consumers request for a service.

1. Sensor Service Provider: - Initially the Sensor Service Providers will register the Wireless sensor service in the Sensor Storage System and provides functional and non-functional information (QoS) about the offered services.

2. Sensor Service Requester: - Sensor Service Requester is the consumer who requests for the Sensor service discovery by providing his query for the service. The request made may be for Semantic Search or Advance Search.

i. Semantic search- In Semantic search, the user have to provide only his query as we usually do in Google.

ii. Advance search- In Advance Search, the user have to provide both query (functional) as well as the QoS (non-functional) parameters. This is a case usually found in job search interface of different company websites.

3. Service Discovery System (SDS):- The SDS will accept the request from the requester and scans the query string and applies necessary algorithms along with the SQL queries to extract all possible services from the Sensor Storage System.

4. Service Registry (SR):- Service Registry acts as an information registry for all the Sensor services which gets registered. All the functional and non-functional information provided by the Service provided will be stored in SR.

5. Semantic Repository (SMR):- Semantic Repository stores only the semantic information related to the services.

6. Dictionary Database (DD):- Dictionary approach is used in our model to capture real world knowledge if the user by mistakenly doesn't provide the query correctly.

7. Service Rate System (SRS):- After the SDS applies the extraction methods onto the Sensor Storage System, a list of Sensor Web services according to the user requirements will be returned back to SRS. The SRS then arranges the services according to their Service Rate which is nothing but the frequency of a particular service. The frequency in our context is the number of times the page is accessed. So the SRS arranges the discovered services according to their frequency in a descending order.

8. Finally the organized sensor service by the Service Rate System (SRS) is returned to the user.

3.2 Proposed Algorithm

1. Request for a desired Sensor web service
2. Split words by white space and store in the array.
3. Discover Service:
   i. Fire SQL query to perform a traditional keyword based search to find the requested Sensor web services. If there is no match found or if the user is not satisfied with the match results then go to Step **3.ii** else go to Step **4**.
   ii. Fire SQL query integrated with codes to match those array of words with that of the semantic descriptions of services stored in semantic repository database. If found go to Step **4** else go to Step **3.iii**
   iii. Fire SQL query to match the query of words with the substring of words present in the dictionary database. If found go to step **4** else displaying "No result found".
   iv. After processing the above two SQL queries, a list of sensor services are discovered from the Service registry.

4. Service Rate calculation.
   The rate of service is calculated according to the service bits.
5. Invoke appropriate service
   Most relevant sensor web service invoked and is displayed according to their service rating in descending order.

3.3 Various Parameters Used in Our Model

We should model a search engine in such a way so as to extract a satisfying result for all the requester. An efficient Sensor service system should contain the functional parameters as well as the non-functional parameters for the service discovery. The Functional Parameters used are the Sensor Service Name, Sensor Service Address, Sensor Service Description. The Non-functional Parameters used are QoS Data like Response Time (RT), Throughput (TP), Availability (AV), and Cost of Service(C).

4 Experimental Implementation

The proposed system for discovering Sensor Web service can be programmed using PHP technology which is a distributed, loosely-coupled, Platform-independent system, which runs on multiple operating systems, such as Linux, Windows, or Solaris. The Sensor Service storage system is designed using MySQL database package using which all the Sensor service providers can register and store their services.

5 Conclusions and Future Works

To avoid the serious bottlenecks in WSDL and UDDI registries we are successfully completed a system which is much more advance in dynamic service discovery and selection, with some completely distinct features. This paper presents a semantically enhanced Sensor Storage System consisting of two crucial parts, the Sensor service registry and Semantic repository. All possible diagrams and important test cases are provided for a better understanding of our model. These Test cases are most likely to occur during a Sensor service discovery. This model can be extended in multiple directions. We can integrate some security features in our model which we have not focused currently. An Auto- Suggester as well as an Auto-Previewer can be added using the implementation of Ajax in the Sensor web service page as an extra work in the future.

References

1. Pahlevan, A., Müller, H.A.: Static-Discovery Dynamic-Selection (SDDS) Ap-proach to Web Service Discovery. In: Proceedings of 7th IEEE International Conference on Web Services (ICWS 2009), pp. 769–772 (July 2009)
2. Chu, X., Buyya, R.: Service Oriented Sensor Web. In: Mahalik, N.P. (ed.) Sensor Networks and Configuration: Fundamentals, Standards, Platforms and Applications, pp. 51–74. Springer (2007) 978-3-40-37364-3

3. Al-Masri, E., Mahmoud, Q.H.: Discovering the best Web service. In: Proceedings of the 16th International Conference on World Wide Web, WWW 2007, pp. 1257–1258. ACM, New York (2007)
4. Park, J., Han, J., Kang, K., Lee, K.H.: The Registry for Sensor Network Discovery. In: Proceedings of the 12th IEEE International Conference on Engineering Complex Computer Systems (ICECCS 2007), Auckland, New Zealand, July 11-14, pp. 129–137 (2007)
5. Parhi, M., Acharya, B.M., Puthal, B.: An Effective Mechanism to Discover Sensor Web Registry Services for Wireless Sensor Network under x-SOA Approach. In: Proceedings of 2nd IEEE International Conference on Trendz in Information Science & Computing (TISC 2010), Chennai, India, December 17-19 (2010)
6. Ibrahim, N., Le Mouël, F., Frénot, S.: Mysim: a spontaneous service integration middleware for prevasive environments. In: Proceedings of the 2009 International Conference on Prevasive Services, ICPS 2009, pp. 1–10. ACM, New York (2009)
7. Pandey, K.K., Patel, S.V.: A Design of Sensor Web Registry for Wireless Sensor Networks with SOA Approach. In: Proceedings of the 1st IEEE International Conference on Computational Intelligence, Communication Systems and Networks (CICSYN 2009), Indore, India, July 23-25, pp. 247–252 (2009)
8. Nath, R., Kumar, H.: Building Software Reuse Library with Efficient Key-word based Search Mechanism. International Journal of Computing Science and Communication Technologies 2(1) (2009) ISSN 0974-3375
9. Dong, X., Halevy, A., Madhavan, J., Nemes, E., Zhang, J.: Similarity search for web services. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases, VLDB 2004, VLDB Endowment, pp. 372–383. (2004)

Prevention of Man in the Middle Attack by Using Honeypot

Mayank Tiwari[1], Tushar Sharma[1], Pankaj Sharma[1], Shaivya Jindal[1], and Priyanshu[2]

[1] Information Technology Department
[2] Electronics and Communication Engineering Department,
ABES Engineering College, Ghaziabad (U.P.), India
{mayank190590,tushsharma13,sharma1pk,
shaivyajindal,priyanshu329}@gmail.com

Abstract. In this emerging trend of Internet, wireless communication/network gained so much popularity due to its faster accessing speed and portability, but it is insecure too. The objective of this paper is to identify intruder and prevent man in the middle attack (MitM) by using mantrap/honeyd honeypot(a tool).some tools/approaches are there to cope out with problem of such type of attacks but existing technologies are not so efficient. honeypot proves boon in handling with active attacks.

Keywords: clean slate, mantrap, honeyd, fakeAP, Iframe injection, network stumbler, perl script.

1 Introduction

The Internet is full of excellent resources that describe wireless technologies, wireless threats, wireless security offerings and honeypot technologies. This paper won't cover those points, but will instead focus on the core of the subject: wireless security using honeypots. In this paper, one can suppose you know what wireless networks are, that wireless security issues certainly exist and that there are security resources called honeypots to help mitigate this threat (man in the middle attack) [7].

In previous years, many researches had been done on preventing wireless media from man in the middle attack. some examples are clean slate approach, But no one is able to provide accurate result or in the mean way proper countermeasures had been proposed. Our research is basically to provide prevention from such type of intrusion attacks and iframe injections, which cause lot of loss to an individual. In this paper, we revisit 'Man-in-the-Middle' attacks[9] and examine in detail a frightening category of MitM attacks that targets Web Applications. We will discuss in detail how an attacker can steal users' private data for any site the attacker chooses when the victim uses a public network, even though all the victim does is whether the wireless network is encrypted read the latest news headlines on a harmless site. The attack methods we will describe work regardless of or not.

We use mantrap honeypot for protecting network from intrusion attacks and malicious misuse. we generate fakeAps by using network stumbler and create a cluster of wireless links, intruder attacks on this fake network and get trapped.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 593–600.
springerlink.com

If you glance at the web site of Lance Spitzner, leader of the Honeynet Project, you'll read the definition of a honeypot : "*A honeypot is an information system resource whose value lies in unauthorized or illicit use of that resource.*" [1].

So, a wireless honeypot could simply be a wireless resource that would wait for attackers or malevolent users to come through on your wireless infrastructure.

We will first describe what a wireless honeypot could do, and then move on to addressing our related goals. Then we will focus on theoretical aspects and design possibilities, before looking at some technical examples.

2 Related Work

Microsoft suggests various rules for using public wireless networks safely. For example, the user is advised to use a firewall, not to connect to unencrypted networks, and not to submit sensitive information. These are the commonly known precautions for using a public network securely. They protect against Passive attacks, but are not enough to protect against Active ones.[8]

In the Active attack (MitM) scenario [9], a malevolent third party manipulates a response within a legitimate session in a way that tricks the client into issuing an unwanted request (unknown to the user) that discloses sensitive information.

The attacker can then apply a regular Passive attack on this information. It is important to emphasize that this is made possible by a design flaw, not an implementation error or bug[7].

1. Injecting an IFrame- Active attacks can be initiated in several ways. One way, which we will use throughout this paper, is by injecting a specially crafted object such as an IFrame. Injecting an IFrame in the response inside an HTTP session will cause the user's browser to send out a request for the SRC of the IFrame along with the site's credentials. When the victim browses a news site, for example, the attacker might return a modified attack page that is identical to the original page, except for an extra line containing a malicious, and probably invisible, IFrame.

1.1 Html Page with an Injected Iframe[6] -

<html><head> <title>World Wide News </title>
</head>

<body> World Wide News Original Content

<iframe src="http://abc.com" width="0" height="0"></iframe>

</body> </html>

When the browser renders the response, it will automatically send a request for the site specified as the SRC of the IFrameActive attack Flow-

The technique described above can be illustrated with the following attack flow.

1. The victim browses his favorite news site (a request is sent to "http://news.channel")
2. The attacker intercepts the response from "http://news.channel", and injects an IFrame with SRC set to "http://abc.com" (a site for which the victim has credentials that the attacker wants to steal).
3. The user's browser renders the IFrame object in the response and sends an automatic request to the source of the IFrame (http://abc.com) with the user credentials for the site (i.e. the user's cookies). The attacker has now obtained the user's credentials, and controls the response from http://abc.com. The attacker can now perform Passive attacks such as impersonating the victim using the cookies he has obtained, Inject arbitrary JavaScript into the response, and execute transaction on behalf of the victim.

2. Remediation- We have shown that users who follow all the commonly recommended security instructions are nonetheless vulnerable to Active attacks. However, there are some practical steps that - though they do not prevent these attacks - can limit their impact.

2.1 End Users

"Clean Slate" Approach[3] -As we have seen, Active attacks:

- Endanger even information that the user did not choose to expose during the current session
- Are persistent beyond the current session.

To avoid the risk of such attacks, whenever we use an un- trusted network we should take care to connect and disconnect with a "clean slate". When we connect there should be no sensitive information on our computer that is accessible to the browser, and at the end of the session all potentially malicious information created should be deleted. Suggested safe browsing practices:

Method 1 –

1. Before connecting to any un-trusted network, delete all browser cookies & cache files. That way there is nothing for an attacker to steal.

2. After disconnecting from any un-trusted network, delete all browser cookies & cache files. That way, if any of the cookies or cache files have been poisoned, the attack will not persist in future browsing.

Method 2 –

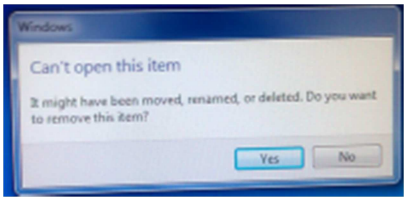
Another way of accomplishing the same thing is by always using two different browsers (not tabs or instances!), one for trusted networks and one for un-trusted networks. That way the browser used for untrusted networks will never have access to sensitive.

3 Disadvantages/Deficiency of Proposed System

As we have studied clean slate system for protecting our wireless media from MitM. We have found it is secure upto some extent but it has some defects too. Some of them are:

1. Some updates OK others not- Well a couple of days later we receive a support ticket that one of the computers was stuck at 35% update in our college lab and a day later it still was sitting at the Configuring updates 35% complete. No HD light activity. we force shutdown the unit and restarted it stopped at the same point. we fight on this and while doing that we start updating all other computers.1 of them worked fine,3 others had problems .

2. Don't log off- After installation we came back to resolve some problems with other software and while there someone sat down to the computer and when logging in to the user they went to click on an icon on the taskbar and pow a nice error popped up, "Can't open this item - It Might have been moved, renamed, or deleted. Do you want to remove this item?", and they couldn't do anything. we told them to use another computer and sat down at the computer.



The computer refused to do anything other than letting us open the Start Menu and logoff(We had changed the default to be logoff on the button since logging off is quicker than restarting and patrons are impatient just like us). we went to login as the public user again and the same problem. We then restarted and the problem went away. But after logging off and back in the problem occurred again. The long and short of it, Logging out and back in didn't work. Clean Slate wasn't doing all it needed through a log off.

This proposal (clean slate) instead of securing our system slows down the processing speed and non-updation of programs like anti-virus software makes clear holes for intruders to trapped into our system and gain unauthorized access of our application and browsing sessions.

4 Proposed System

To overcome such type of problems we suggest use of honeypot in lieu of clean slate for protecting MitM attack. Basically we use mantrap (for commercial purpose) and honeyd(for research purpose) to track intruder. When we want to track MitM attack we start honeyd [4] it starts creating dummy database, attackers will try to

scan and/or listen to wireless networks, so you may be interested in sending out fake packets, asserting the presence of wireless networks [5]. Or, you may be interested in deploying fake wireless resources dedicated to some honeypot infrastructure. A very interesting option would be to simulate traffic through the waves of your honeypot, but at this time no automatic or easy-to-use public tool has been released. One could use something like automated scripts simulating network sessions between an Access Point and its clients, as we'll see below, or use tools that replay recorded packets such as tcpreplay. Honeynet sometimes use Perl scripts that automate dialogs between clients and servers with random sessions and commands. The following example offer such automation, generating random sessions and commands that simulate wireless traffic:

```
1) #!/usr/bin/perl # initiated by
priyanshu , mayank tiwari # example
of script to simulate an automatic
FTP session # feel free to modify it
and add random activity # launch it
from your clients (use cron, etc)
use Net::FTP; $ftp = Net::FTP-
>new("192.168.16.98"); if ($ftp ==
NULL) {     print "Could not connect
to server.\n";    exit(9);        }
if ($ftp->login("barbu1",
"StEugede"))   {    $ftp-
>cwd("/home/rpm/");      $ftp-
>get("Readme.1st");      $ftp-
>quit(); } else {     print   "Could
not login.\n";     exit(7); }
```

Simulating traffic can be a more important issue on a wireless network dedicated to honeypot activity than on a wired one, because attackers need to see traffic in order to perform some of theirattacks. Bypassing 802.1X, bypassing MAC address filtering, cracking malformed WEP keys, looking at beacons, looking at SSID in the frames used for connection by clients, and so on all require existing traffic to be analyzed.

1. Wireless architectures -

You will first need at least one device that offers wireless access. If you choose to use a real Access Point, then you can safely plug it on a wired network (with at least one computer) with visible resources playing the role of targets on this fake network, and invisible resources to record data and detect intrusions (data capture). To monitor wireless- specific layer 2 attacks, one can use data capture on a wireless invisible client in mode Monitor, using software such as Kismet. An example architecture is shown below in Figure 1:

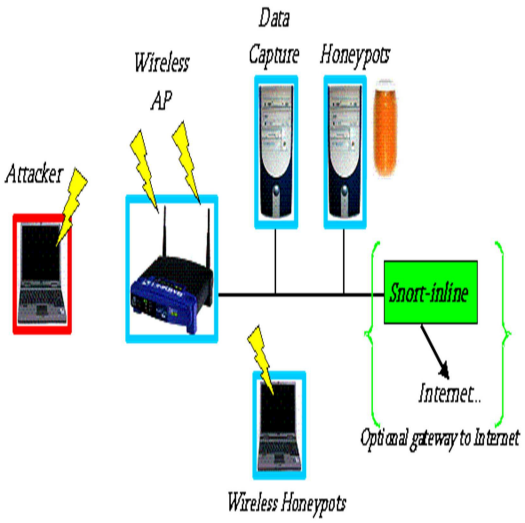


Fig. 1. Sample WiFI honeypot architecture

2. Tracking intruders –

Simulating a wireless AP -

One other interesting possibility of Honeyd [4] is the creation of fake TCP/IP stacks to fool remote fingerprinting tools such as nmap or xprobe, and this is an easy way to create your own fake services. For example, by copying well-chosen web pages used to manage an access point, one could really simulate an AP. This technique can be useful to monitor attackers who would try to connect to the management interface using well-known default passwords, or who would try other opened services (such as attacks over DNS, TFTP, etc).

For example, here is a quick test that could be tried on a laptop with a wireless card turned in Master mode and Honeyd listening on it. Suppose you want to simulate a Linksys WRT54 Access Point with a web server used for administration. Just ask Honeyd to simulate this stack and web server, as follows:

```
 create Linksys  set linksys
personality " Linux Kernel 2.4.0
 - 2.5.20"  add linksys tcp port
80 "/bin/sh
scripts/fakelinksys.sh"   add
linksys udp open 53 open add
linksys udp open 67 open add
linksys udp open 69 open set
linksys tcp action reset bind
192.168.1.1 linksys
```

FakeAP -

If you remember the movie called War Games, the young adolescent was using a modem on the phone line to scan remote phone numbers and find open lines like BBSes. This activity was called wardialing, and by transposition in the wireless world, people talking about wireless scanners or wireless listeners as wardriving, or even warwalking. Wardrivers try to find open networks. A good first idea to delude those potential intruders would be to simulate as many fake networks as possible for them to lose time and patience. Targeting one network is an easy task, whereas dealing with a cloud of targets could be more difficult.

This proof of concept was done with a tool called FakeAP [5]. This tool can send specific wireless network traffic to fool basic attackers. As a wardriving countermeasure, it generates 802.11b beacon frames as fast as possible, by playing with fields like BSSID (MAC), ESSID, channel assignments, and so on. This trick is easily created by playing with the tools used to manage a wireless card (under Linux, that's like manually playing with: iwconfig eth1 ESSID Random SSID channel N...). A remote, passive listener should then see thousands of fake access points!.The idea behind this simple tool was quite good when it was first released, and we could even detect NetStumbler users by looking at 802.11b probe requests/responses. Whereas now, most updated tools can advise the attacker that the detected access points are unusually strange, such as these cases where no traffic is generated on the found networks. Figure 2, below, indicates a Net Stumbler scan on one of these honeypots:

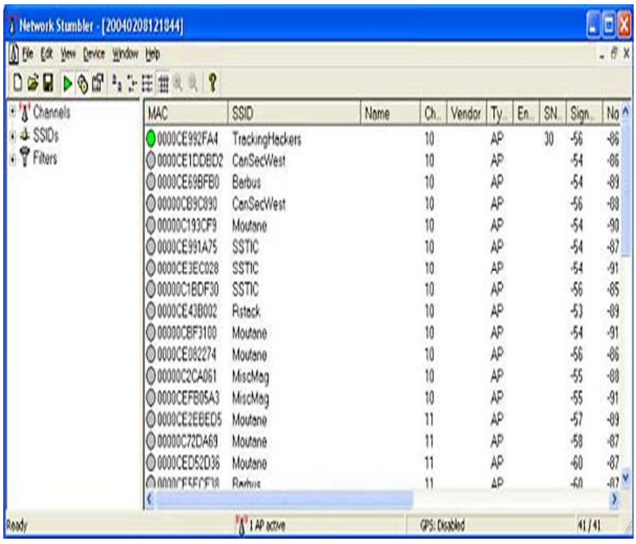


Fig. 2. NetStumbler scan on a FakeAP honeypot

3. MANTRAP- Produced by Recourse Mantrap [2] is a commercial honeypot. Instead of emulating services, Mantrap creates up to four sub-systems, often called 'jails'. Security administrators can modify these jails just as they normally would with any operating system, to include installing applications of their choice, such as an Oracle database or Apache webserver. This makes the honeypot far more flexible, as it can do much more. The attacker has a full operating system to interact with, and a variety of applications to attack. All of this activity is then captured and recorded. Not only can we detect port scans and telnet logins, but we can capture rootkits, application level attacks, IRC chat session, and a variety of other threats.

5 Conclusion and Future Work

By proposing this paper we mean to secure our wireless networks by MitM attack by use of young technology called honeynet/mantrap honeypot.It ensures secure transaction, sessions or browsing over Wireless media.it is costless and definitely prove a boon to network security. We can also secure e-commerce sessions and virtual private network by using of this proposal. It will also be implemented to mobile devices to protect m-commerce sessions and definitely if this technology works then no doubt more user trust on the secureness of wireless media.

References

[1] Lance Spitzner's web site, http://www.trackinghackers.com
[2] Martin, W.W.: Honey Pots and Honey Nets-Security through Deception. In: CISSP, May 25 (2001)
[3] http://www.fortresgrand.com/products/cls/cls.htm
[4] Honeyd project, by Niels Provos: wireless honeypots examples, http://honeyd.org/configuration.php
[5] FakeAP tool, by BlackAlchemy, http://www.blackalchemy.to/project/fakeap/
[6] Google Technical Report provos-2008a on All Your iFRAMEs Point to us by Niels Provos, Panayiotis Mavrommatis Moheeb, Abu Rajab, Fabian Monrose
[7] Man in the middle attack demos by Alberto ornaghi, Marco valleri, blackhat confrence usa (2003)
[8] Microsoft Public Wireless Networks-How To stay Safe, http://www.microsoft.com/protect/yourself/Mobile/publicwireless.mspx
[9] SSL Man-in-the-Middle Attacks, peter Burkholder (v2.0), February 1 (2002)

Slicing of Programs Dynamically under Distributed Environment

Santosh Pani[1], Shashank Mouli Satapathy[2], and G.B. Mund[3]

[1] School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, Odisha, India
santosh_pani@hotmail.com
[2] Department of MCA
Raajdhani Engineering College
Bhubaneswar, Odisha, India
shashankamouli@gmail.com
[3] School of Computer Engineering
Kalinga Institute of Industrial Technology
Bhubaneswar, Odisha, India
mundgb@yahoo.com

Abstract. A dynamic program slice is the part of a program that affects the computation of a variable of interest during program execution on a specific program input. Dynamic slices are usually smaller than static slices and are more useful in interactive applications such as program debugging and testing. The understanding and debugging of multithreaded and distributed programs are much harder compared to those of sequential programs. The nondeterministic nature of multithreaded programs, the lack of global states, unsynchronized interactions among processes, multiple threads of control and a dynamically varying number of processes are some of the reasons for this difficulty.

Different types of dynamic program slices, together with algorithms to compute them have been proposed in the literature. Most of the existing algorithms for finding slices of distributed programs use trace files and are not efficient in terms of time and space complexity. Some existing algorithms use a dependency graph and traverse the graph when the slices are asked for, resulting in high response time. This paper proposes an efficient algorithm for distributed programs. It uses control dependence graph as an intermediate representation and generates the dynamic slices with fast response time.

Keywords: Program slicing, Dynamic slice, Control Flow graph, Debugging, Distributed programming, Message Passing, Active Concurrent Slice.

1 Introduction

Program slicing is a well known decomposition technique for extracting the statements of a program related to a particular computation. A slice of a program P can be constructed with respect to a slicing criterion. A slicing criterion is a tuple <s, V> where s is a program point of interest and V is a subset of the program's variables

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 601–609.
springerlink.com

used or defined at s. The slice can be obtained by deleting the statements from the program P which have no effect on any of the variables in V as execution reaches statement s. There are two types of slices depending on the input to the program: static slice and dynamic slice. A static slice of a program P with respect to a slicing criterion <s,V> is the set of all the statements of program P that might affect the slicing criterion for every possible inputs to the program. In contrast, a dynamic slice contains only those statements of program P that actually affect the slicing criterion for a particular set of inputs to the program. Hence a dynamic slice is smaller in size and more useful for interactive application like program testing and debugging.

Now-a-days most of the application programs are distributed in nature and run on different machines connected to a network. The emergence of message passing standards, such as MPI, and the commercial success of high speed networks have contributed to making message passing programming common place. Development of real life distributed programs presents formidable challenge to the programmer so as to the debugging and testing process.

Any dynamic slicing algorithm to be useful in a distributed environment, the construction of slices should be made in a distributed manner. Each statement in a distributed system should contribute to the slice by determining its local portion of the global slice in a fully distributed fashion.

Weiser (1982) [9] introduced the concept of a static program slice and presented the first intraprocedural static slicing algorithm. His method used a Control Flow Graph (CFG) as the intermediate representation of the program, and was based on iteratively solving data-flow equations representing inter-statement influences. This algorithm did not handle programs having multiple procedures. Korel and Laski [8] extended Weiser's CFG based static slicing algorithm to compute dynamic slices. Their method computes dynamic slices by solving the associated dataflow equations. The method of Korel and Laski needs O(N) space to store the execution history, and $O(N^2)$ space to store the dynamic flow data, where N is the number of statements executed (length of execution) during the run of the program. Larson and Harrold were the first to consider object orientation aspects in their work. They introduced the class dependence graph which can represent a class hierarchy, data members, inheritance and polymorphism. They have constructed the system dependence graph (SDG) using the class dependence graphs to satisfactorily represent object oriented programs. Larson and Harrold have reported only a static slicing technique for sequential object-oriented programs, and did not address the concurrency and dynamic slicing aspects. Zhao, Song and Huynh, Wang et al. and Xu and Chen have addressed the issues of dynamic slicing of object-oriented programs, but they have not addressed the concurrency issues in object-oriented programs. Mohapatra et al. [4] have proposed an algorithm which uses a modified program dependence graph i.e. distributed program dependence graph (DPDG) for intermediate representation of programs. They have extended the basic techniques of the edge marking dynamic slicing algorithm of Mund et al. (2003) [6] to find out the distributed dynamic slices of a multithreaded java program thereby increasing the response time. Mund et al. (2006) [3] present an efficient interprocedural dynamic slicing algorithm for structured programs. They propose an intraprocedural dynamic slicing algorithm, and subsequently extend it to handle interprocedural calls. The interprocedural dynamic slicing algorithm uses a collection of control dependence graphs (one for each procedure) as the intermediate

program representation, and computes precise dynamic slices. The proposed interprocedural dynamic slicing algorithm is more efficient than the existing dynamic slicing algorithms with faster response time. We use the basic concepts of Mund et al. [3] algorithm and propose an efficient algorithm for dynamic slicing of distributed programs computed in a distributed manner with fast response time.

The rest of the paper is organized as follows. In next section, we describe some basic definitions that are used by our algorithm. The dynamic distributed slicing algorithm is discussed in the next section followed by the analysis of the algorithm and comparison with related work. The next section concludes the paper.

2 Basic Concepts and Definitions

This section describes some basic notations and definitions that are used in our algorithm. Some of them are already available in Mund et al. [3]. We present them here for the sake of completeness.

2.1 Control Flow Graph

The control flow graph (CFG) G of a program P is a graph G = (N,E), where each node n Є N represents a basic block of statements in the program P. For any pair of nodes x and y, (x,y) Є E iff there is possible flow of control from x to y. This Control Flow Graph can be used to extract control dependency that can exist among statements in a program.

2.2 ControlDependentOn(u)

Let u be a statement of the program P. ControlDependentOn(u) = s iff the statement u is control dependent on s.

2.3 ActiveControlSlice(s)

Let s be a test statement (predicate statement)of a program P and UseVarSet(s) = {var1. . . vark}. Before execution of the program P, ActiveControlSlice(s) = Φ. After each execution of the statement s in an actual run of the program, ActiveControlSlice(s) = {s} U ActiveDataSlice(var1) U …U ActiveDataSlice(vark) U ActiveControlSlice(t), where ControlDependentOn(s)= t. If s is a loop control statement, and the present execution of s corresponds to exit from the loop, then ActiveControlSlice(s) = Φ.

2.4 ActiveConcurrentSlice

ActiveConcurrentSlice is updated with every type of interaction that takes place between multiple machines in a distributed system e.g. Send and Receive, Lock and Unlock permissions. In case of communication between statements, the slicer computes the ActiveConcurrentSlice for the sender and sends it to the slicer at the receiver end. This ActiveConcurrentSlice received by the receiver slicer helps in updating it's own ActiveConcurrentSlice.

Table 1. A distributed system with two machines

Machine A	Machine B

main()
{
….
…..
10. send(p);
….
…
}
	main()
	{
	….
	…..
	5. receive(d);
	….
	…
	}

Let Uactive be the statement for interaction in a machine then,

For Machine A
ActiveConcurrentSlice = Uactive U ActiveConcurrentSlice U ActiveControlSlice(t)
where ControlDependentOn(Uactive) = t.

For Machine B
ActiveConcurrentSlice = ActiveConcurrentSlice U ActiveConcurrentSlice(Received from A)
ActiveDataSlice(d) = ActiveDataSlice(p) U ActiveConcurrentSlice(B) U ActiveControlSlice(t)

2.5 ActiveDataSlice(var)

Let var be a variable in a program P. Before execution of the program P, ActiveDataSlice(var) = Φ. Let u be a Def(var) node, and UseVarSet(u) = {var1, . . ., vark}. Consider an actual run of the program with a given set of input values. After each execution of the node u in the actual run of the program, ActiveDataSlice(var) = {u} U ActiveDataSlice(var1) U…U ActiveDataSlice(vark) U ActiveControlSlice(t), where ControlDependentOn(u) = t.

2.6 DyanSlice(machineno, s, var)

Let s be a node of the CDG GP of a program P having identified by machineno, and var be a variable in the set DefVarSet(s) U UseVarSet(s). Before execution of the program P, DyanSlice(machineno, s, var) = Φ. Consider an actual run of the program with a set of given input values. After each execution of the node s in the actual run of the program, the dynamic slice DyanSlice(machineno, s, var) with respect to the slicing criterion <s, var> identified by machineno corresponding to the execution of s is updated as DyanSlice(machineno, s, var) = ActiveDataSlice(var) U ActiveControlSlice(t), ControlDepenedentOn(u) = t.

2.7 ActiveCallSlice

Let P be a multi-procedure program. Before execution of the program, ActiveCallSlice = Φ. Consider an actual run of the program with a given set of input values. At an instance of the actual execution of the program, let Uactive represent the active call statement. Then ActiveCallSlice = {Uactive} U ActiveCallSlice U ActiveControlSlice(t), where ControlDependentOn(Uactive)= t.

2.8 CallSliceStack

A stack called CallSliceStack is used to store a relevant sequence of ActiveCallSlices during an actual run of the program. During execution of the program the top element of the stack always represents the ActiveCallSlice. Before execution of each call statement, the ActiveCallSlice corresponding to the execution of the call statement is computed and pushed onto the stack CallSliceStack.

2.9 ActiveReturnSlice

Let P be a structured multi-procedure program. Before each execution of the program P, ActiveReturnSlice = Φ. Let x be a RETURN statement in GP, and UseVarSet(x) ={var1, . . ., vark}. Then, before each execution of the RETURN statement x, ActiveReturnSlice = {x} U ActiveCallSlice U ActiveDataSlice(var1) U…U ActiveDataSlice(vark) U ActiveControlSlice(t), where ControlDependentOn(x)= t.

2.10 Formal(x, var), Actual(x, var)

Let P1 be a procedure of a program P having multiple procedures, and x be a calling statement to the procedure P1. Let f be a formal parameter of the procedure P1 and its corresponding actual parameter at the calling statement x be a. Formal(x,a) = f and Actual(x, f) = a. Note that Formal(x,a) = f iff Actual(x, f) = a.

3 Algorithm for Finding Dynamic Distributed Slicing

3.1 Algorithm (For Each Machine)

1. Construct the CFG GP of the program P statically only once.
2. Do the following before each execution of the program.
For each statement u of program P do the following
If u is a test (predicate) statement, then ActiveControlSlice(u) = Φ .
Update ControlDependentOn(u).
For each variable var Є DefVarSet(u) U UseVarSet(u) do
DyanSlice(Pid,u, var) = Φ.
For each variable var of the program P do
ActiveDataSlice(var) = Φ.
CallSliceStack = NULL.

ActiveCallSlice = Φ

ActiveConcurrent Slice= Φ.

3. Run the program P with the given set of input values, and repeat steps 4, 5, 6 and 7 until the program terminates.

4. Do the following before execution of each call statement u.

Let u be a call statement to a procedure Q.

(a) Update CallSliceStack and ActiveCallSlice.

(b) For each actual parameter var in the procedure call Q do

ActiveDataSlice(Formal(u,var)) = ActiveDataSlice(var) U ActiveCallSlice.

5. Do the following before execution of each RETURN statement u.

Update ActiveReturnSlice.

6.Do the following before execution of each concurrent statement u of the program P

(a)If u is a send (var) statement where data is a variable of program P then

Update ActiveConcurrentSlice and send it along with ActiveDataSlice(var) to the recipient machine in the distributed system.

(b)If u is a receive(var) statement then Update ActiveConcurrentSlice and Active-DataSlice(var)

(c)If u is a Wait/Block statement then Update ActiveConcurrentSlice and send it to othere machines in the distributed system.

(d)If u is a Notify/UnBlock statement then Update ActiveConcurrentSlice and send it to othere machines in the distributed system.

7. Do the following after each statement u of the program P is executed.

(a) If u is a Def(var) statement and not a call statement then

Update ActiveDataSlice(var).

(b) If u is a call statement to a procedure Q then do

For every formal reference parameter var in the procedure Q do

ActiveDataSlice(Actual(u,var)) = ActiveDataSlice(var).

if u is a Def(var) statement then

ActiveDataSlice(var) = ActiveReturnSlice.

or every local variable var in the procedure Q do

ActiveDataSlice(var) = Φ .

Update CallSliceStack and ActiveCallSlice.

Set ActiveReturnSlice = Φ.

(c)If u is a receive(var) statement then do

ActiveDataSlice(var)= ActiveDataSlice(var) U ActiveDataSlice(var1) where Acti-veDataSlice(var1) is received from the sender machine.

(c) For every variable var Є DefVarSet(u) U UseVarSet(u) do

Update DyanSlice(machineno, u, var).

(d) If u is a test statement, then update ActiveControlSlice(u).

8. Exit when execution of the program P terminates.

4 Working of Proposed Algorithm

Table 2. A distributed system with four machines

Machine A	Machine B	Machine C	Machine D

main()
{
….
…..
….
…..
 	
10. send(p);
11. notifyAll();
….
…
}
	main()
	{
	….
	4. wait();
	5. receive(x);
	….
	…
	}
		main()
		{
		….
		…..
		7. wait();
		8. receive(y);
		….
		…
		}
			main()
			{
			….
			…..
			….
			….
			….
			….
			….
			16. wait();
			17. receive(z);
			……
			}

4.1 Analysis of Algorithm

In this example, we are having four machines each running different parts of a distributed program. Each machine is executing its part of the program P instrumented with its local slicer. The slicer updates ActiveConcurrentSlice with every communication that takes place between other machines in the distributed system.

In the above example, Machine A is broadcasting the message and notifying to other machines. After the notification, all other machines will receive the message and update their own slicer.

For Machine A
ActiveConcurrentSlice = Uactive U ActiveConcurrentSlice U ActiveControlSlice(t) where ControlDependentOn(Uactive) = t.
The ActiveConcurrentSlice and ActiveDataSlice(p) are sent to the Machine B Slicer.

For Machine B
ActiveConcurrentSlice = ActiveControlSlice U ActiveConcurrentSlice(Received from A)
ActiveDataSlice(x) = ActiveDataSlice(p) U ActiveConcurrentSlice U ActiveControlSlice(t)

For Machine C
ActiveConcurrentSlice(C) = ActiveControlSlice(C) U ActiveConcurrentSlice (Received from A)
ActiveDataSlice(y) = ActiveDataSlice(p) U ActiveConcurrentSlice(C) U ActiveControlSlice(t)

For Machine D
ActiveConcurrentSlice(D) = ActiveControlSlice(D) U ActiveConcurrentSlice (Received from A)
ActiveDataSlice(z) = ActiveDataSlice(p) U ActiveConcurrentSlice(D) U ActiveControlSlice(t)

4.2 Complexity of the Algorithm

The space complexity of our algorithm is mainly due to the space requirement for storing the CDG G of the local part of the program P. If program P has n number of statements then maximum O(n2) space is required to store the graph G. It can be easily shown that the other data structures used by our algorithm requires maximum O(n2) space with disposal of the runtime data structures when not required. The time complexity of our algorithm remains O(n).

4.3 Comparisons with Related Algorithms

The advantage of this algorithm is that, it does not use a trace file to store the execution history. It does not traverse a dependency graph and uses some data structures to capture the runtime dependencies that exist in a distributed system. Initially our algorithm constructs a Control Flow Graph to capture the control dependencies, used variable set and defined variable sets of a node. As the Control Flow Graph is not traversed by our algorithm even it can be disposed after the information is extracted from it. Our algorithm uses some run time disposable data structures which are updated with execution of each statement in the program. Hence the slices are available before it is asked for resulting in fast response time.

5 Conclusion

In this paper we present an algorithm for slicing distributed programs. We use the basic concepts of the inter-procedural dynamic slicing algorithm and remodel it to extract slices of distributed programs with introduction of some additional data structures. We believe that to extract precise slices of distributed programs the slicing algorithm must also work in a distributed manner. The future scope of this paper lies in designing a testing tool for the distributed slicing algorithm.

References

1. Pani, S.K., Arundhati, P., Mohanty, M.: An Effective Methodology for Slicing C++ Programs. International Journal of Computer Engineering and Technology 1, 72–82 (2010)
2. Mund, G.B., Mall, R.: Program Slicing. In: Compiler Design Handbook: Optimization and Machine Code Generation, pp. 14.1–14.35. CRC Press (2008)
3. Mall, R., Mund, G.B.: An efficient interprocedural dynamic slicing method. The Journal of Systems and Software 79, 791–806 (2006)
4. Mohapatra, D.P., Kumar, R., Mall, R., Kumar, D.S., Bhasin, M.: Distributed Dynamic Slicing of Java Programs. The Journal of Systems and Software 79, 1661–1678 (2006)

5. Mall, R.: Fundamentals of Software Engineering. Prentice – Hall, India (2003)
6. Mund, G.B., Mall, R., Sarkar, R.S.: Computation of Intraprocedural Dynamic Program Slices. Information and Software Technology 45, 499–512 (2003)
7. Sarkar, S., Mund, G.B., Mall, R.: An efficient dynamic program slicing technique. Information and Software Technology 44, 123–132 (2002)
8. Korel, B., Laski, J.: Dynamic Slicing of Computer Programs. Journal of Systems and Software 13, 187–195 (1990)
9. Weiser, M.: Program Slicing. IEEE Transactions on Software Engineering 10(4), 352–357 (1984)

An Efficient Incentive Compatible Mechanism for Paid Crowdsourcing

Shalini Gupta[1], Sajal Mukhopadhyay[1], and D. Gosh[2]

[1] Department of information Technology,
National Institute of Technology, Durgapur, India
[2] Department of Computer Science,
National Institute of Technology, Durgapur, India

Abstract. In paid crowdsourcing environment an organization post a task/problem at various platform like *freelancer.com* and the mass (crowd/developer) is invited to complete the problem. This environment is becoming an emerging trend to solve a problem with the brain of mass. At present the allocation and the payment made to the developers are mostly based on first price auction. However there is always a chance for manipulation in first price auction and also what punishment a developer should get if he try to do the same i.e. he can't complete the task/problem within the stipulated time mentioned in his bid (if they manipulate by day to get the project) or try to manipulate the money demanded for project completion is not addressed. In this paper we have developed an incentive compatible mechanism that will prevent the developer from doing manipulation and also a novel penalty scheme is incorporated in our mechanism so that, the punishment-to-developer problem could be handle in an efficient way.

Keywords: Crowdsourcing, Algorithmic Mechanism Design, Reputation System, Reverse Auction, VCG Auction.

1 Introduction

Crowdsourcing is emerged in past years as a online distributed problem solving production model.

Crowdsourcing was first defined by Jeff howe[3] as the act of taking a job traditionally performed by a designated agents(usually as employee)and outsourcing it to an undefined and generally large group of people in the form of open call.

More elaborately business people/organization post the task/problem at online platform and a vast number of contributor give solution toward the respective problem but when these solution is use by organization for their own benefit and to be sold by them to make profit, organization compensate contributor in many formats like cash prices and this process is termed as paid crowdsourcing. Due to paid crowdsourcing organization not only save time but also efforts to get satisfactory outcome (results).

Notable examples of this model include PeoplePerHour, Freelancer etc. is a global online labor marketplace where more than 2.5 million organizations can utilize a global network of over 2.6 million of the world-wide problem solvers.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 611–620.
springerlink.com

Crowdsourcing platforms like freelancer is promising as a new method for a problem solving where reverse auction (procurement auction) is going on. An organization posts a task to make it from the developer. In this, an organization gives money interval in (lower bound & upper bound) and max time (duration) for task completion where developers (bidders) makes bid to procure a project. However, the social efficiency of it is questionable. Since developers may behave selfishly and manipulate the bid so that efficient task allocations may be not realized.

Furthermore, no much research has done in this area. Since, mechanism design gives the optimum solution to the problem where multiple rational agents have some private information, we consider this problem as a mechanism design problem. Where, there is N finite number of rational developers/agents and they make competition to procure a single task of an organization and submit bid in open bid format. A developer $i \varepsilon N$ is the winner of an auction who bid lowest. Whereas i's bid is composed of two variables i.e. money for project development and reported time of project completion which is the private information for that developer. Since, every developer is rational therefore he is interested in his own utility and not interested in social efficiency. Developer is trying to give false bid if it gives a positive expected utility. viz. situation may occur that after project procurement, if i is deviating from his reported time for project completion i.e. he make delay to submit a project in announced time or he may manipulate the money which he demanded for the project completion. In this situation organization bear loss.

The above situation leads us to consider the following problem:

- How to allocate a project to a truthful developer?

In our paper we propose a truthful defrayal and penalty mechanism to allocate the job in socially efficient way where we try to reward developer for his truthfulness as well as punish him if he deviates.

2 Related Work

As far to concern with our research, that how to allocate resource efficiently in paid crowdsourcing on-line job market, a very less research went on.

In [3] author defines crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. Here in this paper author tried to explain crowdsourcing with the help of examples viz. InnoCentive, iStockphoto, Threadless. In this article author introduce to crowdsourcing what it is, how it works, and its potential.

The work mostly close to us is [4,12] in which John J. Horton et al. Solve the problem the problem that buyers and sellers en-counter in arriving at prices in a distributed labor market. But we can consider the above approach is to resolve the project allocation through bargaining. He discuss [12] about the problem of how to design an efficient crowdsourcing mechanisms. The concern of efficient mechanism design problem is all around incentives and strategic choices of all crowdsourcing participants. Author argues that designing efficient crowdsourcing mechanisms is not

possible without deep understanding of incentives and strategic choices of all participants. As study of the world's largest competitive software development portal: TopCoder.com,author find significant evidence of strategic behavior of contestants.

3 Background and Motivation Problem

3.1 Mechanism Design

Mechanism design is the sub-field of microeconomics and game theory that considers how to implement good system-wide solutions to problems that involve multiple self-interested agents, each with private information about their preferences [13]. In following way mechanism is-

Let N be the number of agents or participants. Each agent $i \varepsilon N$ can have private information or type or valuation v_i, e.g. in an auction the type of player would be his valuation price for the item ordered. A_i is the set of possible strategies or actions for player *i*. Depending on his type, the player will pick an action or strategy, $a_i \varepsilon A_i$, e.g. in an auction, a strategy of *i* would be a bid of a certain amount.

The mechanism provide as output function $o = o(a_1, a_2, \ldots, a_n)$. Agents $\forall i \varepsilon N$ try to optimize the utility μ_i . The objective or certain outcome can be achieved by using payment p_i given to the agent(s).

3.2 Reverse Auction(Procurement Auction)

In the reverse auction multiple sellers compete (bid) on goods and the evaluation value is shown by the buyer. The buyer wants to procure an item from the bidder who have lowest bid. The buyer benefits from significant price reduction; the supplier benefits because an e-Auction is effectively very open and transparent com-petition where s/he can bid against the other companies. The valuation spaces [5] are given by:

$V_i = \{v_i \mid v_i\,(i\text{-}wins) \leq 0$ and $\forall j \neq i\ v_i\,(j\text{-}wins) = 0\}$, and indeed to procure an item from the lowest cost bidder is maximizing the social welfare.

The well known VCG payment rule would be used for the mechanism to pay to the lowest bidder an amount equal to the second lowest bid, and pay nothing to the others [5] maximize the social welfare.

3.3 VCG Second Price Auction

Arguably the most important positive result in mechanism design is what is usually called Vickery-Clarke-Groves (VCG) mechanism. It is a sealed bid auction. The important property of the VCG mechanism is that it is truthful or incentive compatible [5,6,7].

3.4 Reputation System

Reputation systems have emerged as a method for stimulating adherence to electronic contracts and for fostering trust amongst strangers in e-commerce transactions [1,10,11]. Typically the way these systems work is, once a transaction has been

completed between participants/members, based on their satisfaction, each of them may give their partner a rating. This rating is collected and processed by reputation mechanism and is available for future reference to potential traders who might engage in transactions with each other. According to Resnick et al. [2] reputation mechanisms can provide an incentive for honest behavior and help people make decisions about who to trust.

The two main type of reputation system architecture [9] are centralized reputation system and distributed reputation system.

4 Problem Formulation

4.1 Current Scheme

In present paid crowdsourcing schemes, where individuals are selected from crowd by the organization for work through an auction like in freelancer.com, the scenario is given below in fig1. At present N numbers of developers (bidders) submit open bid for task procurement.

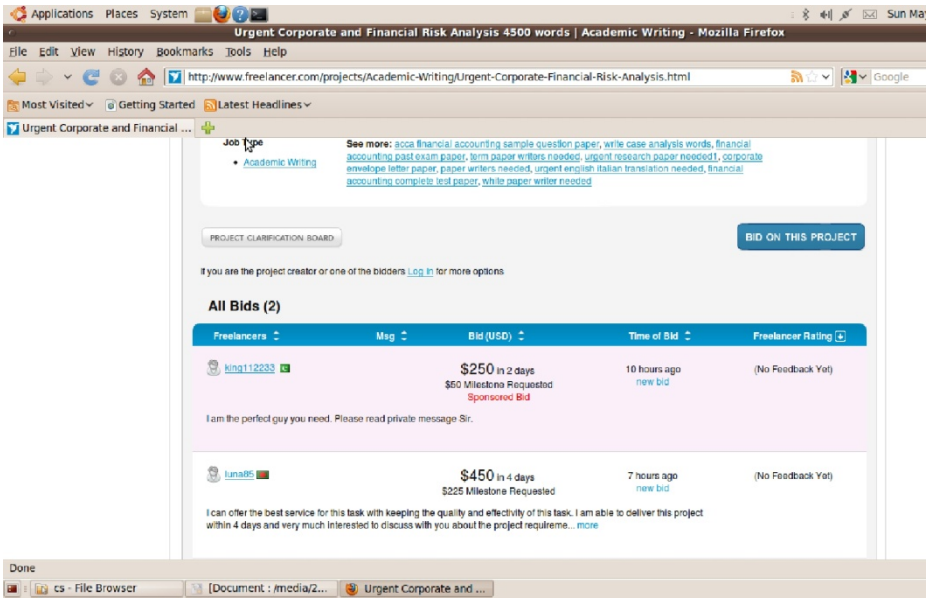


Fig. 1. Bidding Scenario at Freelancer.com

B : set of bid vector.
b_i : bid value of developer i, $i \varepsilon N$ compose of two parts $(m^*, d)_i$ where, m_i^* is total money demand by developer i for task completion and d_i completion time for task

Winner is selected on the basis of some weightage w_i given to money, time and reputation means, selection criteria is either lowest money, lowest time, highest reputation or some combination of all three. And winner gets payment as first price (his own bid price).

4.2 Problem Definition with Respect to Current Scenario

1 $\mu_i^m = m_i^* - v_i^m$, utility of developer *i* with respect to money. Where m_i^*, v_i^m is the total money demanded and true valuation with respect to money of developer *i*.

2. $\mu_i^d = d_i - v_i^d$, utility of developer *i* with respect to time. Where v_i^d is the true valuation with respect to time and d_i is the announced task completion time of developer *i*.

Here for every developer quasi-linear utility is considered. Therefore, each developer tried to maximize his own utility means, he may lie and not give his true valuation in his bid. Under this scenario it is not possible to allocate task to the most indigent or deserving developer.

AIM: To achieve maximum social welfare[5] *viz.* in task procurement auction if winning criteria for developer is lowest money, then to choose a developer who value the given project lowest is maximizing social welfare condition.

5 Our Mechanism

Basic Definition:

- A centralized reputation system is maintained *viz.* (by crowdsourcing platform) which facilitates organization to rate various developers after the completion and submission of full problem solution. $R_i : n_i \rightarrow r_i$, implies the reputation corresponding to developer $i(r_i)$.
- An organization posts a task, which is allocated to developer through auction mechanism. Hence, we denote the collection of social alternatives as X = { single posted problem by an organization }.
- There are N number of developers make competition to procure a single task of an organization and submit bid in sealed bid format.
- Let $v_i(x)$ is the valuation of developer i to procure a task where, $x \varepsilon X$. And v_i is consider as the value given to the social alternative x by the i^{th} developer. But we have to note that v_i signify that how much money per day he want to complete the task solution and how much time he want to spend on a task.
- A function $f: v_i \rightarrow x_i$, is called a social choice function which defines the winner, having minimum valuation.
- A developer strategy b_i is composed of two separate parts: m_i money per day and d_i time he want to spend on a problem where, $g(m_i, d_i) : m_i * d_i \rightarrow b_i$ signify that bid of *i* is a product of his or her money per day and time declaration for task completion by an developer.
- The output of mechanism $o = (f, p_1 \ldots . p_n)$ where, p_i is a payment of i^{th} player and f is a social choice function. o = (winner of an auction, payment).
- The utility μ_i of winning developer is b_i-v_i.

Here we propose Defrayal and Penalty mechanism to tackle an issue mentioned in [Problem formulation]. At present scenario, in paid crowdsourcing where an organization post their problem and take bid from developer in open bid format. In our scheme we adopt sealed bid form for bidding rather than open bid. We are proposing our scheme in following way:

- Task Allocation
- Defrayal and penalty Mechanism
- Claim our mechanism is truthful

5.1 Task Allocation

B = $\{b_1, b_2, \ldots, b_n\}$ is a set of bid and $|B| = N$.
L = $\{b_i \mid b_i \leq b_j,\ \forall i, j\ \varepsilon\ N \text{ and } b_i, b_j\ \varepsilon\ B\ \}$

where as $L \subseteq B$. In other words L is a set of bids which are equal and minimum than rest of the bids.While task allocation we consider following two cases:

Case1: $|L| = 1$ In this case set L contain only single element b_i ; $i\ \varepsilon\ N$, where $b_i = \min(b_1, b_2, \ldots, b_n)$. Hence, we allocate task to the developer *i* whose the bid is lowest.

Case2: $|L| > 1$ In this case, hence all bids are equal and lowest among all bids of set B. To allocate task consider reputation of bidders in following way-

Apply reputation points of developer i ε N to allocate task (consider reputation point of only those developers whose bid is in partial order set L).Then arrange these reputation points in non-increasing order. Allocate task to developer *i*, who have the highest reputation points among all the developers $b_i \varepsilon L$. And if more than one developer have most highest and equal reputation points then choose winner randomly among them.

5.2 Defrayal and Penalty Mechanism

In last section we allocate task to $min\ (b_1, b_2, \ldots, b_n)$, where $1, 2, \ldots, n\ \varepsilon\ N$. while problem allocation we assume that bidder is truthful in the context of his m_i & d_i. Let us say,

b_i = bid price of developer *i* and consider *i* is the winner.
$b_j = min(b_1, b_2 \ldots, b_{i-1}, b_{i+1}, \ldots, b_n)$.

5.2.1 Defrayal Function

As our aim of study is to allocate project in socially efficient way and for that we need to elicit the true valuation of every developer with respect to money/day and time. Since VCG mechanism is truthful [5,7], we adopt this mechanism in addition with penalty scheme to design solution. If $b_i < b_j$ then *i* is the winner and his defrayal (df_i) is describe by following method:

$$df_i(b_i, b_j) = \begin{cases} b_j, & b_i < b_j \\ 0, & else \end{cases} \quad (1)$$

5.2.2 Penalty Scheme

- Penalty Function: In introductory section we point out problem that is, developer may manipulate the bid values. Here we consider this rationality of a developer. We assume that developer is rational in d_i i.e. he lie or choose to perform any project allocated to him in time $d_i' \geq d_i$ where d_i' is the actual execution time of developer i and as a result, organization have to bear a loss. Penalty function is devise to avoid this deviation of developer from his d_i and give punishment for deviation.

$$pf_i(b_i, b_j) = -\begin{cases} (b_j - b_i) + & if\ d_i < d_i' \quad where \quad > 0 \\ 0 & else \end{cases} \tag{2}$$

  The above penalty function (pf_i) indicates that if an agent i is honest and perform allocated project in his announced time d_i then his penalty is zero. But if he deviates from d_i and consume more time, then his penalty is increases as much as he deviates plus some positive value. Here penalty scheme is designed in such a way that developer always gives penalty greater than its valuation if he deviates.

- Reputation function: Along with penalty function, reputation function gives the additional method to punish the deviating developers. Reputation of developer i is increases/decrease with some constant k ε R(real number) factor if he finish/not finish allocated task in his announced time d_i.

$$R_i(r) = \begin{cases} r_i^+ & if\ d_i \geq d'_i \\ r_i^- & if\ d_i < d'_i \end{cases} \tag{3}$$

5.2.3 Payment and Utility

Payment function is take account of developer's rationality in money and time. Payment of developer for project development is based on his truthfulness. Net payment is the sum of two terms compensation and penalty.

$$p_i(b_i, b_j) = df_i(b_i, b_j) + p\,f_i(b_i, b_j)$$

Here we have to note that utility/profit μ_i of a developer i is based on others bid value and his time d_i.

Hence $\mu_i = p_i - v_i$

$\mu_i = (df_i + pf_i) - v_i$

5.3 Truthfulness

5.3.1 Theorem

The Defrayal and Penalty mechanism is truthful implementation of social choice function in paid crowdsourcing

5.3.2 Proof

Part1

m_i *:* Money per day of developer i.

d_i *:* Total days of developer i.

Here we assume that m_i, d_i of developer i is his own valuation of problem.
m_{-i} : Lowest money per day of the developer other than i.
d_{-i} : Lowest total days of the developer other than i.
If $g(m_i, d_i) < g(m_{-i}, d_{-i})$ then, developer i is the winner.
$\mu_i = g(m\text{-}_i, d\text{-}_i) - g(m_i, d_i)$
And his payment $p_i = g(m_{-i}, d_{-i}) > g(m_i, d_i)$ if developer is truthful

- Case1: If developer i deviates by days (decrease days to d_i'')
  $= p_i = g(m_{-i}, d_{-i}) - [g(m_{-i}, d_{-i}) - g(m_i, d_i'')] - \varepsilon$
  $= g(m_i d_i'') - \varepsilon$
  $< g(m_i, d_i)$; since > 0 & $d_i'' < d_i$
  $< m_i * d_i$
  hence, $p_i < m_i * d_i$ therefore, $\mu_i < 0$

- Case2: If developer i deviates by days (increase days to d_i''')
  In some of cases $g(m_{-i}, d_{-i}) < g(m_i, d_i''')$ since $d_i''' > d_{-i}$
  therefore, developer i loose the auction and his utility $\mu_i = 0$

- Case3:If developer i deviates by money(decrease money per day to m_i'')
  Naturally in this case $d_i < d_i'$.
  $p_i = g(m_{-i}, d_{-i}) - [g(m_{-i}, d_{-i}) - g(m_i'', d_i)] - \varepsilon$
  $= g(m_{-i}, d_{-i}) - g(m_{-i}, d_{-i}) + g(m_i'', d_i) - \varepsilon$
  $= g(m_i'', d_i) - \varepsilon$
  $< g(m_i, d_i)$; since $\varepsilon > 0$ & $m_i'' < m_i$
  $< m_i * d_i$
  hence, $p_i < m_i * d_i$
  therefore, $\mu_i < 0$

- Case4: If developer i deviates by money (increase money per day to m_i''')
  In some of cases $g(m_{-i}, d_{-i}) < g(m_i''', d_i)$ since $m_i''' > m_{-i}$
  therefore, developer i loose the auction and his utility $\mu_i = 0$

- Case5: If bidder i deviates by both money per day (increase money per day to m_i''' and decrease days to d_i'')
  Unless $g(m_i''', d_i'') \leq g(m_i, d_i)$
  $p_i = g(m_{-i}, d_{-i})$
  $= m_{-i} * d_{-i}$(same)
  $\mu_i = g(m_{-i}, d_{-i}) - g(m_i, d_i)$(same)
  Otherwise, μ_i=0; since developer i loose the auction

- Case6: If developer i deviate by both money (decrease money per day to m_i'' and increase days to d_i''')
  Unless $g(m_i'', d_i''') \leq g(m_i, d_i)$
  $p_i = g(m_{-i}, d_{-i})$
  $= m_{-i} * d_{-i}$ (same)
  $\mu_i = g(m_{-i}, d_{-i}) - g(m_i, d_i)$(same)
  Otherwise, μ_i=0; since developer i loose the auction

Part2: If developer *i* deviate from his announced task completion time d_i then he may loose his reputation because, in this situation the organization who allotted the task to developer gives negative rating to the developer. And if in case bid of developer *i* belong to partial order set L i.e |L| > 1 then, probability to win the auction/task is decreases.

Therefore overall, dominant strategy for any developer is to bid his true valuation (m_i, d_i).

5.4 Time Complexity of Mechanism

The overall time complexity of our mechanism could be given as follows:

1. Time to sort out all N developers according their bid b_i in non-decreasing order is: $\Theta(n \lg n)$.
2. If |L| > 1 then, time to sort out all bidders (developers) according their reputation r_i in non-increasing order is: $\Theta(n \lg n)$.
3. To calculate utility of winning developer is: $\Theta(k)$ where $k < n$ and $k \varepsilon R$.
4. Total time taken to calculate step1, step2 and step3 is: $\Theta(n \lg n) + \Theta(n \lg n) + \Theta(k)$.
5. Over all time complexity is: $\Theta(n \lg n)$.

6 Conclusion

Proposed scheme use a mechanism design approach for paid crowdsourcing environment where reverse auction is going on (many developer are bidding to procure a single problem from an organization) and provide a truthful defrayal and penalty mechanism to resolve, how the private value of these selfish developers is reveal out so that, project is allocate to the most indigent developer. The overall time complexity of our mechanism is polynomial (*n lg n*).

7 Future Work

In future we will try to give a truthful mechanism for the situation where developer may be interested in more than a single task and give combine bid to procure more than single task[multiple tasks(say m tasks) to be allocated to several developer(say n users)]. This problem of task allocation becomes combinatorial in nature. More formally there is a set of m indivisible tasks that are concurrently auctioned among n developers.

References

1. Patton, M.A., Jsang, A.: Technologies for Trust in E-Commerce. In: Proceedings of the IFIP Working Conference on E-Commerce, Salzburg, Austria (June 2001)
2. Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K.: Reputation systems. Communications of the ACM 43(12), 45–48 (2000); Brent Frei: Paid Crowdsourcing - Smartsheet, Produced by Smartsheet.com. A leading provider of paid crowdsourcing technology & services, Version 1.00.00 - Release Version. pp. 4, 6 (2009)

3. Brabham, D.C.: Crowdsourcing as a Model for Problem Solving. Convergence: The International Journal of Research into New Media Technologies 14(1), 75–90 (2008)
4. Horton, J.J., Zeckhauser, R.J.: Algorithmic Wage Negotiations: Applications to Paid Crowdsourcing. In: Crowd Conf. 2010, San Francisco, CA (2010)
5. Nishan, N., Roughgarden, T., et al.: Algorithmic Game Theory. Cambridge University Press (2007)
6. Nisan, N., Ronen, A.: Algorithmic Mechanism Design. Games Econ. Behav. 35, 166–196 (2001)
7. Vickery, W.: Counter Speculation, Auctions and Competitive Sealed Tenders. J. Economic Theory, 187–217 (1961)
8. Vukovic, M.: Crowdsourcing for Enterprises. In: 2009 World Conference on Services-I, Los Angeles, CA (2009)
9. Azer, M.A., El-Kassas, S.M., Hassan, A.W.F., El-Soudani, M.S.: A Survey on Trust and Reputation Schemes in Ad Hoc Networks. In: Third International Conference on Availability, Reliability and Security (2008)
10. Resnick, P., Zeckhauser, R., Swanson, J., Lockwood, K.: The Value of Reputation on eBay: A Controlled Experiment. Experimental Economics 9(2), 79–101 (2006)
11. Hood, W., Wilson, C.S.: The Value of Reputation: The Literature of Bibliometrics, Scientometrics, and Informetrics. Scientometrics 52(2), 291–314 (2001)
12. Archak, N.: Money, Glory and Cheap Talk: Analyzing Strategic Behavior of Contestants in Si-multaneous Crowdsourcing Contests on TopCoder.com. In: International World Wide Web Conference Committee, IW3C2 (2010)
13. Myerson, R.: Perspectives on mechanism design in economic theory. Prize Lecture, Department of Economics, University of Chicago (2008)
14. Narhari, Y.: Lecture notes on Mechanism Design. IISc Banglore, pp. 1–11 (2008)

Doubling Runtime Estimations to Improve Performance of Backfill Algorithms in Cloud Metaschedular Considering Job Dependencies

Ankur Jindal and P. Sateesh Kumar

Indian Institute of Technology, Roorkee-247667, India
ankur2.iitr@gmail.com,
drpskfec@iitr.ernet.in

Abstract. Job scheduling is a very challenging issue in cloud computing. Traditional backfill algorithms such as Easy and conservative are extensively used as job scheduling algorithms. Backfill algorithms require the shorter job to come forward if sufficient resources for the execution of this job are available and run in parallel with the currently running jobs provided it does not delay the next queued jobs. This technique is highly dependent on runtime estimations of job execution. Moreover in real life scenario it has seen that submitted job's may or may not be independent to each other. In this paper we have proposed a technique that uses dynamic grouping method to consider job dependencies and doubling runtime estimation method in cloud metaschedular to improve performance of backfill algorithm. Results have shown that doubling runtime estimations can significantly improve performance of backfill scheduling algorithms provided that the runtime estimations are correct.

1 Introduction

Cloud computing [4] is a future technology that won't need to compute on local computers, but on centralized facilities operated by third-party compute and storage utilities. Job scheduling is one of the core and challenging issues in a Cloud Computing system. In general two schedulers [5] are available in cloud one is global or metaschedular and another is local scheduler. Local scheduler determines how the processes that reside on a single CPU are allocated and executed. Users submit their jobs to Metaschedular, it is metaschedular responsibility to use information about the system and allocate processes among the different clusters in cloud. In this paper, it is attempted to improve the performance of EASY(the Extensible Argonne Scheduling System) backfill Algorithm[2] by doubling runtime estimation method in cloud metaschedular to maximize the resource utilization and minimize the resource gap of idle resources. We also have taken care that submitted jobs may or may not be independent to each other. The evaluation of EASY algorithm before and after doubling runtime estimations is made. The rest of the paper is organized as follows: the next section discusses the related works. Section 3 presents an overview of cloud metaschedular architecture where EASY with doubling runtime estimations as well as dynamic grouping of jobs is made. Section 4 describes an algorithm which is combination of dynamic grouping and EASY algorithm with doubling of runtime estimations. Results of the performance and parameter study are reported in section 5.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 621–628.
springerlink.com

2 Related Work

Recent years have seen many efforts focused on the efficient utilization of cloud resources by cloud metaschedular that lead satisfaction to both cloud service provider and service users. CloudSim [8] allows modeling and simulations of entities in parallel and distributed computing systems. Aneka[9] form enterprise grid and cloud platform provide following services as task scheduler service for the task programming model, thread scheduler services, for the thread programming model, storage service for file store for applications. Hadoop a popular open-source implementation of the Google's Map Reduce model is primarily developed by Yahoo. The work done by [6], [7] considers Hadoop scheduler can cause severe performance degradation in heterogeneous environments and provide a new scheduling algorithm, Longest Approximate Time to End (LATE) for concurrent jobs in heterogeneous environments. But LATE doesn't always improve the performance.

The work related to [1] considers self adaptive backfill policy for parallel systems using multi queue. And IBM in paper [3] proves the effectiveness of backfill algorithms for parallel systems. The work done by [10] focuses on optimizing the system throughput by maximizing the overall resource utilization and guaranteeing increased performance of the applications. Here an optimal solution for cloud job scheduling is made only better than the traditional First Come First Serve (FCFS), Round robin and failed to fill the resource gap completely. The work related to [2], consider the commonly used method of job scheduling FCFS, along with Backfilling method EASY and CONSERVATIVE algorithms where small jobs are moved ahead in the schedule can fill the resources gap that is generated by FCFS. However it has seen that resource gap is not fully covered using given runtime estimations.

3 Task Scheduling Problem

In general cloud users submit their jobs to cloud metaschedular. It is the cloud metaschedular which make decision to map jobs submitted by cloud users to cloud clusters. Figure1. Shows the scenario.

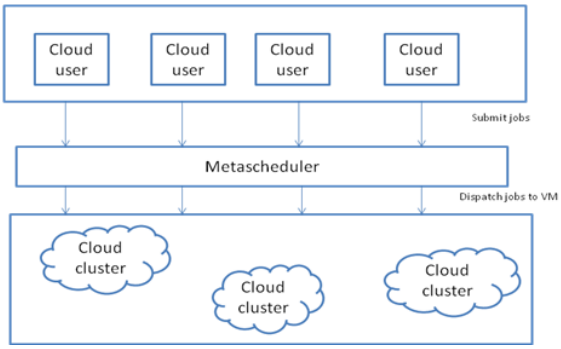


Fig. 1. Cloud metaschedular

The following steps are performed:

Step1. The cloud users submit their request for job completion to the metaschedular
Step2. As per the availability of free nodes in cloud cluster decision for scheduling is made. It is the metaschedular which is responsible for mapping jobs between cloud clusters and cloud users.
Step3. After the jobs are submitted to cloud cluster these are executed by local scheduler.

4 Task Scheduling Algorithm

We are representing our job workflow in the form of a DAG (Directed Acyclic Graph) G (V, E). V (Vertices) represents jobs and E (Directed edges) represents dependencies. We are considering a DAG because if there is a cycle present, we will stick in a situation of deadlock. Following steps are performed to make dynamic grouping.

4.1 Dynamic Grouping Method Algorithm

1. First find all the root nodes means jobs which are not dependent on any other job. Put these jobs in first group.
2. Increment group number and check all the nodes which are directly dependent on all or some jobs of the previous group. Put these jobs in this new group.
3. Check if there are any other nodes left, if yes go to step 2 otherwise step 4.
4. Apply EASY with doubling runtime estimation on each of these groups individually.

Fig. 2. Dynamic Grouping Method

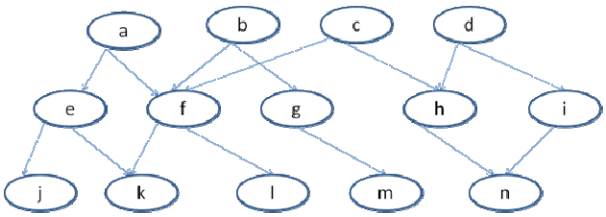


Fig. 3. DAG representing job workflow

Here according to this method we get three groups for DAG shown in Figure2 G1= {a,b,c,d}, G2= {e,f,g,h,i}, G3= {j,k,I,m,n}

Improved Easy Backfill Algorithm:

1. Double the given runtime estimation's of jobs
2. Find the shadow time and extra nodes
    a) Find when enough nodes will be available for the first queued job; this is the shadow time.
    b) If this job does not need all the available nodes, the ones left over are the extra nodes.
3. Find a backfill job
    a) Loop on the list of queued jobs in order of arrival
    b) For each check whether either of these conditions hold
        i. It requires no more than the currently free nodes ,and will terminate by the shadow time , or
        ii. It requires no more than the minimum of the currently free nodes and extra node
    c) The first such job can be used for backfilling

Fig. 4. EASY with doubling runtime estimations

Consider a scenario with 5 jobs in some individual group as shown in Fig 5. According to EASY backfill algorithm with actual runtime estimates the jobs will be executed as shown in Fig 6. The total execution time of all the 5 jobs come here is 950 units. But if we double the runtime estimates of the jobs total job execution time come here is750 units. The corresponding execution sequence is shown in Fig 7-11.These Figures are self explanatory.

Jobs	Number of Pe's required	Expected Runtime
J1	5	400
J2	9	100
J3	2	200
J4	4	300
J5	7	150

Fig. 5. Jobs with expected runtime

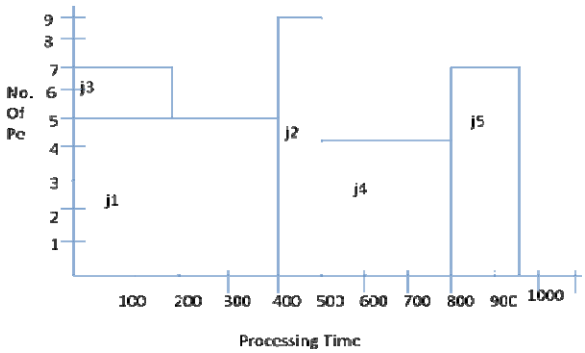


Fig. 6. Job execution according to EASY backfill with actual runtimes

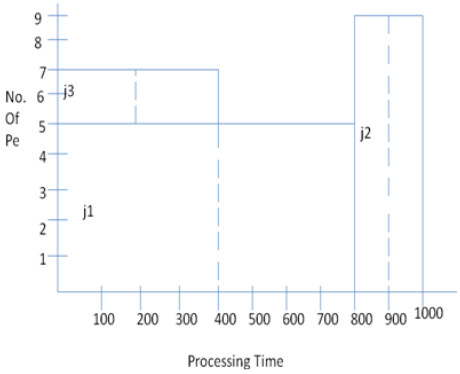


Fig. 7. Job execution according to EASY backfill with doubling j1 and j3 starts execution

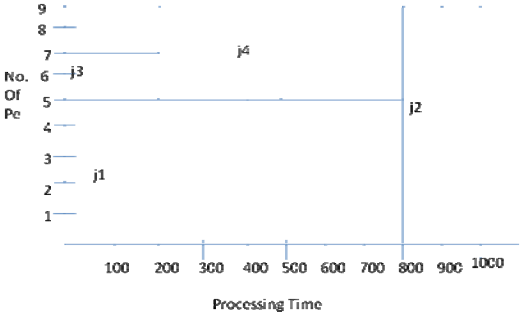


Fig. 8. Job execution according to EASY backfill with doubling after j3 completed execution and j4 and j1 are executing

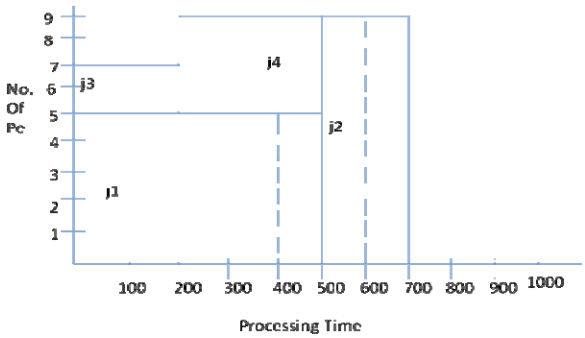


Fig. 9. Job execution according to EASY backfill with doubling after j4 and j1 completed execution and j2 is executing

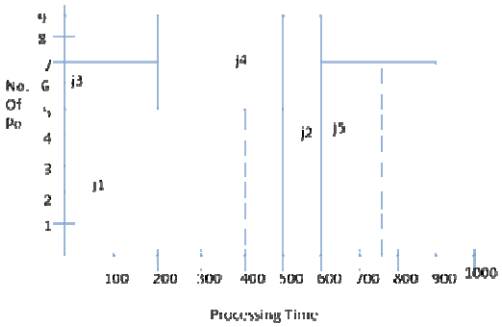


Fig. 10. Job execution according to EASY backfill with doubling after j2 completed execution and j5 is executing

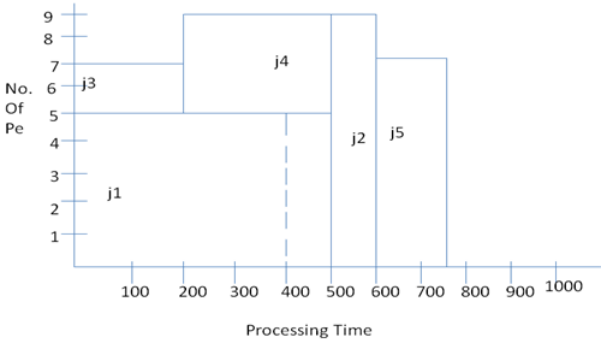


Fig. 11. Job execution according to EASY backfill with doubling after all jobs completed execution

5 Simulation and Results

In this section, the experimental evaluation for the cloud metaschedular is discussed. The Cloudsim toolkit is used to simulate the algorithm with various experimental setups. The default classes in Cloudsim toolkit are extended to implement the proposed policy and other parallel job scheduling strategies. The experimental setup include by varying jobs runtime, speed of processing elements, size of cloudlets and also policies. It can be analyzed by experimental results as shown in Figure 12 that job execution with doubling runtime estimation is faster than backfill algorithms with actual runtime estimations.

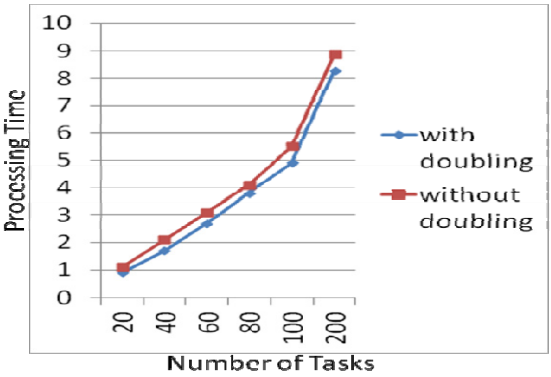


Fig. 12. Performance Backfill Algorithms before and after doubling

6 Conclusion

Doubling approximates an SJF like scheduling by repeatedly preventing the first queued job from being started. Thus, doubling trades off fairness for performance and should be viewed as a property of the scheduler, not the predictor. We have shown doubling technique on EASY Backfill algorithm, but it can be applied to any backfill algorithm.

References

1. Lawson, B.G., Smirni, E., Puiu, D.: Self-adapting backfilling scheduling for parallel systems
2. Feitelson, D.G., Weil, A.M.: Utilization and predictability in scheduling the IBM SP2 with backfilling. In: Proceedings of the First Merged International and Symposium on Parallel and Distributed Processing, Parallel Processing Symposium, IPPS/SPDP 1998, pp. 542–546 (1998)
3. Tsafrir, D., Etsion, Y., Feitelson, D.G.: Backfilling using system-generated predictions rather than user runtime estimates. IEEE Transactions on Parallel and Distributed Systems 18(6), 789–803 (2007)

4. Foster, I., et al.: Cloud Computing and Grid Computing 360-Degree Compared. In: Grid Computing Environments Workshop, pp. 1–10 (2008)
5. Peixoto, M.L.M., et al.: A Metascheduler architecture to provide QoS on the cloud computing. In: 2010 IEEE 17th International Conference on Telecommunications (ICT), pp. 650–657 (2010)
6. Zaharia, M., Konwinski, A., Joseph, A.D., Katz, R., Stoica, I.: Improving map reduce performance in heterogenous environment. In: OSDI (2008)
7. Isard, M., et al.: Quincy: fair scheduling for distributed comptuing clusters. Microsoft Research, SOSP (2008)
8. Buyya, R., et al.: Modeling and simulation of scalable Cloud computing environments and the CloudSim toolkit: Challenges and opportunities. In: International Conference on High Performance Computing & Simulation, HPCS 2009, pp. 1–11 (2009)
9. Buyya, R.: Aneka next generation. net grid/cloud computing company (2009)
10. Sadhasivam, S., Jeya Rani, R., Nagaveni, N., Vasanth Ram, R.: Design and implementation of two level scheduler for cloud computing environment. In: International Conference on Advance in Recent Technologies in Communication and Computing (2009)

Self-managing the Performance of Distributed Computing Systems – An Expert Control Solution

Ravi Kumar G.[1], C. Muthusamy[2], A. Vinaya Babu[3], and Raj N. Marndi[4]

[1] HP Bangalore, JNTUH Hyderabad
ravikgullapalli@yahoo.co.in
[2] Yahoo, Bangalore
chelgeetha@yahoo.com
[3] JNTUH Coll of Engg, Hyderabad
dravinaybabu@yahoo.com
[4] HP Bangalore
rajnarayan99@yahoo.com

Abstract. The advent of internet and cloud computing trends is increasing the complexity of IT Infrastructures very rapidly. The non-functional requirements availability and performance are becoming increasingly important. IT Service providers constantly facing challenges to meet the performance related SLAs defined with Enterprise Customers. To meet such demands, IT environments are built with self-managing capabilities. Autonomic Computing has emerged to support self-managing features using the feedback control systems. There are investigations in using control systems in different areas of computing such as computer networking, database systems, data centers and distributed computing systems in enterprise and cloud environments. We observe that there is a need for an end-to-end solution starting from design and modeling of the software, deploying and runtime management that enables in building self-managing. In this paper we propose an end-to-end Expert Control System Solution for Distributed Computing Systems and discuss its application in Java Enterprise environments.

1 Introduction

The advent of internet has brought rapid change in the way computing applications are developed, deployed and executed [1]. This has a huge impact on IT infrastructures hosting software applications. There are various operations from Business Enterprises to the individual consumers heavily depending upon IT systems. It is important and obvious that such IT systems provide best possible functionality to the users. Availability and Performance are becoming increasingly important requirements [2]. Additionally there are other dimensions such as heterogeneous compositions of hardware and software components, number of servers, geographically dispersed sites, number of users adds further complexity to the IT systems [3] driving self-surviving IT Systems. The IT systems require the ability to pro-actively identify the faults and performance degradation. Such abilities of pro-active control require predicting the workload dynamics, number of users, usage patterns of the IT systems

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 629–638.
springerlink.com

triggered the emergence of Autonomic Computing Systems [4] where the IT Systems are capable of self-surviving dealing with faults and performance issues. There are various concepts and technologies though explored to build Autonomic Computing Systems, Feedback Control Systems has proved very successful [5] and Adaptive Control [5] is used in designing pro-active control systems. Majority of the applications of Adaptive Control systems are explored in Electrical [6] and manufacturing engineering [6]. There are recent research trends in investigating Feedback control in building Autonomic Computing Systems such as computer networking [7], database systems [8], and data centers [9], majorly in Distributed Computing Systems [10].

2 Problem and Related Work

The recent trends in the internet and cloud environments are increasing the complexity of IT systems due to various reasons such as the volume of servers and storage in the data centers, hardware and software applications from different vendors, integration challenges of applications, geographically distributed IT infrastructures. The software applications hosted in such complex IT environments have the challenge of meeting SLAs consisting both functional and non-functional requirements [11]. The most common non-functional requirements include Availability, Performance and Security [2]. To meet these SLAs it is necessary to design all the layers of IT architecture with capabilities of availability and performance management. In this paper we focus on the performance management issues in the Middleware application layer [12]. The typical distributed system middleware layers are Application Servers such as Java based Enterprise Servers (JEE Servers) [13], serve as platforms for rapid distributed application development, deployment and execution [14], play a significant role in building self-surviving applications. If such platforms fail to perform self-corrective actions, service providers may breach the SLAs affecting the business with unfriendly IT experience for the users and huge penalty. There are many systems like self-managing, self-protecting usually called as self-X systems [15] and Autonomic Computing systems. The majority of such systems are built using the Control Systems theory making use of Adaptive Control Systems significantly. There is a significant study conducted in control systems application in Distributed Computing Systems such as Web Servers performance [16], improve the caching in Web Servers and Services [17,18] and EJB Servers performance [19] using PI, PID Controllers, Fuzzy and Neural Controllers [20], hybrid controllers [21, 22]. There are similar attempts in building intelligent solutions using control systems but with a specific emphasis on cloud environments [23]. We observe that there are solutions to specific problems in Middleware environments, and end-to-end self-managing solution is still a potential gap. We believe that self-managing mechanisms have to be built during the design and modeling of the system which will be executed and exercised during runtime. In this paper we propose an Intelligent Control based Expert Control Solution enriching the self-managing Distributed System Middleware and discuss its application in some of the JEE Server components, followed by the analysis.

3 Expert Control System Solution

We propose an end-to-end solution that enables the Distributed System Middleware with autonomic computing abilities inherently as first class features. We build the autonomic computing elements during design and modeling of the application software that execute at runtime environment. We explain the lifecycle and working model of the proposed solution in this section.

3.1 Design and Modeling Lifecycle

The Fig. 1 below shows the life cycle of the proposed autonomic aware UML. It is a Framework available as within the UML Tool constructs required to develop the application to monitor, model, and select controllers. The application developer will use the design created and will generate the code, our solution generates java code with all the required autonomic aware constructs. After the application is implemented, it will be built, packaged and deployed into the Application Servers.

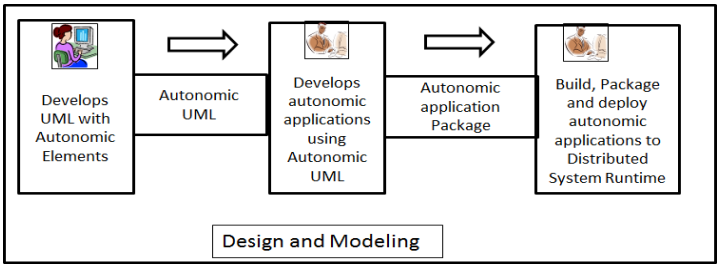


Fig. 1. Design and Modeling Lifecycle

3.2 Runtime Lifecycle

The autonomic aware software once designed the software developer implements and deploys to the runtime. The following Fig. 2 shows the lifecycle of the autonomic aware software till deployed into the runtime. The Expert Control System will be packaged and integrated with the Distributed System Middleware. Our Expert Control System will be an integral part rather than a COTS product.

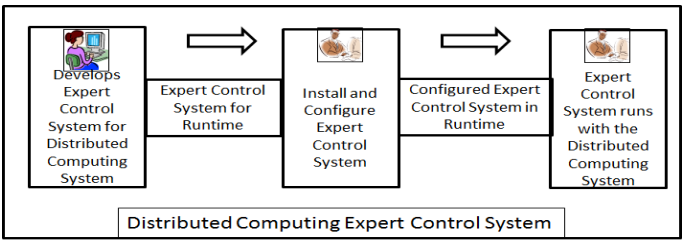


Fig. 2. Runtime Lifecycle

4 Design and Modeling

Design and Modeling is most initial important step in the process of developing any software application and service. In order to build any self- application, it is required to plug-in the autonomic computing elements during the modeling of software system. In this direction, we propose a framework that can be introduced into the UML Tools, such that it can be used like any other UML design construct. The objective of this framework is to enable various self-managing constructs as explained below.

4.1 Architecture

The Fig. 3 below shows the architecture of the autonomic aware framework which provides various design constructs for developing self-managing applications. Some of such constructs include control class design to be controlled, pre-define the standard system metrics like CPU, memory, message throughput, dynamically create and modify the models. The following are the different components of the framework.

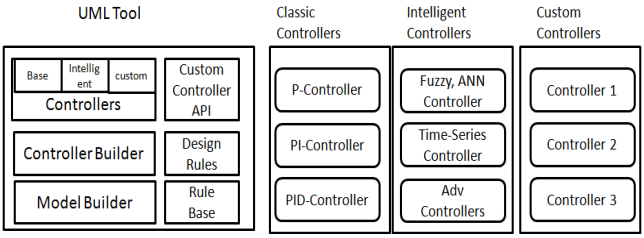


Fig. 3. Autonomic aware UML architecture

- Model Builder: In Control Systems it is essential to model the system to be controlled. The framework has the Model Builder to generate a model of the application or service to be controlled. It requires the input and output parameters to be provided by the designer during the system design.
- Controller Builder: When the software designer chooses a specific controller, the corresponding controller class is generated.
- Base Controllers: A Set of base controllers such as P, PI and PID Controllers are part of the framework. This enables the designers to choose during system design
- Intelligent Controllers: There are a set of advanced controllers that are available as part of the framework based on Machine learning techniques such as Fuzzy Logic or Data Mining based techniques such as Time-Series controllers, Episode discovery, outlier type of controllers.
- Custom Controller API: The framework provides flexibility to develop custom controllers and add such controllers to the existing controllers.

5 Runtime Expert System

In this section we discuss the Expert control solution that consists of different components which ensure the pro-active control mechanisms and ability to self-manage the performance of the different components of the Distributed System Middleware.

5.1 Architecture

The Fig. 4 is the Expert Control System Architecture for the Distributed System Middleware which works in conjunction with the Autonomic aware UML framework to provide end-to-end solution for building Autonomic Systems. The Autonomic aware UML Package parser transforms the autonomic constructs present in the autonomic aware software into runtime constructs.

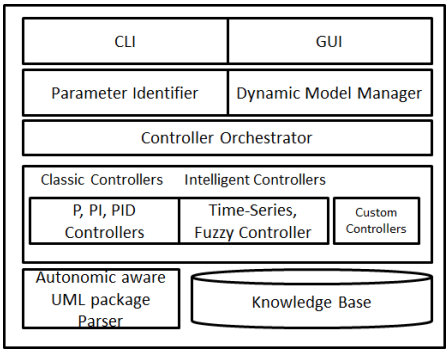


Fig. 4. Intelligent Control Solution

The solution has built-in classic and intelligent controller sets, with a facility to develop custom controllers available at runtime. The behavior of the controlled system, the workload patterns are captured in the knowledge base. Controller Orchestrator will enable the adaptive control mechanisms to choose a right controller based on the knowledge captured. For example a hybrid control may have a PI Controller, Fuzzy Controller and Time-Series Controller. The Dynamic Model Manager retunes the model parameters at runtime. Parameter Identifier will identify the new set of input parameters which are not included during the design but affecting the system behavior. We also intend to develop CLI and GUI based interfaces to perform queries on the Expert Control System.

6 Case Study – Java Enterprise Edition Servers

We implemented a subset of the proposed solution in JEE Server components mainly in JDBC Drivers [24, 25] and JMS Servers to improve the cache hit ratio and message throughput respectively. In this section we present an overview of these solutions. We have used ARMA modeling represented by the Equation (1)

$$yr(t+1) = ay(t) + bu(t) \quad (1)$$

$y(t)$ = The current output of query throughput or message Throughput
$u(t)$ = the current input of maximum number of queries or Subscribers
a and b = the model parameters to be estimated
$yr(t+1)$ = the output in the next step

The model parameters a and b are estimated using the sample data and are assumed to be fixed during performance tuning. A hybrid control approach is used to regulate the query throughput in JDBC drivers and message throughput in JMS Providers as shown in Fig. 5. The output $yr(t+1)$can be regulated by adjusting the controller gains. The Adaptive Control is an implementation of Time-Series exponential smoothing in conjunction with the Fuzzy control

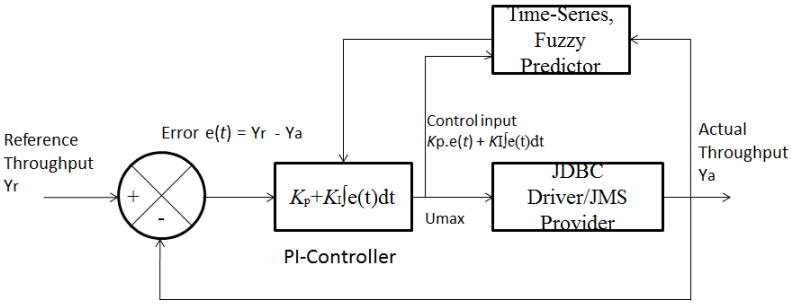


Fig. 5. Self-Managing Database driver and JMS Provider

7 Implementation and Analysis

In this section we explain the implementation details of the Autonomic aware UML framework discussed in the section 4 and the performance analysis of the JDBC Driver query throughput and message throughput of the JMS Providers.

7.1 Modeling

The autonomic aware UML framework is implemented using the USE UML Tool [26]. The current implementation though is an external plug-in, intend to modify the UML Tool source such that framework is a first class feature of the UML. The Fig. 6 below shows the UML diagram of the framework. The details this framework are discussed in [27].

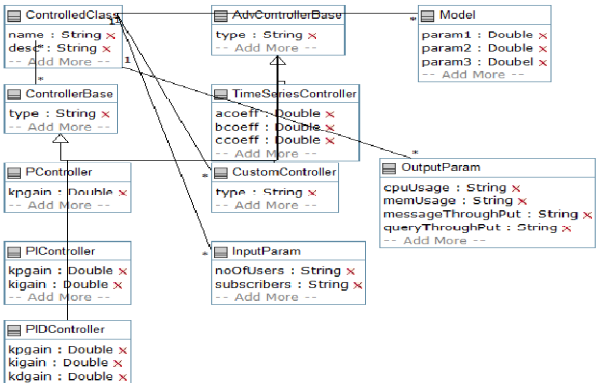


Fig. 6. Autonomic aware UML Framework

7.2 JDBC Driver

We implemented feedback control based solutions to improve the cache hit ratio in JDBC Drivers and evaluated the different approaches. The same is shown in the Fig. 7 below where we can clearly observe that the Time-Series Control in conjunction with Fuzzy control has a better cache hit ratio as against a LFU and Time-Series control. The detailed analysis of the solution is discussed in [24, 25].

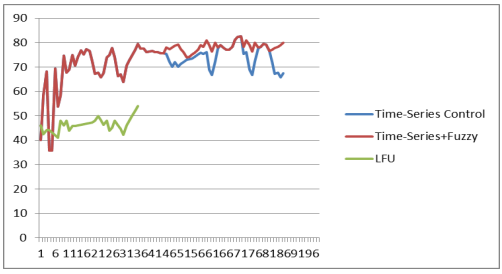


Fig. 7. Cache Hit Ratio

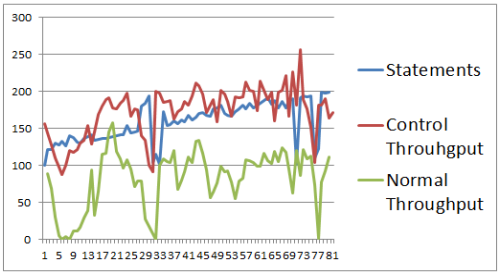


Fig. 8. Controlled Query Throughput

The Fig. 8 above shows query throughput at the JDBC Driver level run in large scale enterprise environments. The results shown below are generated on experiments in simulated environment where. A hybrid control using PI and Fuzzy control is implemented to regulate the query throughput where the statements are generated randomly in a range of 100-150 statements and the corresponding query throughput is measured.

7.3 JMS Providers

We implemented a P-Controller and a Time-Series Adaptive Controller and extended with a hybrid control mechanism consisting of PI, Fuzzy control to regulate the message throughput against varying subscribers over a period of time. The sample data is collected by running Active MQ Server and the solution is applied offline to measure the performance. The Fig.9 shows distinct performance improvement when the control system solution is used as against the throughput under normal conditions. The details of modeling, control algorithms and implementation analysis are discussed in [28, 29].

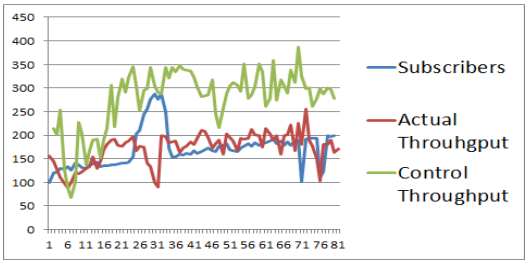


Fig. 9. Controlled Message Throughput

8 Conclusion and Future Work

Our investigations proved that control system applications in building self-managing software are very encouraging. Our experiments have shown improvement in the cache hit ratio and query regulation of database drivers, improved the message throughput in Message Servers using classic controllers and partially hybrid controllers. Our major objective is to build an Expert Control System which handles intelligent performance tuning of the major components of JEE Servers and also provide interactive system where the control system elements are first class features of the Distributed System Middleware servers. We are working in applying our proposed solution in EJB, ESB and OSGi servers. We are also working in extending the autonomic aware UML as integral part of UML Tool. We are working in run time modification of the model parameters based on the running conditions of the system.

References

1. Akerkar, R., Lingras, P.: An Intelligent Web: Theory and Practice, 1st edn. Johns and Bartlett, Boston (2008)
2. Adaptive Systems Research, http://www.research.ibm.com/PM/
3. Nami, M.R., Sharifi, H.: Autonomic Computing: A New Approach. AMS (2007)
4. Russell, L.W., Morgan, S.P., Chron, E.G.: Clockwork A new movement in autonomic systems. IBM Systems Journal 42(1), 77–84 (2003)
5. Gullapalli, R.K., Muthusamy, C., Vinaya Babu, A.: Control Systems application in Java based Enterprise and Cloud Environments – A Survey. IJACSA 2(8) (2011)
6. Astrom, K.J., Wittenmark, B.: Adaptive Control. Pearson Education, Control Engineering(2009), http://en.wikipedia.org/wiki/Control_engineering (accessed March 30, 2012)
7. Shiy, D., Tsungz, F.: Modelling and diagnosis of feedback-controlled processes using dynamic PCA and neural networks. INT Journal of Prod. and Res. 41(2) (2003)
8. Ryu, S., Cho, C.: PI-PD-controller for robust and adaptive queue management for supporting TCP congestion control, pp. 132–139, 18-22 (2004)
9. Kang, K.-D., Oh, J., Son, S.H.: Chronos: Feedback Control of a Real Database System Performance. In: Real Time Systems Symposium, pp. 267–276 (2007)
10. Zhu, X., Uysal, M., Zhikui, Wang, Singhal, S., Merchant, A., Padala, P., Shin, K.: What Does Control Theory Bring to Systems Research? ACM SIGOPS Operating Systems Review 43(1) (2009)
11. Li, B., Nahrstedt, K.: Impact of Control Theory on QoS Adaptation in Distributed Middleware Systems. In: American Control Conference, vol. 4, pp. 2987–2991 (2001)
12. Zhang, L., Ardagna, D.: SLA Based Profit Optimization in Autonomic Computing Systems. In: ICSOC (2004)
13. Middleware, http://en.wikipedia.org/wiki/Middleware (accessed March 30, 2012)
14. JEE, http://www.oracle.com/technetwork/java/javaee/tech/index.html (accessed March 30, 2012)
15. Application Server, http://en.wikipedia.org/wiki/Application_server (accessed March 30, 2012)
16. Klein, C., et al.: A Survey of Context Adaptation in Autonomic Computing. In: ICAS (2008)
17. Lu, Y., Abdelzaher, T., Tao, G.: Direct Adaptive Control of A Web Cache System. In: American Control Conference, Denver, Colorado (2003)
18. Robertsson, A., Wittenmark, B., Kihl, M., Andersson, M.: Design and evaluation of load control in web server systems. In: IEEE American Control Conference
19. Abdelzaher, T., Lu, Y., Zhana, R., Henriksson, D.: Practical Application of Control Theory to Web Services. In: American Control Conference (2004)
20. Zhang, Y., Qu, W., Liu, A.: Adaptive Self-Configuration Architecture for J2EE-based Middleware. In: HICSS 2006, vol. 9 (2006)
21. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A.: A Study of Intelligent Controllers Application in Distributed Systems. INDJCSE 2(4)
22. Lama, P., Zhou, X.: Autonomic Provisioning with Self-Adaptive Neural Fuzzy Control for End-to-end Delay Guarantee. In: IEEE International Symposium on Modeling, Analysis and Simulation of Computer Telecommunication Systems (2010)
23. Patikirikorala, T., Colman, A.: Feedback controllers in the cloud, Swinburne University (2011)

24. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A., Marndi, R.N.: Autonomic Database driver – An Adaptive Control Solution. In: ICITEC, Bangalore
25. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A., Marndi, R.N.: A Feedback Control Solution in Improving Database Driver Caching. IJEST 3(7) (2011)
26. USE UML Tool, http://www.db.informatik.uni-bremen.de/projects/USE/ (accessed December 25, 2011)
27. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A.: Design and Modeling Autonomic aware Software in UML – A Control System Solution. In: ICCIT, Tirupati (2012)
28. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A.: Throughput Regulation of Messaging Servers – An Intelligent Control Solution. IJACSA 3(1) (2012)
29. Ravi Kumar, G., Muthusamy, C., Vinaya Babu, A.: Self-Managing Message Throughput in Enterprise Messaging Servers. IJARCSSE 2(4) (2012)

An Embedded Navigation System for Aiding People with Alzheimer's Disease

Siddalingesh Navalgund[1], Jayashree Taralabenchi[2], Kavana Hegde[2], and Soumya Hegde[2]

[1] Department of Electronics and Communication Engineering, S.D.M. College of Engineering and Technology, Dharwad-02, Karnataka, India
siddunavalgund@yahoo.com
[2] Students, Department of Electronics and Communication Engineering, S.D.M. College of Engineering and Technology, Dharwad-02, Karnataka, India
jayashree.taralabenchi@gmail.com, kavanahegde6@gmail.com, soumyahegde7@gmail.com

Abstract. Alzheimer's disease is a progressive, irreversible brain disease that destroys memory and thinking skills. It is a part of dementia, a loss of memory and intellect that interferes with the daily life activities of Alzheimer patients. They even forget what they are doing and what they want to do. In this paper an attempt is made to provide solution for the memory part of the problem. This paper is aimed to design program based hardware navigation system, which helps the people with Alzheimer disease by alerting them if they are not moving in the predefined path. For example, if a person wants to go to the college and he goes somewhere else, then this system will inform him with proper message. If the person is performing the intended task, then the system just monitors it. This prototype design can be implemented on 8051 microcontroller using GPS module.

1 Introduction

1.1 Alzheimer Disease

Alzheimer's disease is a slowly progressive disease of the brain that is characterized by impairment of memory and eventually by disturbances in reasoning, planning, language and perception. Many scientists believe that Alzheimer's disease results from an increase in the production or accumulation of a specific protein (beta-amyloidal protein) in the brain that leads to nerve cell death. The main risk factor for Alzheimer's disease is the age. It is known that Ten percent of people over 65 years of age and 50% of those over 85 years of age have Alzheimer's disease [1]. Unless new treatments are developed to decrease the likelihood of developing Alzheimer's disease, the number of individuals with Alzheimer's disease will be increasing. Alzheimer's disease is a neurological brain disorder named after a German physician, Alois Alzheimer, who first described it in 1906 [1].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 639–643.
springerlink.com

1.2 Global Positioning System (GPS)

GPS is a space-based satellite navigation system that provides the location and time information anywhere on or near the Earth. GPS satellites circle the earth twice a day in a very precise orbit and transmit signal information to the earth. GPS receivers take this information and use triangulation to calculate the user's exact location. Essentially, the GPS receiver compares the time when a signal was transmitted by a satellite with the time it was received. The time difference tells the GPS receiver how far the satellite is. Now, with the distance measurements, the receiver can also determine the user's position and display it on the unit's electronic map [2].

1.3 P89V51RD2 Microcontroller

The P89V51RD2 is an 80C51 microcontroller with 64 kB Flash and 1024 bytes of data RAM. A key feature of the P89V51RD2 is its X2 mode option. The design engineer can choose to run the application with the conventional 80C51 clock rate. The Flash program memory supports both parallel programming and in serial In-System Programming (ISP). The P89V51RD2 is also In-Application Programmable (IAP), allowing the Flash program memory to be reconfigured even while the application is running. The operating voltage is 5V and the operating frequency is from 0 to 40 MHz [3].

2 Motivation and Design

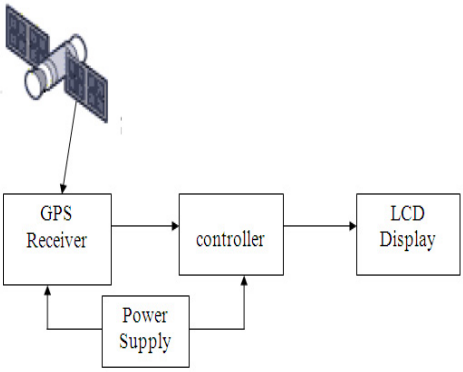


Fig. 1. Block diagram of the system

Now a days due to lack of nutritious food and pollution in the environment, people are suffering from many health hazards. Alzheimer disease is one among them, which is mostly found in old ages. They have tendency to forget daily routines. Sometimes the Alzheimer disease leads to death of the patients. So, there is a need for the system that helps them to do their daily needs. The design of the prototype system, which guides the patient to reach the destination efficiently, is attempted in this paper. The block diagram of the proposed system is as shown in the figure 1.The system alerts the patient/user, if he is moving in the undesired path.

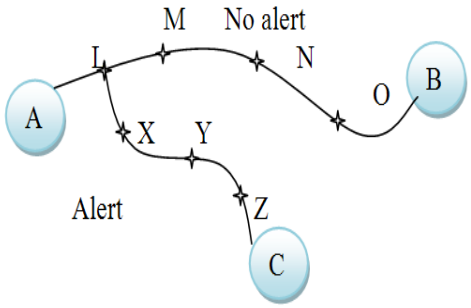


Fig. 2. Considering the intermediate points of the destination

In this paper, two methods are suggested to implement proposed idea i.e. to the guide the user. One method guides the user by informing him the exact location of the destination, which is as shown in the figure 4. Consider that user wants to move from point A (source) to point B (destination). Here some predefined points (their longitude and latitude) are saved to guide the person. The longitude and latitude of the source and destination are also stored. If the person moves in a different path, then the system alerts him and displays the information of the right point. In figure 2 the intermediate positions are shown by the points L, M, N and O. If person moves through these intermediate points then system will not alert him. But if he moves through X, Y, Z points, then system will inform him to follow the correct path.

The second method is as shown in figure 3. In this method an imaginary circle of predefined radius is considered around the source. The region within this circle is considered as the safe region. This method finds more applications in practical scenario due to uneven paths between sourceand destination. The difference between the longitude and latitude of source and longitude and latitude of the destination is calculated for every new GPS data. If the difference is decreasing then it is considered that the person is moving towards the destination hence no alert should be produced.

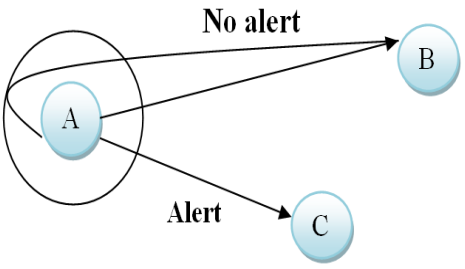


Fig. 3. Considering the imaginary circle

For example, a person wants to move from position A to position B as shown in the figure 3. If he moves in correct path it will not alert. Even if he deviates from the destination within the predefined circle then the system will not alert immediately. But, if he moves away from the circle then the system alerts him by informing about the destination. It also alerts him if he is moving within the circle for more than the specified time.

3 Implementation

The overview of the system is as shown in the figure 4. It consists of microcontroller, GPS module, LCD, Buzzer, battery, voltage regulator, crystal oscillator and voltage converter. For tracking and alerting Alzheimer patients, the GPS must be

Fig. 4. Overview of the system

interfaced with the 8051 microcontroller. Here the GPS receiver gets the information of the location by tracking the satellites and loads it to the microcontroller.

The output of the GPS is in the NMEA (National Marine Electronics Association) format [4]. This format gives the information about the latitude, longitude, speed and the direction of the GPS receiver. The output of the GPS is sent to the microcontroller at the frequency of GPS L50[5]. Then every GPS output of the system is compared with the initially stored information about the destination. If it matches with the stored data, then the system takes that person has already reached the destination. If differences of present latitude, longitudes and the stored latitude and longitude of the destination are decreasing, then the person is moving in the right direction. Hence, it will not alert the person in both the cases. If difference is increasing above the tolerance value (here tolerance value is 100m) then a beep is produce as a sign of alert.

4 Advantages of the Proposed System

Today we have many alerting systems. But these existing systems alert the person based on the time rather than the position i.e. it alerts the person at a particular time irrespective of his position. It alerts him even if he even moves in the right direction, which is unnecessary disturbance to the person who may be busy with his work and get confused by the wrong alert. In the existing system, there is wastage of power due to unnecessary alert. The designed system takes less power. Because it consumes 5V of power and alerts only when the person moves away from the right path at that particular time. Hence, the designed system avoids the wastage of time and power.

5 Results

The system operates in two different modes. They are view mode and store mode. In view mode, the latitude and longitude of current location, GMT (HH:MM), date (DD/MM), number of satellites in view (SV) and valid bit (F) is displayed. It is as shown in figure 5. If the valid bit is 1 then, we can store the value of that location else the data is insufficient to store. Figure 5a shows the data before tracking the satellite. As no data is received by the receiver hence no data is displayed. As soon as receiver gets the data, it is displayed on LCD as shown in figure 5b. Here receiver could track 9 satellites. The value of latitude and is 15 (degree) 27 (minutes) 3 (seconds) towards north and the value of longitude is 75 (degree) 01 (minutes) 3 (seconds)

towards east. We neglect the last two digits of seconds as they keep on fluctuating. Hence the accuracy is limited to 75meters i.e. we cannot differentiate two positions with distance less than 75m.

a b c

Fig. 5. LCD output at different instances a) before tracking the satellites b) after tracking the satellites c) during the storage of data

The latitude and longitude of required location can be stored as shown in figure 5c. GPS outputs should be taken in an open space so that the GPS antenna can receive minimum 3 satellites and give the expected result.

Fig. 6. LCD output showing the alert message

When person is deviating from the desired path, the system alerts the user. The alert message is as shown in the figure 6. This helps the user to reach the destination. Here the location shown in the figure 6b is taken as the destination point. When person moves away from it i.e. if the latitude is 15 (degree) 27 (minutes) 3 (seconds) then the system alerts the user.

Acknowledgments. The authors thank the authorities of Sri Dharmasthala Manjunatheshwara College of Engineering and Technology, Dhavalgiri, Dharwad, Karnataka, India for encouraging us for this work. Authors would like to thank ICAdc 2012 International conference Advances in computing, M.S.Ramaiah Institute of Technology Bangalore, India, for giving an opportunity to present the paper and publish this paper. Thank to the unknown referees whose critical reviews help improve the quality of the paper.

References

[1] http://www.medicinenet.com/alzheimers_disease_causes_stages_and_symptoms/article.htm#tocc
[2] http://www8.garmin.com/aboutGPS/
[3] http://saimete.edu.vn/attachments/259_p89v51rd2.pdf
[4] http://www.tronico.fi/OH6NT/docs/NMEA0183.pdf
[5] http://peti.webege.com/pdf/L50_HD_V1%200.pdf
[6] CSC 522 Embedded Systems Summer (2006)
[7] http://nancyfanwell.en.made-in-china.com/product/vMHxIeVErykF/China-GPS-Navigation-Device-Receiver-Antenna-LR9500-.html

Software Licensing Models and Benefits in Cloud Environment: A Survey

Mohan Murthy M.K., Mohd Noorul Ameen, Sanjay H.A., and Patel Mohammed Yasser

Nitte Meenakshi Institute of Technology, Bangalore
{maakem,ameen1176,sanju.smg,yasserpatel87}@gmail.com

Abstract. The impact of Cloud on Software industry is significant. There are many advantages with a software hosted in cloud. Software licensing models should consider the new parameters which are introduced by the cloud. In the cloud world user can use his own software license to deploy the software in VM (Bring Your Own License) or he can purchase a Virtual Machine with the required software, in this case the software license charges are included with the Virtual Machine price (License Charges Included). The service providers use different software licensing models, user has to select licensing model and licensing option (BYOL or LCI) according to his requirement.

Keywords: Cloud, SaaS, Bring Your Own License, Software License.

1 Introduction

In cloud computing world computing resources, network and software are given as metered services. Based on the service provided, three types of service models are defined in cloud: Software as a Service (SaaS), Infrastructure as a Service (IaaS), and Platform as a Service (PaaS). In this work we have considered the SaaS model. In SaaS software will be given as service. Ex - Salesforce.

The raise of cloud computing has brought significant changes to the traditional software model. Software licensing also got affected by the cloud. There are two licensing options available in cloud.

Bring Your Own license (BYOL) – In this option user can use his existing software license to deploy his application in cloud (Virtual Machine).

License Charges Included (LCI) – In this option the Virtual Machine is preloaded with the software. The software licensing charges are included in the Virtual Machine price. Obviously these Virtual Machines are costlier when compare to the Virtual Machines without software.

There are many licensing models available. User should choose the licensing option and licensing model according to his requirement.

The aim of this paper is to identify the benefits of the cloud for the software and software licensing and also to list some popular software licensing models. We will compare the pricing of BYOL and LCI licensing options for the same set of requirements. When comparing the prices we have considered IBM DB2 and Oracle database software. According to us, this is first attempt towards this work. In this work we

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 645–650.
springerlink.com

have identified the important benefits of cloud for software and software licensing and also we will list some popular software licensing models. This survey will put more light on the implications of cloud on software licensing and also helps in comparing the prices for BYOL and LCI licensing options.

Rest of the paper is organized as follows, section 2 explains the benefits of cloud with respect to software and software licensing, section 3 lists some popular licensing models and section 4 compares the pricing of BYOL and LCI licensing options.

2 Benefits of Cloud with Respect to Software

This section lists some important benefits of the cloud with respect to software.

2.1 Software Maintenance

In the cloud computing model user need not worry about the software update and applying the patch. The service provider automatically updates the software and applies a patch when they are available. Software will be automatically updated without user interventions.

2.2 Infrastructure for the Software

For hosting the software in traditional software model, user should setup the appropriate infrastructure. If the infrastructure capacity is more than the software requirement then user has unnecessarily invested his money in setting up the infrastructure, if the infrastructure capacity is less, then the application performance will be degraded also in many cases it is difficult to upgrade the infrastructure. In cloud setting up the infrastructure is taken care by the service provider.

2.3 Upgrading the Software

In traditional software model to upgrade the software to a higher version user should buy the license for new software because most of the software vendors do not support free up gradation and if the new software version needs more computing power (CPU, ram etc.) user should upgrade or change his infrastructure. This is a costly and time consuming procedure and also the license fees for the older version of the software that user has paid will be wasted. Cloud model addresses these issues. If the user wants to go for the higher version of the software, user can simply terminate his existing subscription and he can opt for the higher version software. Setting up the environment for the new version of the software is done by the service provider.

2.4 Upgrading the Software License

There may be a situation to upgrade the software license. For example in the beginning user may have purchased the license for basic functionalities of the software but after some days he may want to go for some more functionalities of the software. Since the new software has more functionality it requires more computing power. In traditional software model user should upgrade his infrastructure to host his new

application. As mentioned earlier this is a time consuming and costly job. In cloud user can upgrade his license very quickly.

2.5 License Mobility

User can use his existing license in the cloud. For example Amazon has introduced bring your own license concept, using this user can use existing software license in cloud. Microsoft also supports the license mobility.

2.6 Cloud Enabling Is Simple

Infrastructure providers are helping the software vendors to make their software cloud enabled. For example Amazon has a program called AWS Independent Software Vendor (ISV) Program [2]. Using Amazon's ISV program software vendors can make their software cloud enabled.

2.7 Cost Reduction

If the user is interested in short term usage of a software definitely the cloud model will be cheaper. In traditional software model even though the user is going to use the software for short term he has to pay a huge amount of license fee but in cloud model user has to pay only for the duration he has used the software.

3 Software Licensing Models

Many traditional software licensing models are available on the cloud with little modification or no modification at all. There is no standard approach for software licensing in cloud. According to Gartner [2] at present organizations with the majority of their IT infrastructure in the cloud, or via SaaS is 3%, Gartner expects that this will increase to 43% in four years. Different software vendors use different metric for pricing. For example IBM uses Processor Value Units. A Processor Value Unit (PVU) is a unit of measure used to differentiate licensing of software on distributed processor technologies (defined by Processor Vendor, Brand, Type and Model Number). [3]. When using Oracle applications on the Amazon EC2, Oracle pricing is based on the size of the EC2 instances. EC2 instances with 4 or less virtual cores are counted as 1 socket, which is considered equivalent to a processor license. For EC2 instances with more than 4 virtual cores, every 4 virtual cores used (rounded up to the closest multiple of 4) equate to a licensing requirement of 1 socket [4]. Soft partitioning which is segmenting the operating system using OS resource managers [5], is not permitted as a means to determine or limit the number of software licenses required for any given server.. User should consider all these factors before selecting the required software. During this survey we have come across with many licensing models. Some of the interesting license models are listed below.

3.1 Pay As You Go

In this model. user is going to pay for what he has used. As the software usage increases the billed amount will be increased. If the user requirement is short term and number of users are less user can go for this model. This model is observed in Amazon Relation Database Service (RDS) for oracle.

3.2 Subscription Model

If the user is interested for long term usage, he can go for subscription model. For example if the user wants to use a software for three months he can subscribe for the software for three months provided that service provider should support three months subscription option for that software. If that option is not present user can select the nearest time period subscription option. Most of the software vendors support both Pay as you go and subscription based model. The models explained in the next sections are combined with either Pay as you go or subscription model.

3.3 Processor Based

In this model the license price is directly propotional to the processor capacity. This model is cost effective when the number of users are more. Ex – This model we can observe in IBM DB2. Based on the processor spec IBM defines Processor Value Units for that processor. These Processor Value Units will be used to decide the pricing of the license.

3.4 Based on the Number of Users

As the number of users using the software increases price also increases in this model. This model is cost effective when the number of users are less. Ex - Microsoft supports Subscriber Access License (SAL) model, which is based on the number of end users connected.

3.5 Based on the Number of Transactions

This model can be observed in the database systems offered as SaaS. The price will be determined by the number of transactions made by the user. As the number of transactions increases price will get increased.

3.6 Based on the Subscription to the Functionalities

Enterprise software will be having many modules and functionalities. User may not be interested in all of those functionalities. In this case some software providers give the flexibility to select the individual module from the software system. Only for those modules which the user has selected will be charged. Ex – Based on the functionalities of its CRM (Customer Relationship Management) software Salesforce.com has introduced different editions of its software. Each edition is having different pricing. Sometimes software provider will be having different types of the same software Ex – Basic, Standard, Enterprise etc. These software types are different in terms of

functionalities, number of users supported etc. Price is also different for each of the software types. User can select the software type which suits his requirement. Ex – Oracle provides software types like standard enterprise etc

3.7 Free Software, Pay for Support

There are some software vendors which provide their software for free but the user has to pay for the support. Red Hat Enterprise Linux uses this model.

4 Pricing of BYOL and LCI

In BYOL licensing option user can use his existing license to host his application in cloud or he can purchase the software license separately and host the application in the cloud. Amazon has the option of BYOL. The table gives the pricing of Amazon EC2 standard small instance along with licensing charges of IBM DB2 Express Edition. From the tabulated data of table 1 we can conclude that if the user is going for a long term subscription then it is better to buy the software separately and host it on the cloud.

Table 1. Pricing details of IBM DB2 in Amazon EC2

Duration in months	Price in US Dollars without software license charges	Price in US Dollars including license charges	License charges in US Dollars
1	82.8	280.8	198
3	248.2	842.4	594.2
6	496.4	1684.8	1188.4
12	992.8	3369.6	2376.8

Table 2 gives the pricing details of Amazon RDS (Relational Database Service) for Oracle Database. We have considered Amazon's standard DB instance for Oracle's Standard edition one. Here also we can observe that for short term usage of the software the Amazon RDS along with the Oracle license is cheaper. For long term usage user can go for the Bring Your Own Licensing option. User should be careful before going to Bring Your Own Licensing option, because as we mentioned earlier he is going to lose several benefits of the cloud.

Table 2. Pricing details of Amazon RDS for Oracle

Duration in months	Price in US Dollars without software license charges	Price in US Dollars including license charges	License charges in US Dollars
1	75.6	111.6	36
3	226.8	334.8	108
6	453.6	669.6	216
12	907.2	1339.2	432

5 Conclusion

Traditional software model is affected by the cloud computing. In cloud world instead of purchasing, user will rent a software. User need not worry about the software maintenance, infrastructure and its maintenance in cloud environment. License charges are included in the Virtual Machine price. Pay as you go and subscription models are popular among the SaaS providers. Some of the service providers are giving the flexibility for the users to bring their own license to the cloud (BYOL). When the user is interested in long term usage of a software he can go for the BYOL licensing option. But he should be careful when going for the BYOL licensing option since he is going to lose several benefits of cloud. There are many licensing models available. User has to select the best licensing model according to his needs. In future we are going to work on a framework for software licensing in cloud environment.

References

1. Hurwitz, J., Bloor, R., Kaufman, M., Halper, F.: Cloud Computing for Dummies
2. http://betanews.com/2011/01/24/gartner-most-cios-have-their-heads-in-the-clouds/
3. http://www-01.ibm.com/software/lotus/passportadvantage/pvufaqgen.html
4. http://www.oracle.com/us/corporate/pricing/cloud-licensing-070579.pdf
5. http://www.oracle.com/us/corporate/pricing/partitioning-070609.pdf
6. http://aws.amazon.com/solutions/solution-providers/program/isv/

VLSI Architecture of Spread Sprectrum Image Watermarking Decoder

Navonil Chatterjee, Moudud Sohid, and Sudipta Chakraborty

Bengal Engineering and Science University, Shibpur, Howrah, India
navonilchatterjee@yahoo.in, hi.maddy08@gmail.com,
sudipta1788@gmail.com

Abstract. Multifarious techniques have been adopted in the field of image watermarking which offers some desirable characteristics to the watermarked image. In this paper, the combined technique of channel coding and Spread Spectrum modulation has been exerted to implement the proposed image watermarking algorithm. A block based spatial domain watermarking scheme for digital image using Spread Spectrum has been employed for the implementation of the algorithm. In practice, with the help of pseudo random noise the binary watermark image has been distributed using channel coding and spatial bi-phase modulation technique. With the extensive use of Spread spectrum modulation technique the security issues related to image watermarking can be effectively dealt with, and at the same time robustness and imperceptibility of watermarked image can be enhanced appreciably. VLSI implementation using Field Programmable Gate Array (FPGA) has been developed the decoding portions of the algorithm.

1 Introduction

The process of embedding hidden information into a digital signal which may be used as a means of concealed communication between the sender and recipient is called invisible Digital watermarking. The information may be embedded in the digital signal such as- audio, image or video [1]. These days the use of internet has become indispensable in man's life. The ever increasing efficacy of the internet and the ease of availability of inexpensive tools render it possible to manipulate the digital content. Therefore the requirement for protection of digital content against unauthorized replication and modification has increased at an alarming rate. Digital Rights Management techniques deal with the ownership rights of the digital content. Digital watermarking though not complete DRM mechanism can be utilized in DRM systems to prevent unauthorized access and undesirable manipulation, thereby assuring authentication [2],[3].

In Digital watermarking, information is embedded in the form of symbol, text or a number. In case of invisible watermark it must be ensured that the modification of the media, as a result of embedding information, is imperceptible. Depending on perceptibility, digital watermarking has been divided into two types- : (i) Visible Watermark (ii) Invisible Watermark. Classification can also be made on the basis of whether or

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 651–658.
springerlink.com

not any Transformation technique has been implemented, namely- (i) Spatial Domain Watermarking [1] (ii) Transform Domain Watermarking [5].

Information transmitted through perilous communication channels is prone to contamination with undesirable and pernicious data or information may be tampered. To prevent this, the concept of Pseudo random Noise sequence is widely used in the domain of image watermarking and in steganography. In order to ensure secure data communication through the insecure communication channels, it is therefore necessary to utilize some stern and effective mechanism. Spread Spectrum watermarking technique is such a mechanism where, the information is spread over a larger frequency so that during extraction the integrity of the information is not surrogated. Spread spectrum (SS) watermarking may be used in fragile and semi fragile watermarking by keeping the chip rate low. The term chip rate indicates the number of cover signal's sample over which the watermark bit is spread [6], [2].

The objective of this paper is to design VLSI architecture for the given image watermarking algorithm that caters to the need of media authentication as well as secure transfer of image. Hardware implementation of digital watermarking provides several advantages over its software counterpart in terms of- less area requirement, low execution time, and less power consumption. The example of TV broadcast will highlight the significance where digital media is to be marked in real time and hardware is the only solution [9]. The solution to establish the chain of custody for forensic digital photographers can be cited as another example.

The architecture that has been created for the embedding portion of the given watermarking algorithm, enables watermarking to be implemented instantaneously at the time of capturing the image rather than using a software procedure that calls for greater execution time [10]. This semi-fragile image Watermarking algorithm is based on the combined technique of channel coding and Spread Spectrum modulation. In this paper, concentration has been focused on developing the architecture of the decoding section of the above mentioned watermarking algorithm. The architecture that has been employed to implement the decoding algorithm can be characterized as simple with low computation expenses and low IOBs. Its greatest advantage is provided by the fact that it can be easily implemented in hardware and thereby promoting it's acceptability in the field of watermarking.

The content of the paper has been organized as follows- Section 2 expounds the proposed watermarking algorithm. The architecture corresponding to the decoding portion of the proposed algorithm has been delineated in Section 3. An account of the digital design of the decoding algorithm and the results has been provided in Section 4. Section 5 illustrates the conclusion.

2 Proposed Watermarking Algorithm

The following two subsections describe different steps in watermark embedding and decoding process respectively. Here, Binary watermark is embedded directly to the pixel values of each block of the cover image using SS modulation. During decoding, watermark information is extracted using normalized correlation and the extraction of binary watermark is done using channel decoding and spatial bi-phase demodulation.

2.1 Watermark Embedding

The Spread spectrum (SS) watermarking using binary watermark in spatial domain is discussed in details in the following section. Different steps for watermarkembedding are described as follows:

Image Partitioning. The cover image is taken as F, where F = $\{F_{ij}, 1 \leq i \leq F_{length}, 1 \leq j \leq F_{width}\}$, while $F_{ij} \in \{0, 1, \ldots, 255\}$, F_{length} is the image length and F_{width} the width of image. Now we partition the cover image into (m x m) which is non-overlapping, where b = 4, 8, 16, 32 etc. Suppose we call them as F_{ij}, where 'i' is the number of rows and 'j' is the number of columns. Here we are partitioning the image keeping row major. Each of this (m x m) blocks are broken partition into (m/2 x m/2) blocks which are non-overlapping in nature. Suppose we call them as H_{ij}.

Formation of message vector. The message image is taken as W, where W = $\{W_{ij}, 1 \leq i \leq W_{length}, 1 \leq j \leq W_{width}\}$, while $W_{ij} \in \{0, 1\}$, W_{length} is the image length and W_{width} the width of image. We partition the watermark image into (L x L) non overlapping blocks, taking row major. We get a matrix G_{LxL}, therefore we drive a vector B = $\{b_1, b2, \ldots\ldots, bj\}$, $b_j \in \{0, 1\}$, the same size as the partitioned image.

Generation of PN code. We derive the vector S_{LxL} from the PN (Pseudo Noise) sequence generated from the polynomial defined for a particular image length over which the message would be embedded. S = $\{s_1, s2, s3, \ldots..,s_{LxL}\}$, $s_i \in \{0, 1\}$. The vector Z is created by $z_i = 2s_j - 1$, where $z_i \in \{1,-1\}$. If there are equal numbers of zeroes and ones are present is S then the vector Z will be a vector with zero mean. We have to generate 4 PN (Pseudo Noise) codes of length (n x n), where n = 4, 8, 16 etc.

Watermark Embedding. We now embedded the cover image with the binary water mark using the Spread Spectrum (SS) watermarking scheme. The rule is given as:

$F^e = F + KS$ if b_j = ***'0'***
$F^e = F - KS$ if b_j = ***'1'***

Where

F^e = Embedded image in spatial domain.
F = Cover image.
K = Modulation Index.
S = PN code.

The value of modulation index is calculated through experimental result evaluation.

2.2 Watermark Image Extraction & Message Decoding

The watermark recovery process requires the sets of PN matrices that were used for embedding the message into the cover image. The watermarked image is divided into blocks of (M x M) size and each (M x M) is further divided into (M/2 X M/2) blocks which would 1 bit information of the message image. The process of extracting the information from the embedded image is given below:

Watermarked Image partitioning. The received image is taken as F*, where error have been introduced by attackers or through channel noise. F* is represented as F* = $\{F^*_{ij}, 1 \leq i \leq F^*_{length}, 1 \leq j \leq F^*_{width}\}$, while $F^*_{ij} \in \{0, 1, \ldots, 255\}$, F^*_{length} is the image length and F^*_{width} the width of image. The received image is broken into (n x n) non overlapping blocks, where n is same as that for cover image. Now each (n x n) is broken down into H_{ij} that is (n/2 x n/2) block.

Zero Mean Calculation. Zero mean normalization is done on each block of H_{ij}, suppose we consider k^{th} block. We have H^k_{ij}, where $H^k_{ij} = \{h_{1.j}, \ldots\ldots\ldots, h_{i,1}, \ldots\ldots h_{ij}\}$. The zero mean normalization is given as:

$$h^{new}{}_{ij} = \frac{h_{ij} - \mu_{H_{ij}}}{\sigma_{H_{ij}}}$$

Zero mean and unit variance normalization is quite simple; we just remove the mean intensity value from an image and then scale it with its variance. Here the mean is taken as μ and the variance is given as σ.

Correlation Coefficient Calculation and Message Extraction. We calculate the correlation coefficient values for each block H^k_{ij}. The correlation coefficient, sometimes also called the cross-correlation coefficient, is a quantity that gives the quality of a least squares fitting to the original data. The image equality assessment is carried out using this methodology. The correlation coefficient is calculated as:

$$r = \frac{\sum_m \sum_n (H_{ij} - \overline{H}_{ij})(K - \overline{K})}{\sqrt{\left(\sum \sum \left(H_{ij} - \overline{H}_{ij}\right)^2 \left(K - \overline{K}\right)^2 \right)}}$$

Here we see that due to zero meaning of the watermarked image the value of $\overline{H}_{ij}$ is equal to zero and also if the number elements of *K* (Pseudo Noise) is equal and opposite {1,-1} then $\overline{K}$ is equal to zero. We extract the message embedded in the watermarked image by comparing the value of correlation coefficient as show:

(i) $r \leq 0$ then the extracted bit is '1'
(ii) $r > 0$ then the extracted bit is '0'

3 Architectural Design of the Proposed Algorithm

The architecture of the proposed watermark algorithm is shown in figure 1. The flowchart of the proposed architecture is shown in figure 2. The main building blocks are as follows: (I) Controller (II) Linear Feedback Shift Resister (III) Memory (i) Watermark Memory (ii) Memory Embed (IV) Normalizer (V) Correlator.

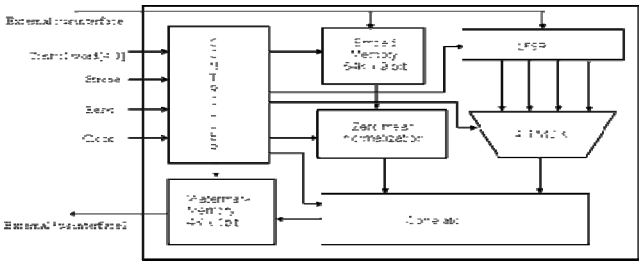


Fig. 1. Decoding Architecture

3.1 Controller

The controller is controlled by the phase control word. The control word is defined by a 4 bit register. The configuration and operation of each bits is described in Table 1.These phases are triggered by the corresponding bit pattern of the control word. In Phase LFSR the four LFSRs are loaded with the seed value and the PN sequence is generated. The PN sequences formed are stored in four 5 bit registers. In the next phase i.e. Phase Initial, data (embedded data) is loaded in the Memory Embed. In Phase Middle, normalization of partitioned embedded image takes place. B1, B2, B3, B4 (the four bits) is calculated by performing correlation between the sub-blocks and respective PN sequence. The output of the correlator is stored in the Memory Watermark. Data is extracted from the Memory Watermark through the external bus interface2 in Phase End. The strobe signal is used for activation of the system.

Table 1. Phase Control Word

CONTROL WORD	STATE OF OPERATION	DESCRIPTION
0001	Phase LFSR	Set LFSR
0010	Phase Initial	Load data in Embed Memory
0100	Phase Middle	(i) Decoding Operation (ii) Resultant data is stored in Watermark Memory
1000	Phase End	Data read from Watermark Memory

3.2 Linear Feedback Shift Register

The LFSR will be enabled by the controller in the Phase LFSR when it sets set_seed = '1'. The seed is fed by the external interface by the user. The PN (Pseudo Noise sequence) is generated using the feedback mechanism. In the above stated algorithm we used PN sequence of length 16 bits. Depending on the polynomial the LFSR work in accordance to the taps to generate the PN sequence.

3.3 Memory Block

The Memory Embed is 2^{16} x 9 bits memory which stores the embedded image. The MSB of embedded data is the sign bit. Next we have the memory water which is 2^{12} x

1 bit long. The extracted binary watermark image data is stored in this memory in Phase Middle.

3.4 Normalizer

The zero mean normalization takes place in Normalizer block. The partitioned watermark (8x8) is broken into subsets (4x4). Mean of each subsets are calculated and the result is subtracted from individual pixel using an ALU. Each element of these subsets is divided by corresponding variance. The elements of normalized block (8x8) are represented by 13 bits, where last 4bits are for fraction, the next 8 bits represent integer value and MSB is the sign bit.

3.5 Correlator

In this block the output data of the Normalizer are provided. Each subset which consists of 16 elements, each of which is multiplied with corresponding PN sequences (S1, S2, S3, and S4) which are 16 bit long. The outputs are added cumulatively and the result is divided by 16. A comparator is used to calculate where resultants are greater than or less than the threshold, if it greater then the extracted bit is 0 else 1.

4 Simulation and Synthesis of Digital Design

The simulation process is initiated by providing the seed values to the lfsr. Four such LFSR outputs are stored in four SISO registers. The watermarked embedded image is loaded into the embed memory. Using the controller data is read from the embed memory and the algorithm proposed is implemented. The output result is stored in watermark memory (64x64). In the simulation result is shown in figure 3, the four LFSR output that is fed into the SISO (serial in serial out) register are shown. The data_in bus is used to load the embed memory. The intermediate Normalizer and Correlator signals are also shown. The phase is governed by the user to initiate the control signal for the execution of the decoding procedure. The design is also facilitated with a strobe which gives the user an extra degree of freedom in controlling the functionality. The result is read from the watermark memory using data_bus_out.

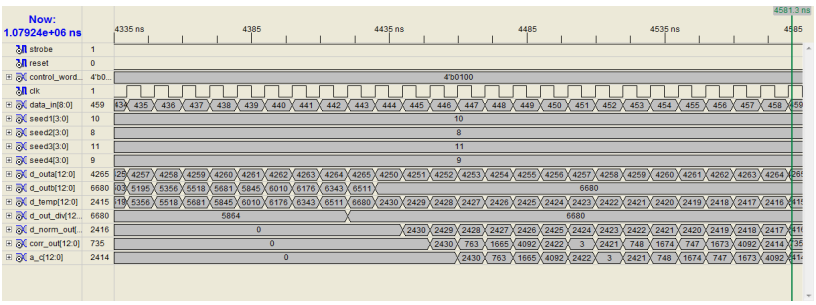


Fig. 2. Behavioral Simulation output shown in ISE for the top level module

The synthesis of both the watermark embedding and decoding have been implemented on Xilinx (ISE version 8.1) based FPGA. We have chosen Virtex series of FPGA to fit the complexities of the design. The device used is xc2vp30-7ff896 for the implementation of the design and the language used is VHDL. The device utilization summary is stated in Table:2. The behavioral simulation was done with ISE simulator to verify the functionality of the design. A test bench is also written in VHDL to give the input vectors for the simulated program. Timing summary for the design: (i) Minimum period: 23.631ns (Maximum Frequency: 42.317MHz) (ii) Minimum input arrival time before clock: 3.087ns (iii) Maximum output required time after clock: 4.325ns.

Table 2. Device Utilization Summary

Logic Utilization	Used	Available	Utilization
Number of Slices	327	13696	2%
Number of Slice Flip Flops	363	27392	1%
Number of 4 input LUTs	520	27392	8%
Number of BRAMs	37	136	27%
Number of Bonded IOBs	45	556	8%
Numbger of GCLKs	1	16	6%

5 Conclusion

The paper proposes a spatial image watermarking scheme implementing spread spectrum modulation technique increasing robustness and perceptual quality of watermarked image. The algorithm is simple with low computation cost and can be easily implemented in hardware. Digital design of the proposed algorithm using FPGA is also developed and thus makes it suitable for real time authentication as well as secured communication. Current work is going on to develop a private key that would enhance the security factor and try for ASIC implementation.

References

1. Ghosh, S., Ray, P., Maity, S.P., Rahaman, H.: Spread Spectrum Image Watermarking with Digital Design. In: IEEE International Advance Computing Conference (IACC 2009), pp. 868–873 (2009)
2. Special issue on copyright and privacy protection. IEEE Journal on Selected Areas in Communication (JSCA) 16(4) (May 1998)
3. Special issue on enabling security technologies for digital right management. Proceedings of IEEE 92(6) (June 2004)

4. Paquet, A.H., Ward, R.K., Pitas, I.: Wavelet packet-based digital water-marking for image verification and authentication. Signal Processing 83, 2117–2132 (2003)
5. Cox, I.J., Kilian, J., Leighton, T., Shamoon, T.: Secure spread spectrum wa-termarking for multimedia. IEEE Transaction on Image Processing 6(12), 1673–1687 (1997)
6. Maity, S.P., Kundu, M.K., Das, T.S.: Robust SS Watermarking with improved capacity. In: Pattern Recognition Letters "Advances in Visual Information Processing", vol. 28, pp. 350–357. Elsevier (2007)
7. Maity, S.P., Kundu, M.K.: A blind CDMA image watermarking scheme in wavelet domain. In: Proc. of IEEE Int. Conf. on Image Proc., Singapore, pp. 2633–2636 (2004)
8. Mathai, N.J., Kundur, D., Sheikholeslami, A.: Hardware implementation perspectives of digital video watermarking algorithms. IEEE Transaction on Signal Processing 51, 925–938 (2003)
9. Maity, S.P., Banerjee, A., Abhijit, A., Kundu, M.K.: VLSI design of Spread Spectrum Image Watermarking. In: 13th National Conference on Communications (NCC-2007), Indian Institute of Technology, Kanpur, India, January 26-28, pp. 251–257 (2007)
10. Maity, S.P., Kundu, M.K., Mandal, M.K.: Performance Improvement in Spread Spectrum Watermarking via M-band Wavelets and N-ary Modulation. In: IET International Conference Visual Informal Engineering (VIE 2006), September 26-28, pp. 35–40 (2006)

Strategy Driven Approach for the AD HOC Network Participants Using the Notion of Trust and Activity

Shabana Sultana[1] and C. Vidya Raj[2]

[1] Department of Computer Science & Engg.,
The National Institute of Engg., Mysore, India
Shabnamkbn2k@yahoo.co.in
[2] NIE Institute of Technology, Mysore, India
vidya_rajc@yahoo.com

Abstract. A wireless ad hoc network is characterized by a distributed, dynamic, self organizing architecture. one of the problems that cause severe degradation in system performance of ad hoc network is dropping packets by misbehaving nodes. Non-cooperative actions of nodes are usually called selfishness; selfish nodes try to benefit from other nodes but refuse to forward packets of other nodes. using the notion of trust and activity, a strategy driven approach to enforce cooperation among participants of ad hoc networks is proposed in the paper.

Keywords: Manets, Game Theory, Nash Equilibrium, Strategy, Pay-off.

1 Introduction

An ad hoc network is a collection of wireless mobile hosts forming a temporary network without the aid of any established infrastructure or centralized administration[1]. In such an environment, it may be necessary for one mobile host to enlist the aid of other hosts in forwarding a packet to its destination, else to the limited range of each mobile host's wireless transmissions. Indeed, as apposed to networks using dedicated nodes to support basic networking function like packet forwarding and routing, in ad hoc networks these functions are carried out by all available nodes in the network. However, there is no good reason to assume that the nodes in the network will eventually cooperate, since network operation consumes energy, a particularly scarce resource in a battery powered environment like MANET. The lack of cooperation between the nodes of a network is a new problem that is specific to the adhoc environment and goes under the name of node selfishness. A selfish node does not directly intend to damage other nodes by causing network portioning or by disrupting routing information but it simply does not cooperate to the basic network functioning, saving battery life for own communications.

Game theoretic methods are applied to study cooperation. Game theory is a powerful tool for modeling interactions between self interested users and predicting their choice of strategy. Each player in the game maximizes some function of utility in a distributed fashion. The games settle at a Nash equilibrium if one exists, but since

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 659–665.
springerlink.com

nodes act selfishness, the equilibrium point is not necessarily the best operating point from a social point of view.

In game theoretic terms cooperation in mobile network can be interpreted as a dilemma[2] nodes (players) are tempted to get benefit (ability of sending packets) without cost (contribution to packet forwarding). However if such behavior is noticed by other nodes then selfish node may end up at being excluded from the network selfish behavior would be risk free if a cooperation enforcement mechanism did not exit.

In this paper, we address the problem of the selfish behavior in self policing ad hoc networks our approach aims at enforcing cooperation. We propose a strategy driven behavior of network participants. The decision whether to forward or discard packets depends on the reliability of a source of the packet. The calculation of reliability is based on the notion of trust and activity. The calculation of trust based on the ratio of packets discarded Vs packets forwarded by the node, while activity is based on the amount of time spent in the idle node.

We propose a new game theoretic-based model of the ad hoc network whose goal is to evaluate strategies. This model has some similarities with the Iterated Prisoner's Dilemma under the Random Pairing (IPDRP) game in which randomly chosen players receive payoffs that depend on the way they behave [3]. A GA is used to search for good strategies.

2 Strategy Based on Trust and Activity

2.1 Evaluation of Trust

We assume that each node uses an omni-directional antenna with the same radio range. A source routing protocol is used, which means that a list of intermediate nodes is included in the packet's header. In one model the reputation information is gathered only by nodes directly participating in the packet forwarding. Each node monitors the behavior of the next forwarding node.

Reputation data is collected in the following way let's suppose that node A wants to send a packet to node E using intermediate nodes B, C and D. (Fig 1a)

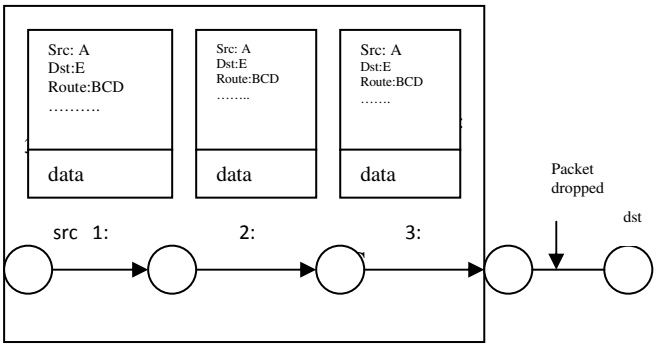


Fig. 1(a). Node A sending a packet to node E using Intermediate nodes B, C and D

If the communication is successful then node E receives the packet and all nodes participating in that forwarding process update reputation information about each other. If communication fails (for example node D decides to discard the packet) this event is recorded by the watching mechanism of the node C. In such case node C forwards alert about selfish behavior of node D to the node B and then node B forwards it to the source node A. Analysis of such a way of collecting reputation data can be found in [4].

Suppose that node B wants to verify how trustworthy is node A (using available reputation data concerning node B). In order to do this, first the fraction of correctly forwarded packets by node B is calculated (forwarding rate) and then the trust lookup table is used Fig.1(b).

fr(B,A) $\frac{P_{fA}}{P_{SA}}$ →

fr	TL_{BA}	
1 – 0.9	3	← highest trust level
0.9 – 0.6	2	
0.6 – 0.3	1	
0.3 - 0	0	← lowest trust level

Fig. 1(b). Trust Lookup TABLE

fr(B,A) – forwarding rate of the node A
P_{fA} – number of packets forwarded by the node A
P_{SA} – number of packets sent to the node A (request to for forwarding)
TL_{BA} – Trust level of the node B in the node A.

As a result one of the four possible trust levels is assigned. For example, forwarding gate of 0.95 results in the trust level B.

If a source node has more than one path available to the destination it will choose the one with the best reputation. Such reputation is based on both forwarding rate and activity of intermediate modes. The reason for taking into account the activity level is related to the fact that there is the risk that a low activity node will switch to idle mode. As a result a path contacting that node will no longer be available.

A path rating is calculated a multiplication of all known forwarding rates and activity weights of all nodes belonging to the route. An unknown node has a forwarding rate set to 0.5.

A path rating is calculated as a multiplication of all known forwarding rates and activity weights of all nodes belonging to the route. An unknown node has a forwarding rate set to 0.5.

2.2 Evaluation of the Activity Level

The calculation of the activity level is based on the total number of packets forwarded by a node. Activity level serves as an indication of the time spent in idle mode. Three activity levels are defined: low (LO), medium (ME) and high (HI) levels.

Those levels are calculated using the same reputation data as used for trust evaluation. In order to verify the activity level of a source node an intermediate node calculates the average number of all packets forwarded by all known nodes (denoted as av).

This value is next compared with the number of packets forwarded by the source node. If this number belongs to a range < av -0.2 * av …. Av +av *0.2> then the medium activity level is assigned low activity level is assigned in the case when this number is smaller than that range while high level in the case when this number is above that range. The activity level is also used for the computation of the path rating. For this purpose so called activity weights are defined as follows: 0.2 for activity level 0, 0.5 for activity level 1 and 0-1 for activity level 2. The goal of activity weights is to promote paths containing more active nodes (as more reliable paths)

2.3 Coding the Strategy

The decision whether to forward or discard the packet is determined by the strategy represented by a binary string of length 13. An example of a strategy is shown in Fig1. The exact division is based on two elements: trust level in the source node and its activity level. There are 12 possible combinations of trust and activity levels. Decisions for each case are represented by bits no. 0-11. Bit no. 12 defines behavior against unknown node. Decision F means "Forward packet" while D stands for the opposite (discard the packet). For example , suppose that node B receives a packet originally coming from node A. Assuming that node B has a trust level 3 in node A and node's A activity is high (HI) then according to the strategy shown in Fig (2). The decision would be to forward the packet (F, bit no.11)

Trust →	Trust 0			Trust 1			Trust 2			Trust 3			
Activity →	LO	ME	HI	LO	ME	HI	LO	ME	HI	LO	ME	HI	
Decision →	D	D	D	D	D	D	F	F	F	F	F	F	D
	0	1	2	3	4	5	6	7	8	9	10	11	12

D: discard intermediate packet
F: Forward intermediate packet

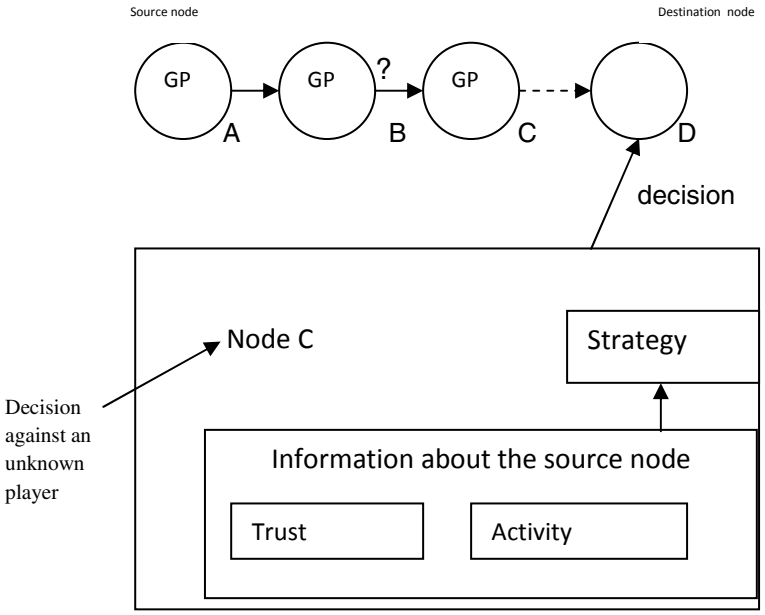
Fig. 2. Coding of the strategy

3 An ADHOC Gaming Model

3.1 Description of an Ad hoc Network Game

We define an Ad Hoc Network Game as a situation in which one node (player) is sending a packet to some destination and intermediate nodes have to decide whether to forward or to discard it. The number of game participants (GP) depends on the length of the path leading from the source to the destination node. Game participants

are composed of the source node and all intermediate nodes. All intermediate nodes are chosen randomly. This simulates a network with a high mobility level, in which the topology changes very fast. The destination node is not a part of the game. Each player is said to play his own game when being a source of a packet and is said to be a participant of other player's game when being an intermediate node. After the game is finished all its participants receive payoffs according to the decisions they made. In the example shown in Fig.(3) the game is composed of 3 nodes: node A, B and C.



GP: game participants.
Node A: Playing its own game.
Node B and C: Participants of other player's game.

Fig. 3. A single ad hoc network game

Node A is the source of the packet while nodes B and C are intermediate nodes asked to forward the packet. After the reception of the packet node B has to decide whether to forwarded or to discard the packet received from the node A. If node B decides to discard the packet then the game ends. Otherwise, it is the turn of node C to divide what to do with the packet. If all intermediate nodes divide to forward the packet, the communication is successful. After the game is finished all its participants receive payoffs according to the decisions they made.

4 Payoff Table and Fitness Function

The goal of payoffs is to capture essential relations between alternative decisions and there consequences. We define three payoff tables. The first is applied for the

intermediate node asked to forward packet Fig. 4(a) and the second for the node being the source of the packet Fig. 4(b) Last table Fig. 4(c) defines a payoff related to the status of node's network interface.

Table 1. Intermediate node asked to forward packet

Reliability of the src node		Payoff	
trust	activity	Forward(F)	Discard (D)
	HI	14	0
3	ME	7	2
	LO	3	8
	HI	12	1
2	ME	6	4
	LO	2	10
	HI	10	2
1	ME	4	6
	LO	1	12
	HI	8	3
0	ME	2	7
	LO	0	14

Table 2. Node being the source of the packet

Transmission status	Payoff
Success (S)	15
Failure (F)	0

Table 3. For rounds in a sleep mode

Mode of the interface in a round	Pay off (per round)
Sleep	6
idle	0

For the source node the exact payoff depends only on the status of the transmission. If the packet reaches the destination then transmission status is denoted as S(success). Otherwise (packet discarded by one of the intermediate node's), the transmission status is denoted as F(failure) payoffs received by intermediate nodes depend on their decisions (packet discarded or forwarded) and on the reliability (trust and activity) of the source node. Generally, the higher the trust and activity levels are, the higher payoff is received by the node forwarding the packet. Higher reliably of the source node means that in the past this node already forwarded a decent number of packets for the currently forwarding node. So it is more likely that such node will be used in the future. This means that forwarding for such node might be considered

as an investment of trust for the future situations. When a node decides to discard a packet it is rewarded for saving its battery life. he payoff table for intermediate nodes reflects the use of the reputation based cooperation enforcement system by network participants. Without such system, the payoff for discarding packets would always be higher than for forwarding. Players also receive additional payoff for the time spent in sleep mode (reward for saving the battery). The fixed payoff is received for each round spent in a sleep mode. The fitness value of each player is calculated as follows:

$$\text{fitness} = \frac{\text{tps} + \text{tpf} + \text{tpd} + \text{tps}}{\text{ne}} \tag{1}$$

Where tps, tpf, tpd, tps are total payoffs received respectively for sending own packets, forwarding packets on behalf of others, discarding packets and staying in a sleep mode. The ne is a number of all events (number of own packets send, number of packets forwarded, number of packets discarded and number of rounds spent in a sleep mode).

Conclusion

The performance of MANETs depends on cooperation among network participants. Such networks are usually deployed in vulnerable environments without any authority controlling all devices. Due to the limited resources of the batteries, nodes might not want to share packet forwarding responsibilities such behavior might seriously threat the existence of the network. In this paper we have proposed a strategy driven approach to enforce cooperation among ad hoc network participants using the notion of trust and activity. It forces nodes both to reduce the amount of time spent in sleep mode and forwarded packet.

References

1. Sivaram Murthy, C., Manoj, B.S.: Adhoc Wireless Networks, 2nd edn. Pearson Education (2005)
2. Self-policing mobile ad-hoc networks by reputation systems. IEEE Communications Magazine, Special Topic on Advances in Self-Organizing Networks 43(7) (July 2005)
3. Namikawa, N., Ishibuchi, H.: Evolution of cooperative behavior in the iterated prisoner's dilemma under random pairing in game playing. In: Proc. IEEE Press Congress on Evolutionary Computation (CEC 2005), pp. 2637–2644 (2005)
4. Marti, S., Giuli, T., Lai, K., Baker, M.: Mitigating routing misbehavior in mobile ad hoc networks. In: Proc. ACM/IEEE 6th International Conference on Mobile Computing and Networking (MobiCom 2000), pp. 255–265 (2000)
5. Seredynski, M., Bourvy, P., Klopotek, M.A.: Preventing Selfish Behavior in Adhoc Networks. In: 2007 IEEE Congress on Evolutionary Computation (CEC 2007), pp. 3555–3559 (2007)

Multicriteria Decision Analysis for Intrusion Detection Data

Sanjiban Sekhar Roy, Omsai Jadhav, Saptarshi Chakraborty, Swapnil Saurav, and Madhu Viswanatham

School of Computing Science and Engineering, VIT University, Vellore
sanjibanroy09@gmail.com, omsaijadhav@gmail.com, sssaptarshii@gmail.com, swapnishu@rediffmail.com, vmadhuviswanatham@vit.ac.in

Abstract. Uncertain data can be dealt with rough set theory and dominance based rough set approach which is an extension to the classical rough set theory is a new mathematical technique to deal with multicriteria indecisive data. Here in this paper we have shown how dominance based rough set approach can be useful for analysis and evaluation of intrusion detection data set.

Keywords: Dominance based rough set, Intrusion Detection.

1 Introduction

Z Pawlak proposed rough set theory in 1982 which deal with vagueness and uncertainty[1]. Rough set theory reduces the needed number of attribute values to produce a more compact decision rule set and increases efficiency. This theoretical framework is based on the concept that every object in the universe is attached with some kind of information. It includes algorithms for generation of rules, classification and reduction of attributes. It is hugely used for knowledge discovery and reduction of knowledge. Let, T = (U, A) and let $B \subseteq A$ and $X \subseteq U$, then we can approximate X by using the information contained in B by building the lower and upper approximations of X, represented $\underline{B}X$ and $\bar{B}X$ respectively, where

$\underline{B}X = \{x \mid [x]_B \subseteq X\}$, $\bar{B}X = \{x \mid [x]_B \cap X \neq \phi\}$

The accuracy of the approximation is given by,

$\alpha_B(x) = \frac{card(\underline{B}(x))}{card(\bar{B}(x))}$. If $\alpha_B(X) = 1$, then X is a crisp set or if $\alpha_B(X) < 1$, then X is rough set.

In classical rough set theory the boundary region B of X is given by, $BN_B(X) = \bar{B}X - \underline{B}X$ consists of those objects that we cannot decisively classify in B. A set is called rough if its boundary region is non-empty, otherwise the set is crisp. If we assume, $c \in C.$, c is dispensable in T, if $POS_C(D) = POS_{(C-\{c\})}(D)$, otherwise

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 667–672.
springerlink.com

attribute c is indispensable in T. The C-positive region of D: $POS_C(D) = \bigcup_{X \in U/D} \underline{C}X$. It is true that the rough set theory proposed by Z. Pawlak is used to solve many decision tribulations but is not able to find solutions in the cases where data are with inclination-ordered attribute domains and decision classes. Therefore, there is a need of multi-criteria decision analysis (MCDA) of rough set loom. The classical rough set is not sufficient to solve attributes with preference-ordered domains of uncertain data. To overcome rough set limitations Greco *et al* [3][4]introduced noble approach which is able to deal with inconsistencies typical to exemplary decisions in MCDA problems namely dominance based rough set approach. Dominance based rough set approach is an extension of classical rough set theory. Here our paper discusses on a systematic framework for analyzing inspection data of intrusion detection models using dominance based rough set technique. The resulting activity patterns of intrusion detection are then utilized to guide the selection of system features and used for construction of additional time-based statistical features for future learning. Classes based on these selected attributes are then computed (inductively learned) using the appropriate, formatted audit data. Here we have shown that classes can be introduced using dominance relation among conditional attributes used in an intrusion detection models since they can decide whether an observed system activity is "authentic" or "disturbing".

2 Rough Approximation by Dominance Relations

All conditional attributes are actually criteria's in multi-criteria classification, which includes order of preference among its domain[2]. In case of dominance based rough set approach outranking relation plays an important role. An outranking relation $\geq_q$ on U symbolize a fondness on the set of objects with respect to criterion q i.e. the logical meaning of $x \geq_q y$ is "x is at least good in comparison with y on the basis of criteria q". The same statement can be said in different way as x dominates y with the criteria q i.e. here $P \subseteq C$ and criteria q, $\forall q \in P$ which sometimes defined as $xD_p y$.

In multi-criteria decision analysis there exists a preference order in the set of classes Cl .The approximation happens for upward and downward classes only. Let, us also consider the following upward and downward unions of classes, respectively,

$Cl_t^{\geq} = \bigcup_{s \geq t} Cl_s ; Cl_t^{\leq} = \bigcup_{s \leq t} Cl_s$

$Cl = \{Cl_t, t \in T\}, 1 \leq t \leq n$ be a set of classes of U,

In dominance based rough set approach, a collection of entities dominating x, named as P dominating set can be given as $D_p^+(x) = \{y \in U : yD_p x\}$ and exactly the opposite ,a collection of entities x, dominated by a set named as P Dominated Set is referred as $D_p^-(x) = \{y \in U : xD_p y\}$ provided $P \subseteq C$ and $x \in U$.

Therefore, approximation values of P-lower and P-upper of $Cl_t^{\geq}$;where $t \in T$ with respect to $P \subseteq C$ given as $\underline{P}(Cl_t^{\geq})$ and $\overline{P}(Cl_t^{\geq})$ correspondingly, which are as follows

$\underline{P}(Cl_t^{\geq}) = \{x \in U : D_p^+(x) \subseteq Cl_t^{\geq}\}$,

$\overline{P}(Cl_t^{\geq}) = \bigcup_{x \in Cl_t^{\geq}} D_P^+(x) = \{x \in U : D_p^-(x) \cap Cl_t^{\geq} \neq \phi\}$

Similarly, P-lower and P-upper approximations of $Cl_t^{\leq}, t \in T$, where, $P \subseteq C$ is referred as $\underline{P}(Cl_t^{\leq})$ and $\overline{P}(Cl_t^{\leq})$ correspondingly, are given as :

$\underline{P}(Cl_t^{\leq}) = \{x \in U : D_p^-(x) \subseteq Cl_t^{\leq}\}$,

$\overline{P}(Cl_t^{\leq}) = \bigcup_{x \in Cl_t^{\leq}} D_P^-(x) = \{x \in U : D_p^+(x) \cap Cl_t^{\leq} \neq \phi\}$

$\underline{P}(Cl_t^{\geq}) \subseteq Cl_t^{\geq} \subseteq \overline{P}(Cl_t^{\geq})$;

$\underline{P}(Cl_t^{\leq}) \subseteq Cl_t^{\leq} \subseteq \overline{P}(Cl_t^{\leq})$;

these properties hold true along with its complimentary properties.

$\underline{P}(Cl_t^{\geq}) = U - \overline{P}(Cl_{t-1}^{\leq})$, t=2,…,n

$\underline{P}(Cl_t^{\leq}) = U - \overline{P}(Cl_{t+1}^{\geq})$, t=1,…,n-1

$\overline{P}(Cl_t^{\geq}) = U - \underline{P}(Cl_{t-1}^{\leq})$, t=2,…,n

$\overline{P}(Cl_t^{\leq}) = U - \underline{P}(Cl_{t+1}^{\geq})$, t=1,…,n-1

Therefore the P-doubtful regions of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ are defined as:

$Bn_P(Cl_t^{\geq}) = \overline{P}(Cl_t^{\geq}) - \underline{P}(Cl_t^{\geq})$, $Bn_P(Cl_t^{\leq}) = \overline{P}(Cl_t^{\leq}) - \underline{P}(Cl_t^{\leq})$,

The correctness of approximation of $Cl_t^{\geq}$ and $Cl_t^{\leq}$ for all $t \in T$ and for any $P \subseteq C$, is defined as

$$\alpha_P(Cl_t^{\geq}) = \left|\frac{\underline{P}(Cl_t^{\geq})}{\overline{P}(Cl_t^{\geq})}\right|, \alpha_P(Cl_t^{\leq}) = \left|\frac{\underline{P}(Cl_t^{\leq})}{\overline{P}(Cl_t^{\leq})}\right| \text{ and the ratio}$$

$$\gamma_P(Cl) = \left|\frac{U - ((\bigcup_{t \in T} Bn_p(Cl_t^{\geq})) \bigcup (\bigcup_{t \in T} Bn_P(Cl_t^{\leq})))}{U}\right|$$

is known as the quality of approximation of the partition Cl by the set of criteria P or briefly quality of sorting. Therefore, γ_P ratio is the relation among the P-correctly classified substance and the objects in the table. The definition of reduct of C with respect to class Cl is each minimal subset $P \subseteq C$ such that $\gamma_P(Cl) = \gamma_C(Cl)$ and is avowed by $RED_{Cl}(P)$. Therefore, a data table can have many reducts. $CORE_{Cl}$ is the intersection of their reducts.

3 Experimental Result Using Dominance Based Rough Set

We have used dominance based rough set approach for the following data taken from paper [2]. We can name this data table as “connection records of a network”. Classes based on these selected attributes are then computed (inductively learned) using the appropriate, formatted audit data. Here, we have shown that classes can be introduced using dominance relation among conditional attributes used in an intrusion detection models since they can decide whether an observed system activity is “authentic” or

"disturbing". Here the attack model we have shown includes short sequence of connection of records of intrusions evidence. To see which ports are easy to get to invader analytically makes links to each port (service) of a intention host (target host). In the connection records, there should be a host (or hosts) which receive many connections to its "different" ports in a short period of time. There can be links of "REJ" flag as numerous ports are by and large not available as nearby are several patterns to facilitate the proposal of the attack, e.g (destination host = 207.217.205.23, flag = REJ). Therefore the destination host and FLAG value constitute an attack. We have shown the minimum set of attributes which summaries the following data table for intrusion detection.

Table 1. Connection records of a network

Sl. No	Clock	A_1	A_2	A_3	A_4	A_5
1	1.0	30	telnet	150	500	REJ
2	1.6	25	http	300	2500	SF
3	2.4	5	Smtp	200	2500	SF
4	3.0	25	telnet	200	3000	SF
5	3.5	30	telnet	300	2000	SF
6	4.0	30	http	150	1000	REJ
7	4.2	5	http	150	1000	REJ
8	4.5	30	Smtp	300	1000	REJ
9	4.9	25	Smtp	150	3000	SF
10	5.0	5	Smtp	150	500	REJ
11	5.2	5	Smtp	200	2500	REJ
12	5.5	25	telnet	300	3500	REJ

Here, set Q and P contains the following attributes.

$Q=\{A_1,A_2,A_3,A_4,A_5\}$

$P=\{A_1,A_2,A_3,A_4\}$

Attribute A_1 to A_4 called as conditional attributes and attribute A_5 is decision attribute. Here, the second column refers to the arrival time known as timestamp of the packet in the data table, therefore attribute A_1 contains the duration of each raw packet. Thereafter, attributes A_3, A_4 contains services (e.g. http,smtp,telnet),source byte and destination bytes of the raw packets. According to value of all the conditional attributes data packet is rejected (REJ) or successfully accepted (SF) by the network. Now using dominance based rough set approach we will approximate the class $Cl_1^{\leq}$ of "atmost REJ" and the class $Cl_2^{\geq}$ of "atleast SF". As we know $P \subseteq C$, therefore we have taken all the attribute combinations to find the γ_P values [3,4] of all the subsets.

3.1 Calculations

1) C= {A$_1$, A$_2$ }

$\underline{C}(Cl_1^{\leq}) = \{7\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{1,2,3,4,5,6,8,9,10,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{1,2,3,4,5,6,8,9,10,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{1,2,3,4,5,6,8,9,10,11, 12\}$

$\gamma_P(Cl) = 1/12$

2) C= {A$_2$,A$_3$}

$\underline{C}(Cl_1^{\leq}) = \{6,7\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{1,2,3,4,5,8,9,10,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{1,2,3,4,5,8,9,10,11,12\}$

$\gamma_P(Cl) = 1/6$

3) C= {A$_1$,A$_3$}

$\underline{C}(Cl_1^{\leq}) = \{7,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{1,2,3,4,5,6,8,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{1,2,3,4,5,6,7,8,9,10,11, 12\}$

$Bn_C(Cl_2^{\geq}) = \{1,2,3,4,5,6,8,9,11,12\}$

$\gamma_P(Cl) = 1/6$

4) C= {A$_1$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \{5\}$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,9,11,12\}$

$\gamma_P(Cl) = 1/2$

5) C= {A$_2$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,5,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$\gamma_P(Cl) = 5/12$

6) C= {A$_3$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11, 12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,5,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$\gamma_P(Cl) = 5/12$

7) C= {A $_1$,A$_2$,A$_3$}

$\underline{C}(Cl_1^{\leq}) = \{6,7,10\}$

$\overline{C}(Cl_1^{\leq}) = 1,2,3,4,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{1,2,3,4,8,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \{5\}$

$\overline{C}(Cl_2^{\geq}) = \{1,2,3,4,5,8,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{1,2,3,4,8,9,11,12\}$

$\gamma_P(Cl) = 1/3$

8) C= {A $_2$,A$_3$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,5,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$\gamma_P(Cl) = 5/12$

9) C= {A $_1$,A$_3$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,5,6,7,8,9,10,11, 12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,5,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \phi$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$\gamma_P(Cl) = 5/12$

10) C={A $_1$,A$_2$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \{5\}$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,9,11,12\}$

$\gamma_P(Cl) = 1/2$

11) C= {A $_1$,A$_2$,A$_3$,A$_4$}

$\underline{C}(Cl_1^{\leq}) = \{1,6,7,8,10\}$

$\overline{C}(Cl_1^{\leq}) = \{1,2,3,4,6,7,8,9,10,11,12\}$

$Bn_C(Cl_1^{\leq}) = \{2,3,4,9,11,12\}$

$\underline{C}(Cl_2^{\geq}) = \{5\}$

$\overline{C}(Cl_2^{\geq}) = \{2,3,4,5,9,11,12\}$

$Bn_C(Cl_2^{\geq}) = \{2,3,4,9,11,12\}$

$\gamma_P(Cl) = 1/2$

Hence, the attribute sets **{A$_1$,A$_4$}** and **{A $_1$,A$_2$,A$_3$}** of **{A $_1$,A$_2$,A$_3$,A$_4$}** . So intersection of reduct hence CORE is attribute **A$_1$.**

4 Conclusion

This paper shows that dominance based rough set approach can be a useful mathematical tool while dealing with uncertain intrusion data set. Here, in this paper we have shown how dominance based rough set approach can be useful for analysis and evaluation of intrusion detection data set and we have shown that classes can be introduced using dominance relation among conditional attributes used in an intrusion detection models. Finally, we have shown the CORE and accuracy of the data table using dominance based rough set approach.

References

[1] Pawlak, Z.: Rough Sets. International Journal of Computer and Information Sciences 11, 341–356 (1982)

[2] Lee, W., Stolfo Salvatore, J., Kui, W.: A Data Mining Framework for Adaptive Intrusion Detection. In: Proceedings of the IEEE Symposium on Security and Privacy (1999)

[3] Greco, S., Matarazzo, B., Slowinski, R.: Rough sets theory for multicriteria decision analysis. European Journal of Operational Research 129, 1–47 (2001)

[4] Greco, S., Matarazzo, B., Slowinski, R.: Rough approximation of a preference relation by dominance relations. European Journal of Operational Research 117, 63–83 (1999)

[5] Kusunoki, Y., Inuiguchi, E.M.: A unified approach to reducts in dominance-based rough set. Approach Soft Computing 14, 507–515 (2010)

Realization of the Cryptographic Processes in Privacy Preserving

Sumana M.[1] and Hareesh K.S.[2]

[1] Information Science and Engineering Department,
M S Ramaiah Institute of Technology, Bangalore, India
sumana.a.a@gmail.com
[2] Computer Science and Engineering,
Manipal Institute of Technology Manipal, India
harish_dvg@yahoo.com

Abstract. Several data mining processes include privacy preservation in order to avoid disclosure of sensitive information while discovering knowledge. Data mining algorithms uses numerous modification techniques to construct models or patterns from private data. It is essential to evaluate the quality of the data resulting from the alteration applied by each algorithm, as well as the performance of the algorithms. Hence it is required to identify a broad set of decisive factors with respect to which to assess the existing algorithms and determine which algorithm meets specific requirements. This paper discusses the working of the cryptographical techniques and its usage in privacy preserving data mining. The performances of these procedures are analyzed indicating their level of privacy maintained.

Keywords: distributed computing, privacy, data mining, cryptography, performance evalution.

1 Introduction

Privacy preserving data mining (PPDM) algorithms aim in extraction of relevant knowledge from huge volumes of data, while protecting sensitive information at the same time. The existing privacy preserving data mining techniques can be classified according data distribution (centralized or distributed), the modification applied to the data (encryption, perturbation, generalization, and so on) , the data mining algorithm(classification, association mining or clustering) which the privacy preservation technique is designed for, the data type (single data items or complex data correlations) that needs to be protected from disclosure and the approach adopted for preserving privacy (heuristic or cryptography-based approaches) as in [1]. Heuristic-based algorithms proposed aim at hiding sensitive raw data by applying perturbation techniques based on probability distributions and is mainly conceived for centralized data sets. Current trends in several heuristic-based approaches for hiding both raw and aggregated are k-anonymization, adding noises, data swapping, generalization and sampling.

As mentioned in [2] due to the varying typical features of the privacy preserving algorithms there is no one single approach that outperforms the others. An algorithm

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 673–678.
springerlink.com

may perform better than the other with respect to hiding sensitive information, privacy level and data quality as seen in [1] and [5]. Users are provided with a comprehensive set of privacy preserving related metrics which will enable them to select the most appropriate privacy preserving technique for the data at hand; with respect to some specific parameters they are interested in optimizing.

2 Privacy

The PPDM algorithms are constructed based on some of the following privacy definitions. As seen in [1] First, privacy can be defined as the right of an individual which means that the person involved can indicate what data and how much of data can be revealed. Secord, privacy as a controlled access to information and lastly, privacy as limited to the person and all features related to the person, which means there is no compromise to individual's privacy.

3 Criteria for Evaluation

A. Level of privacy -Here the level up to which sensitive information that has been hidden can be evaluated. The aspects under which metrics for privacy level needs to be used are data privacy and results privacy. Data privacy is quantified by degree of uncertainty. Higher the degree of uncertainty betters the data privacy. Results privacy should include metrics to indicate the degree of inferring the private/sensitive data from the results obtained after data mining.
B. Failure in hiding sensitive information-The *hiding failure* parameter is estimated by the percentage of sensitive information that is still discovered, after the data has been hidden. Most of the developed privacy preserving algorithms are designed with the goal of obtaining zero hiding failure. Oliveira and Zaiane define the *hiding failure* (HF) as the percentage of restrictive patterns that are discovered from the sanitized database. It is measured as follows: HF = #RP(D′)/ #RP(D), where #*RP*(*D*) and #*RP*(*D′*) denote the number of restrictive patterns discovered from the original data base *D* and the sanitized database *D′* respectively as has been seen in[2]. Under an ideal situation the HF should be 0.
C. Data Quality -Data metrics should be included to evaluate the state of the individual items contained in the database after the enforcement of a privacy preserving technique such as perturbation or anonymizing data in a database. The quality of the data mining results evaluates the alteration in the information that is extracted from the perturbed or anonymized data on the basis of intended data use. Quality of the result extracted includes various parameters such as accuracy, completeness and consistency.
D. Complexity -This evaluation criterion comprises metrics to measure the efficiency and scalability of a PPDM algorithm. Efficiency measures the space and time requirements of the algorithm to indicate its performance. Evaluation of the time requirements is the average number of operations performed in order to reduce the frequency of appearance of sensitive information. The communication cost incurred during the exchange of information among the collaborating sites should also be

evaluated. Scalability describes the efficiency trends of the algorithm when size of the data sets enhances.
E. Resistance -The level of sanitization for different data mining techniques differs. When a sanitization algorithm is developed on a dataset for a particular data mining technique the endurance of this algorithm so that no other data mining technique can disclose the sensitive information has to be measured. Such a parameter is called as endurance traversal as in[7].

4 Commonly Used Cryptographic Approaches

The commonly used cryptographic private approaches used in Privacy Preserving Data mining are secure multiparty protocols, mentioned in [3] and [5]. Most of these protocols are used based on the data mining task to be performed on the data sets. These protocols are either homomorphic or commutative in nature.

Let $Epk(.)$ denote the encryption function with public key pk and $Dpr(.)$ denote the decryption function with private key pr. A secure public key cryptosystem is called homomorphic if it satisfies the following requirements:

(1) Given the encryption of $m1$ and $m2$, $Epk(m1)$ and $Epk(m2)$, there exists an efficient algorithm to compute the public key encryption of$m1+m2$, denoted $Epk(m1+m2) := Epk(m1) +h Epk(m2)$.
(2) Given a constant k and the encryption of $m1$, $Epk(m1)$, there exists an efficient algorithm to compute the public key encryption of $k \cdot m1$, denoted $Epk(k \cdot m1) := k \times h Epk(m1)$.
An encryption algorithm is commutative if the order of encryption does not matter. Thus, for any two encryption keys E1 and E2, and any message m, E1 (E2 (m)) = E2 (E1 (m)).
A. Secure Sum - Secure Sum securely calculates the sum of values from individual sites[6]. Homomorphic encryption could be used to calculate secure sum.
B. Secure Dot Product- Securely computing the dot product of two vectors [4] is another important sub protocol required in many privacy-preserving data mining tasks..The key idea behind the protocol is to use a homomorphic encryption system. However this protocol works well only for 2-party situation.
C. Secure Union and Secure Intersection -Secure Union [6,8] and Secure Intersection compute the secure union and intersection of vectors. These protocols can be used for multiple parties. They can be either homomorphic or commutative in nature.

5 Realizations

The following protocols have been realized based on their role in privacy preserving data mining algorithm. The complexity of each one of them has also been discussed.

A. Secure Sum Computation
In privacy preserving decision tree classification one of the terminating conditions is to check whether there are attributes to split further while selecting the best attribute. This method has been used for multiple parties holding vertically partitioned data,

which means, there are k parties, P1,…., Pk. There are a total of n transactions for which information is collected. Party Pi collects information about mi attributes, such that $m = \Sigma^{i=1}_{k} m_i$, where m_i is the total number of attributes/features. In this situation only one of the parties will have the class attribute.

This program was realized in java.

1. Party1 uses a Generate Random function to obtain a random number (*rand*). It then adds the number of attributes it has to this random number and sends it to the neighboring party.
2. For i=2 to k

Each party i adds its number of attributes to the number obtained from its party (i-1) and forwards it to its party (i+1).

Note: But if a party has the class attribute number of attributes it adds to the value excludes the class attribute.
3. Party1 receives the added result from party k. If the value obtained is equal to *rand* then it indicates that no attributes are left for split. Else there exists attributes for split.
Based on the above for the secure sum protocol with k parties, the computation and communication costs are both O (k log (|T|)). |T| indicates the maximum size of the messages passed between the parties and the outputs of the parties.

Secure sum maintains aggregate of the individual counts and present only the aggregate counts. Though this approach guarantees confidentiality of individual count, it still exposes the accurate sum to the miner, which compromises the individual record privacy.

B. Dot Product Computation

While performing privacy preserving association mining, we need to find the support count of the item sets being maintained by multiple sites. Assumptions done in this approach is that the vector holding information about the transactions having those items will have a value 1 else 0. Given is the itemset for whose support count has to be computed. If the items in the itemset belong to the same site then compute the product of the vectors at the site itself. Else if they do not belong to the same site then obtain the vectors perform homomorphic encryption on it to result in a vector of 0's and 1's and perform the dot product.

C. Realization of Secure Union

Here we discuss the working of the secure union protocol used to compute the union of the data available on multiple sites. RSA protocol has been used to perform commutative encryption required for secure union. The working of the secure union protocol built is as follows.

It is required that the parties will have the items/data whose union is to be computed. To use RSA for commutative approach the modulus value n is common to all the sites. All sites based on this value compute the encryption key.

1. Each of the sites encrypts its data.
2. Each of the sites sends its encrypted results to the other sites.
3. All sites encrypt the encrypted data received from other sites.
4. In the above mentioned step if there are 4 sites then the data present at each site is encrypted 4 times by all 4 sites. Since commutative encryption is used

E1(E2(E3(E4(E)))) = E2(E3(E1(E4(E)))) = E3(E1(E2(E4(E)))) = E4(E2(E1(E3(E))))

5. Let all the encrypted results be sent and maintained in party 1. Party 1 sorts all the encrypted data.
6. Site1 then eliminates all duplicate elements in the sorted set.
7. The final set obtained after sorting will indicate the union of all the elements.

Similar procedure can be used to attain the secure intersection of the elements. Here in the sorted array if the encrypted elements occurs the number of times as that as the number of sites then element is common to all the sites.

Security of the data is maintained by using this approach. Data is also hidden while mining. However the complexity depends on the type of commutative encryption, and the amount of data to be encrypted. Communication cost depends on the number of items encrypted and to be forwarded to the other sites and the key size used for encrypting data.

6 Conclusions

This paper converses a framework for evaluating privacy preserving data mining algorithms. Such framework allows one to assess the different features of a privacy preserving algorithm according to a variety of evaluation criteria. Parameters like level of privacy, data quality and hiding failure have been defined and the evaluations of such parameters over a set of privacy preserving algorithms have been presented. Communication and security analysis of the algorithms such as secure union, sum and dot product have been performed. Further these above mentioned approaches have been implemented keeping in mind the usage of these methods in privacy preserving. Our implementations work well for semi honest models.
In our future work we would like to evaluate the algorithms for malicious models. Security and quality issues also need to be considered while using an efficient encryption method such as Pohlig Hellman. Further we would build privacy preserving techniques using the secure multiparty protocols realized.

Acknowledgments. We thank M S Ramaiah Institute of Technology for their constant support in submitting and participating in International Conferences.

References

1. Bertino, E., Lin, D., Jiang, W.: A Survey of Quantification of Privacy Preserving Data Mining Algorithms. Scientific Commons (2008)
2. Aggarwal, C.C., Yu, P.S.: Privacy Preserving Data Mining: Models and Algorithms. Springer (2007)
3. Lindell, Y., Pinkas, B.: Secure Multiparty Computation for Privacy-Preserving Data Mining. Israel Science Foundation (grant number 860/06) (2008)
4. Hussein, M., El-Sisi, A., Ismail, N.: Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Data Base, pp. 607–616. Springer, Heidelberg (2008)
5. Clifton, C., Kantarcioglu, M., Vaidya, J.: Tools for Privacy Preserving Distributed Data Mining. SIGKDD Explorations 4(2) (2004)

6. Shrikant Vaidya, J.: Privacy Preserving Data Mining Over Vertically Partitioned Data. A Thesis Submitted to Purdue University for the Degree of Doctor of Philosophy (August 2004)
7. Kenthapadi, K.: Models and Algorithms for Data Privacy. A Dissertation Submitted To The Department Of Computer Science for the Degree of Doctor of Philosophy (September 2006)
8. Du, W., Zhan, Z.: Building decision tree classifier on private data. In: IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining, Maebashi City, Japan, December 9, vol. 14, pp. 1–8. Australian Computer Society (2002)

A Privacy Preserved Integrated Framework for Location Based Tracking for Wireless Sensor Networks

Vinoth Kumar S., Suresh R.M., and Govardhan A.

Dhanalakshmi College of Engineering
mail.vinoth72@gmail.com, rmsuresh@hotmail.com,
govardhan_cse@yahoo.co.in

Abstract. One of the most notable challenges threatening the successful deployment of sensor systems is privacy. When the adversary needs to know the location of the particular node, it can easily do that with the help of the signal that are emitted by the mobile sensor node. Many applications today use the mobile sensor nodes for transferring their data which may be sensitive. In such a scenario, an adversary can easily hack the data by finding the location of the mobile sensor node and replicating the mobile sensor node to depict itself as a destination. This is one of the major disadvantages that the existing system faces. Hence we propose a system model that blurs out its location information to the server itself before it sends its location information to other nodes. Further we examine the performance matrices of the existing system. Hence we have modified algorithm to provide efficient energy consumption. As a result of experiment, we observed that our proposed technique protects the location privacy and can sufficiently reduce computation cost so that the computation technique can be practically applied to location-based services.

Keywords: anonymity, cloaked area, privacy, and mobile sensor node.

1 Introduction

Wireless sensor nodes (WSNs) consist of small nodes with sensing, computation and wireless communications capabilities. These sensor networks communicate by interconnecting with several other nodes when established in large and this opens up several technical challenges and immense application possibilities. The advancement in WSN‘s has resulted in many new application for military and civilian purposes. Most of these applications rely on the information of personal locations. For example, Surveillance and Location system [1]. In location based services, users with location-aware mobile devices are able to make queries about their surrounding anywhere and at anytime. While this ubiquitous computing paradigm brings great convenience for information access, it also raises concerns over potential intrusion into location privacy [2]. The general tracking mechanisms makes use of the various sensors namely identity sensors, counting sensors etc. While the identity sensors reveal the exact location of the users, the counting sensors are used for counting the number of users in the specific area. These are the photoelectric sensors or thermal sensors [3], [4], [5]. Unfortunately, these sensors help in monitoring the exact location of the user. This may lead to privacy breach for the user. An adversary may easily infer the

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 679–690.
springerlink.com

location information of the user and can thus get access to the personal information of the user, for example, by knowing the various medical centers visited by a person, the health information of that person can be predicted. Similarly, other such information of vital importance may be revealed. Also, by capturing the location information of the user, the adversary can generate many replicas from a single captured node. In this attacker model, known as replica node attack, the adversary takes the secret keying materials from a compromised node, generate large number of attacker controlled replicas that share the compromised node's keying materials and id, and then spreads these replicas throughout the network. With a single captured node, the adversary can create as many replica nodes as he has the hardware to generate [6]. The other form of privacy breach that can result from location based tracking services is the revealing of the location information of the attacker himself.

In this paper, we propose the algorithm to preserve privacy in location based tracking. The Section 2 talks about related works that are being carried out in this area. Section 3 is about the anonymity principle. Section 4 describes the proposed methodology. Section 5 is about the system implementation. Finally, Section 6 talks about the performance evaluation.

2 Related Works

Privacy concerns are location based application scenarios are typically addressed in a location broke residing in the middleware layer to our knowledge Spreitzer and Theimer pioneered the development [10]. In this work, each user owns a trusted user agent that acts as an intermediary. It collects location information from a variety of sensors and controls application access to this data.

Here, we review the security issues that are concerned with these wireless sensor networks. Fox and Gribble [11], provide a security protocol that provides secure access to application-level proxy services. Their protocol is designed to interact with a proxy to Kerberos and to facilities porting services that rely on Kerberos to wireless devices.

Privacy issues surfaces in many situations:

- Information regarding individuals that is communicated between two parties.
- Information that is destined for the user.
- Information that the user may emanate both implicitly and explicitly (e.g. signals from the cellular phone)

In this paper, the last point is focused as concerned with anonymizer; the first anonymous system, the MixNet system design was proposed by Chaum [12]. The system hides the correspondence between senders and receivers of message by repeatedly encrypting messages with the public keys of a predefined set of mixes. A mix is a server that decrypts received encrypted messages with its corresponding private key and records them before forwarding them to next mix.

3 Anonymity Principles

The main idea of the paper is to propose a system that does not allow the adversary to predict the exact location of the user. This is done by revealing the aggregate location of the user instead of the exact location. The aggregate location of the information by some other sensor nodes that lies closer to the user. This information is sent to the server. The server then forwards this information to the user who locates the user. This conversion of the exact location information into average location information is known as cloaking. The aggregate area is known as cloaked area. The size of the cloaked area depends upon the level of anonymity required by the user. If higher level anonymity is desired, then the quality of getting the preserving privacy is high. On the other hand, if the anonymity level desired is low, then the quality of aggregate information is low (i.e.) the exact location of the user can be obtained with a few observations. Thus, the algorithm must satisfy a K-anonymity principle, which states that the number of the persons in the cloaked area must be greater than or equal to K (n>=k). This paper aims at developing an algorithm to preserve privacy.

4 Proposed Methodology

In the existing work [1] the location monitored by the sensor nodes gives the exact location when the adversary issues a tracking order. This may lead to privacy breaches in case of extremely sensitive matters. The proposed work aims to preserve this privacy while monitoring the location. This is implemented with the help of the system architecture shown below,

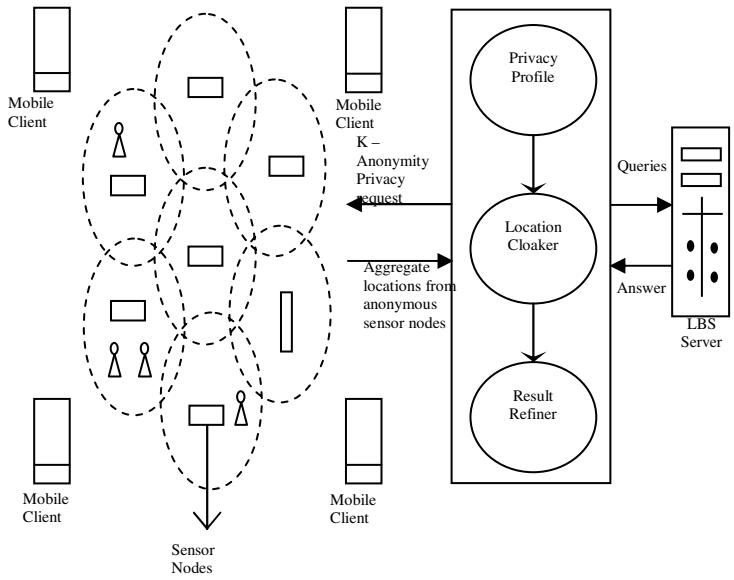


Fig. 1. System architecture

In figure 1 the system architecture that is used in our proposed system is depicted. It consists of the following components:

Sensor nodes: Each sensor nodes determines the number of objects in its sensing area, blurring its sensing area into cloaked area A, which includes at least K objects, and reporting A with the number of objects located in A as aggregate location information to the server.

Server: The server is responsible for collecting the aggregate location reported from the sensor node. The cloaked area that can be used as a rectangle areas because it is a polygon that is widely adapted by existing query processing algorithm.

Here the server is also maintained under the privacy layer. This means that even the server is not able to get the exact location until the node allows it.

1. Mobile user module: This is a web page that can be used by any new user. The new user has to register him with the web site. This information is sent to the database. Any user who has already registered himself could directly login and use the web site. The figure 2 shows the user profile wherein a new user can register, old user can login, set the mode, update the friend list etc.

Fig. 2. User profile

2. **Administrator module:** This is also a web page wherein the user (already registered) can change the location of any other user. The Figure 3 shows the administrator's profile. This special privilege of the administrator is exercised with the help of the location updates tab provided. The implementation is done using a pictorial representation wherein the location is displayed.

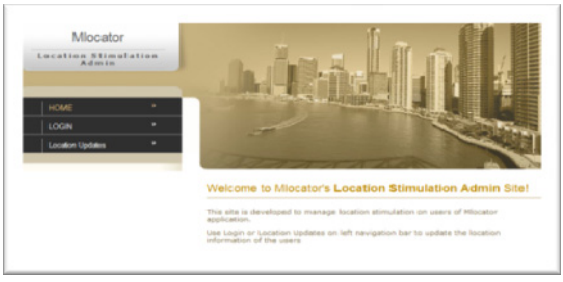
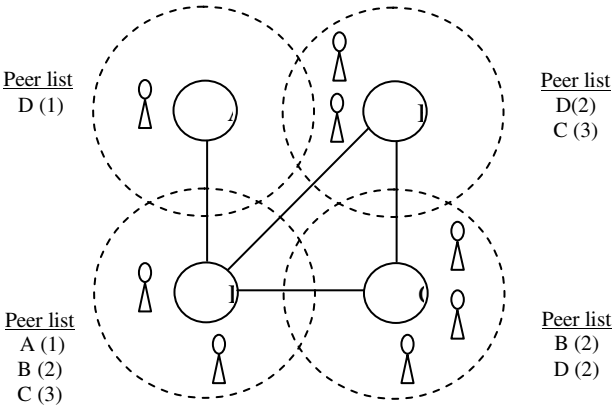


Fig. 3. Administrator's profile

3. **Server module:** Here, we implement the algorithm. Also the server is an interface between the users. Whenever a user is tracked, then the server will check the mode (selective, public or private). If public mode is selected then the server will respond by sending the address. If the private mode is selected then the server responds by a denial to track the user. If the mode is selective then the server will send a notification to the user that he is being tracked and upon reception of approval, the location of the user is sent to the other user. This is the alert message and this is how we preserve privacy. The server by using the algorithm will convert the location information of the user into an averaged area it is technically referred. This cloaked area is also calculated and stored into the database. The server used is apache tomcat server.

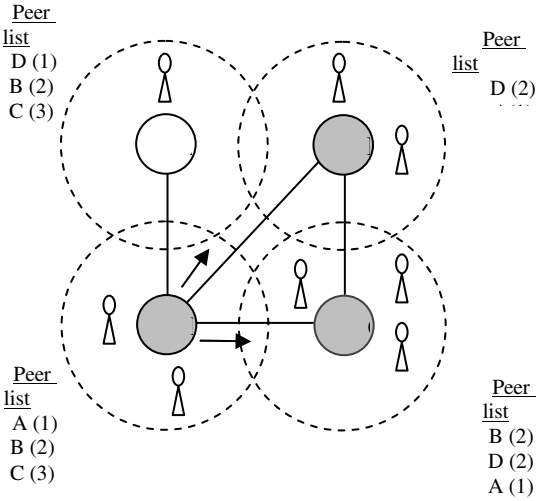
4.1 PP's Tracking Algorithm

Working of the system: Whenever an option to track is given, the anonymity level desired should also be specified. According to the desired anonymity level the peerlist is computed and the corresponding area is reported.

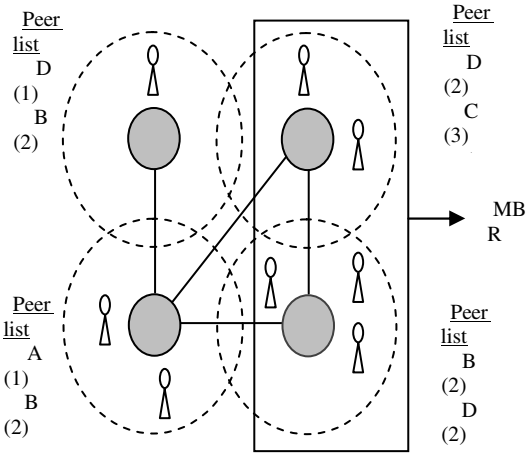


(a) Peer list after the first broadcast

Fig. 4. a) Broadcast by the nodes
b) Rebroadcasting by the node which has not obtained its peerlist
c) Computation of cloaked area



(b) Rebroadcast from sensor node D



(C) Cloaked area of sensor node C

Fig. 4. (*continued*)

PP's Tracking algorithm

```
1.  // Class Declaration
2.  Declare the class DBTest
3.  Declare all the variables
4.  Return the userid
5.  // end of Class Declaration
6.  // Function Definition
7.  Define getUserById ( Int d)
8.  Initialize
9.  Connection con=null;
10. Statement stmt=null;
11. ResultSet rs=null;
12. Define getUserByLogin( String email, String
    pwd)
13. Initialize
14. Connection con=null;
15. Statement stmt=null;
16. ResultSet rs=null;
17. Users oUsers=null;
18. while(rs!=null && rs.next()) {
19. oUsers= new Users();
20. print the Id,name;
21. Define Users updateUserMode( Users oUser,
    String modeStr) {
22. Initialize the variables for connection;
23. Write the query stmt=(Statement)
    con.CreateStatement();
24. }
25. Define public Users UpdateUserFriends(Users
    ousers,  String FriendList)
26. {
27. Class.forName("com.mysql.jdbc.Driver");
28. con =
    DriverManager.getConnection(urlStr,userid,
    password);
29. }
30. public Users insertUser (Users ouser)
31. {
32. stmt = (Statement) con.createStatement();
33. String sqlStr="insert into users (FNAME,
    LNAME,
    EMAIL,PWD,PHONE,DOB,GENDER,ISTRA
    CKABLE,CURRENTLOCATION,FRIENDLI
    STID) VALUES (" + " ' " + ouser.getFname()
    +"' , '"
    +ouser.getLname()+"','"+ouser.getEmail()+"','"
    +ouser.getPassword()+"','"+ouser.getPhoneno()
    +"','"+ouser.getDob()+"','"+ouser.getGender()+
    "','"+ouser.getIstrackable()+"','"+ouser.getCurr
    entloc()+"','"+ouser.getUseridlist()+"');
34. }
35. // End of function Definition
36. //End of Algorithm
```

The idea behind developing the PP's tracking method is the resource aware algorithm [1]. The steps of this algorithm are depicted in figure 4. Our framework consists of,

1) Code for the build file: There are separate build files for the users' page and the administrator page. The build file is written in xml script using the eclipse IDE. When the build file is compiled using the ant build, then the war file is created in the tomcat\Webapps directory.

2) A*dministrator's page's build file*: Compiled using ant build. The algorithm is used to create the users page. Connection, Result set etc. used in the user's page has to be specified in this code as private fields. This is done in order to preserve privacy. For each of the private string specified, we also need to define a corresponding method to do the required actions. These methods however can be made of public visibility.

3) *Attributes the creation of the administrator's page*: There is a major advantage of creating this page. This is because; only the administrator has the privilege to change the location of any of the user. Thus, we can simulate our output to suit the mobile requirements with the help of this administrator's page.

4*) Connectivity step*: Various tools are available for the creation of the database. One such tool that has been used by us is DbVisualizer. We need to connect this tool with the database so that any modifications in data are reflected in the database. We make use of the jdbc.odbc connectivity for establishing connectivity in our database.

5) *Test for connectivity*: Ensures that the connectivity is established correctly.

With all the above steps, we can proceed to get the desired output.

5 System Implementation

We evaluate the system performance with the running time that the build file takes to execute. The running time of the system varies with different integrated development used for compiling and running the system. For instance, the time taken for the MySQL server to start is 1 second. The administrator's build file when build using ant build takes 8 seconds to execute in the eclipse IDE. On the other hand, the user's build file when build using ant build takes 3 seconds. The time for compiling or building these files varies greatly depending upon the order of execution of these files and the amount of disk space occupied. The tomcat server takes 3029 milliseconds to start. Finally, we check the output with the UI depending upon the desired result.

The architecture in the already proposed system consists of the server as centralized system. The existing system is modified in such a way that the server itself is separated from the end user. This ensures that the server itself is hidden from the actual location of the user. We hence present the location to the server as desired by the user. Here is where the mode comes into use. The location presented to the server is based upon the mode. The database for the user is as follows.

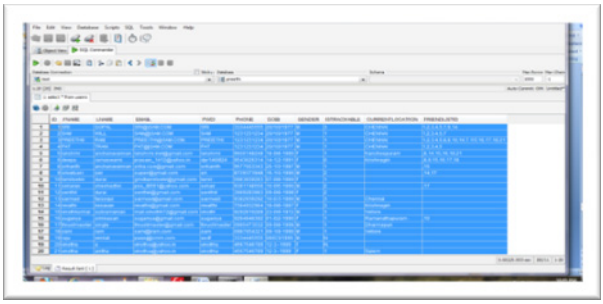


Fig. 5. Back end for list of users

In order to show the results we have distinguished the two users as a common user and an administrator. The common user only has the privilege to register.

The options for mode setting for the user are implemented as follows:

Fig. 6. Mode setting

The users are listed with the help of the list users option provided by the common users login form. The users are listed are as follows:

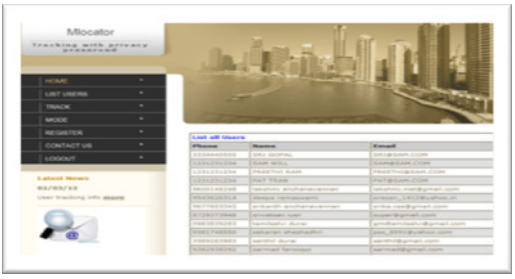


Fig. 7. List users

Finally the tracking option is implemented in collaboration with the mode option. In the mode page there is an additional option which gives the user a privilege to manage friends. This is another pre- requisite to make our system function properly.

For instance let us consider the case wherein the mode selected by the user is allowing friends. In such a case the user has to also select the friendlist. We make use of a foreign key reference here, the primary key being the id of the user. The id of the person who chooses his friends is marked in the friend list field of the users table maintained in the database. In the case considered, when the track option is selected, only the users who have logged in as the friend of the person will be able to view the location of the user. Other users will not be able to find the location of the user. Thus this is how we have implemented the idea of preserving privacy for the users.

6 Performance Evaluation

We evaluate the performance of the system by considering the following factors.

6.1 Effect of Number of Objects

The figure 8 below depicts the performance of our system with respect to increasing the number of objects from 1,000 to 5,000. Figure shows that when the number of objects increases, the communication cost of the resource-aware algorithm is only slightly affected, but the quality-aware algorithm significantly reduces the communication cost.

The broadcast step of the resource-aware algorithm effectively allows each sensor node to find an adequate number of objects to blur its sensing area. When there are more objects, the sensor node finds smaller cloaked areas that satisfy the k-anonymity privacy requirement, as given in Figure b. Thus the required search space of a minimal cloaked area computed by the quality-aware algorithm becomes smaller; hence the communication cost of gathering the information of the peers in such a smaller required search space reduces. Likewise, since there are less peers in the smaller required search space as the number of objects increases, finding the minimal cloaked area incurs less Minimum Bounding Rectangle (MBR) computation. Since our algorithms generate smaller cloaked areas when there are more users, the spatial histogram can gather more accurate aggregate locations to estimate the object distribution; therefore the query answer error reduces (figure c).

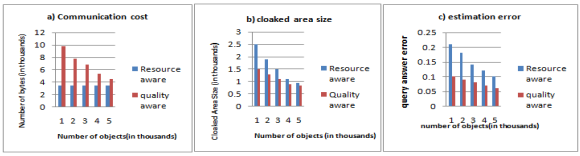


Fig. 8. Effect of number of objects

6.2 Effect of Various Anonymity Levels

The figure 9 depicts the performance of our system with respect to varying the required anonymity level k from 10 to 50. When the k-anonymity privacy requirement gets stricter, the sensor nodes have to enlist more peers for help to blur their sensing

areas; therefore the communication cost of our algorithms increases(figure a). To satisfy the stricter anonymity levels, our algorithms generate larger cloaked areas, as depicted (figure b). For the PP's algorithm, since there are more peers in the required search space when the input cloaked area gets larger, the computational cost of computing the minimal cloaked area by the PP's algorithm and the basic approach gets worse. However, the PP's algorithm reduces the computational cost of the basic approach by at least four orders of magnitude. Larger cloaked areas give more inaccurate aggregate location information to the system, so the estimation error increases as the required k-anonymity increases (figure c). The PP's algorithm provides much better quality location monitoring services than the resource-aware algorithm, when the required anonymity level gets stricter.

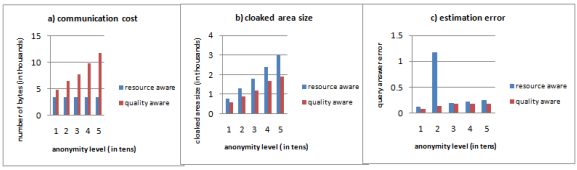


Fig. 9. Effect of anonymity level

7 Conclusion

In this paper, we propose a privacy-preserving location monitoring system for wireless sensor networks. We design an in-network location anonymization algorithm. The implementation is done as a simulated output. This can be further extended to function in a wide area and using the database with more capacity. This can also be implemented to suit android OS. The location can then be obtained based on the signals that are emitted by the mobile devices. The database in this case will be the mobile service providers who will have all the details about the users in the network. This is obtained during the registration. The efficiency of execution of the algorithm is 85% for all the tested user inputs.

References

[1] Chow, C.-Y., Mokbel, M.F., He, T.: A Privacy-Preserving Location Monitoring System for Wireless Sensor Networks. IEEE Transactions in Mobile Computing (October 2011)

[2] Xu, J., Tang, X.: Privacy-Conscious Location-Based Queries in Mobile Environments. IEEE Transactions on Parallel and Distributed Systems 21(3) (March 2010)

[3] Son, B., Shin, S., Kim, J., Her, Y.: Implementation of the real time people counting system using wireless sensor networks. IJMUE 2(2), 63–80 (2007)

[4] One systems Technologies, Counting people in buildings, `http://www.onesystemstech.com.sg/index.php?option=comcontent&task=view%&id=10`

[5] Traf-Sys Inc., People counting systems, http://www.trafsys.com/products/people-counters/thermal-sensor.aspx
[6] Ho, J.-W., Das, S.K.: Fast Detection of Mobile Replica Node Attacks in Wireless Sensor Networks Using Sequential Hypothesis Testing. IEEE Transactions on Mobile Computing 10(6) (June 2011)
[7] Traf-Sys Inc., People counting systems, http://www.trafsys.com/products/people-counters/thermal-sensor.aspx
[8] Gruteser, M., Schelle, G., Jain, A., Han, R., Grunwald, D.: Privacy-aware location sensor networks. In: Proc. of HotOS (2003)
[9] Kaupins, G., Minch, R.: Legal and ethical implications of employee location monitoring. In: Proc. of HICSS (2005)
[10] Location Privacy Protection Act of 2001 (2001), http://www.techlawjournal.com/cong107/privacy/location/s1164is.asp
[11] Title 47 United States Code Section 222 (h) (2), http://frwebgate.access.gpo.gov/cgibin/getdoc.cgi?dname=browse usc&do%cid=Cite:+47USC222
[12] Culler, D., Deborah Estrin, M.S.: Overview of sensor networks. IEEE Computer 37(8), 41–49 (2004)
[13] Perrig, A., Szewczyk, R., Wen, V., Culler, D.E., Tygar, J.D.: SPINS: Security protocols for sensor netowrks. In: Proc. of MobiCom (2001)
[14] Kong, J., Hong, X.: ANODR: Anonymous on demand routing with untraceable routes for mobile adhoc networks. In: Proc. of MobiHoc (2003)
[15] Kamat, P., Zhang, Y., Trappe, W., Ozturk, C.: Enhancing sourcelocation Privacy in sensor network routing. In: Proc. of ICDCS (2005)
[16] Guo, S., He, T., Mokbel, M.F., Stankovic, J.A., Abdelzaher, T.F.: On accurate and efficient statistical counting in sensor-based surveillance systems. In: Proc. of MASS (2008)
[17] Bohrer, K., Levy, S., Liu, X., Schonberg, E.: Individualized privacy policy based access control. In: Proc. of ICEC (2003)

Cluster Allocation Strategies of the ExFAT and FAT File Systems: A Comparative Study in Embedded Storage Systems

Keshava Munegowda[1], G.T. Raju[2], and Veera Manikandan Raju[1]

[1] OMAP Software, Texas Instruments (India) Pvt, .Ltd,
Bagmane Tech Park, Bangalore
{keshava_mgowda,veera}@ti.com,
keshava.gowda@gmail.com
[2] R.N.S. Institute of Technology, Bangalore, India
gtraju1990@yahoo.com

Abstract. The Multimedia card (MMC) and Secure Digital (SD) card associations classify the Extend File Allocation Table (ExFAT) as the standard file system for storage flash cards of more than 32 giga bytes (GB) of size. This paper attempts to explore the cluster allocation strategies of ExFAT file system in comparison with conventional FAT32 file system. The performance improvements by cluster allocation strategies of ExFAT file system are discussed. The adaptation techniques of cluster allocation strategies of ExFAT file system to the FAT32 file system are also discussed.

1 Introduction

The File Allocation Table (FAT) [1] file system is commonly used in embedded storage devices such as MultiMedia Cards (MMC) / Secure Digital (SD) / Micro SD cards, NOR, NAND flash memories and many more. Since the Flash memories are low-priced, smaller size and higher storage capacity, they are used in tablet personal computers, mobile phones, digital cameras and other embedded devices for data storage and multi-media applications such as video imaging, audio/video playback and recording. The initial version of FAT file system was FAT12 by Microsoft Corporation, later it was extended as FAT16 and further as FAT32 to support higher storage size. The FAT file system was initially developed to use on floppy disks and hard-drives. Since most of the Personal Computer (PC) s implements the FAT file system, this file system has become a default and world-wide compatible storage format for embedded devices. Even though FAT file system does not define flash management techniques such as wear-leveling and Bad Block management, the embedded devices implements this file system along with the dedicated flash block management algorithms. In FAT file system, the file or directory is the linked list of the clusters. A cluster is group of blocks or sectors of storage device. The File Allocation Table contains the linked list of clusters of files/directories. The maximum storage size supported by FAT32 file system is 32 GB. But, today the flash storage cards of more than 32 GB are available in market. The ExFAT [2] [3] file system is developed, by

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 691–698.
springerlink.com

Microsoft Corporation, as successor of FAT32 file system. This file system is optimally designed to support large size flash storage cards with higher read and write performance. The maximum storage size supported by ExFAT file system in 128 Peta Bytes (PB). The advanced differentiating features of the ExFAT file system in comparison with FAT File system are

i) Cluster search optimizations
ii) Contiguous clusters read algorithm

2 Cluster Search Optimizations

During file/directory creation and update operations, both FAT and ExFAT file systems searches for new cluster and allocated to a file/directory. The FAT and ExFAT file system uses a different approach for searching free clusters in the storage device. The Conventional FAT file system uses the linked cluster allocation scheme where as ExFAT uses linked cluster allocation scheme and contiguous cluster allocation scheme depending on the availability of the free clusters. The FAT file system examines the status FAT entries and if the status of the entry in the File allocation table is indicating free then the file data is written into the corresponding cluster and the FAT entry will be updated with allocation status. In case of FAT32 file systems, the free cluster search is performed starting from the cluster number specified in the field "FSI_Nxt_Free" (offset 492) [1] of the boot sector. The cluster scheduling method of NFAT [4] (New FAT file system) reduces the free cluster allocation time by using cluster scheduling algorithm operating in a cluster bank area and thus improves the file write performance, but this approach is not compatible with FAT file system. The TFS4 [5] provides contiguous clusters to file write operations by using RB (Red-Black) tree data structure. The RB tree of free clusters are created during file delete operations, during file creation and update operations these trees are traversed for free clusters. Traversing RB tree is much faster operations than searching for free cluster in File Allocation Table, but if there is no previous file delete operation then this TFS4 algorithm has no effect. The RB tree of free clusters are created in the main memory not in the file system hence the information of the binary trees of the free clusters will be lost when the system is restarted. The backward search method of embedded FAT file system [6] always starts the search from tail of clusters chain assuming the availability of contiguous free clusters to allocate to files during write operation, but in this method searching for free clusters in non empty file system requires longer time than conventional free cluster search method starting from the cluster number specified in the field "FSI_Nxt_Free" [1] of the boot sector. The clusters segmentations technique [7] improves the free cluster search by skipping the FAT entries read with in the allocation unit of the clusters segment, but this technique will not have any effect if the allocation unit is just one cluster in the cluster segment. The ExFAT file system optimizes the free cluster search by using "cluster heap". The cluster heap is a group of clusters. Every bit in the cluster heap specifies the status of the data cluster, thus every byte indicates the status of 8 clusters. The binary value "0" indicates that cluster is free and value 1 indicates that cluster is allocated. The cluster heap is also referred as "Allocation Bitmap". The cluster heap is similar to block bitmap and inode bit map structures used in Ext2 [8] and Ext3 [9] [10] file systems.

Figure 1 shows the logical organization of ExFAT file system and the structure of cluster heap. The File allocation tables FAT1 and FAT2 contain the linked list of the data clusters. Note that, FAT2 is not enabled by default in ExFAT; this is used only in another ExFAT file system variant called TexFAT [2] [11] file system. The TexFAT file system provides the transactional file system operations to ensure the file system Meta data consistency during uncontrolled power loss in the storage systems.

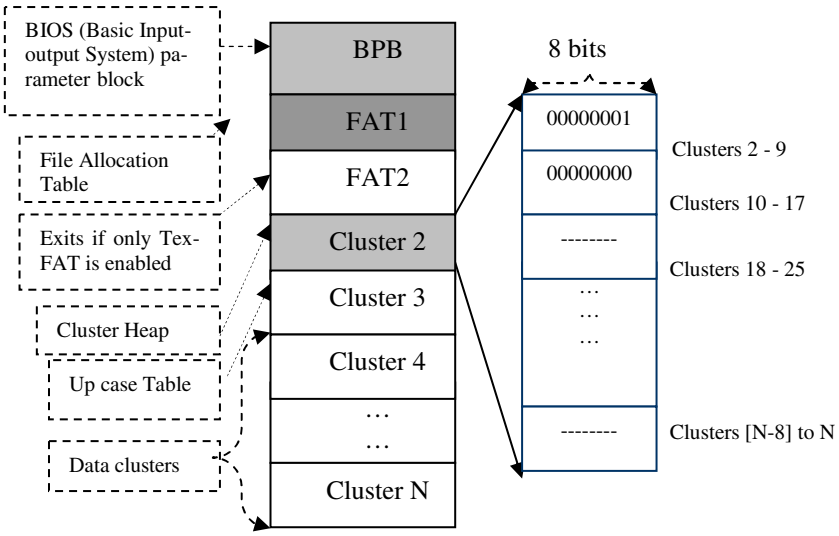


Fig. 1. ExFAT file system Organization and Cluster Heap

The ExFAT uses only one cluster heap where as TexFAT uses two cluster heap structures. The 1st cluster heap is for FAT1 and 2nd cluster heap is for FAT2. During, file/directory creation and update operations, instead of searching 32 bit entries of File Allocation Table, the ExFAT file system search for the free cluster in the cluster heap; Since the cluster heap is smaller in size compared to File Allocation Table, it can be cached in the primary memory and the searching is optimal thus improves performance of the file write. The cluster heap structure of ExFAT file system can be implemented in FAT file system as a file containing the every clusters status such as allocated or free. Note that the Upcase table shown in figure 1 is for case insensitive search of file/directory names during file/directory open operations.

3 Contiguous Clusters Read Algorithm

In ExFAT file system the cluster heap, as mentioned in section 2, is used for the searching for the free clusters while creating and updating files/directories. If the contiguous clusters are available for allocation, then the ExFAT file system does not update the cluster status in FAT entries, instead “No FAT chain” bit is reset to 0, to

indicate the contiguous number of clusters are allocated. The allocated cluster status bit is indicated in cluster heap. The "No FAT chain" bit is a part of generic primary flags of stream extension directory entry of files/directory. Another instance of "No FAT chain" bit field is a part of secondary primary flags of file name extension directory entry of file/directory.

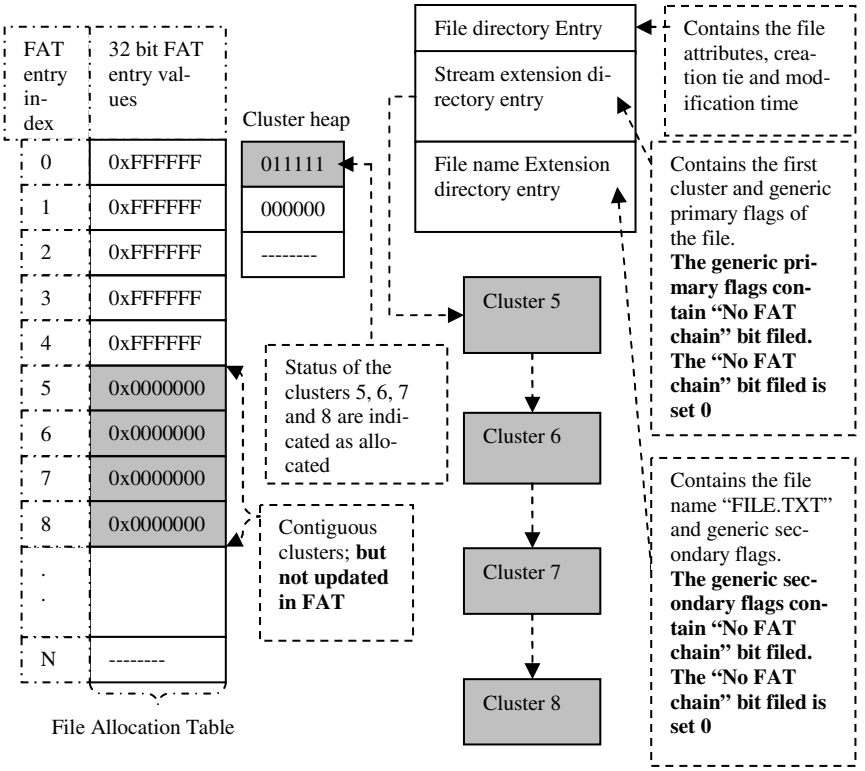


Fig. 2. Contiguous cluster allocation indication in ExFAT File system

As an example, if the FILE.TXT is created in the root directory with contiguous clusters 5, 6, 7 and 8 are allocated from the cluster heap, then the "No FAT chain" bit is reset to 0, and stream extension directory entry contains the first cluster number 5. The status of the clusters 5, 6, 7 and 8 are updated only cluster heap not in FAT; in File allocation table, the status remain as free clusters only. This scheme is shown in figure 2. During file read, the indication of contiguous cluster allocation improves the performance by avoiding the retrieval of FAT entries values and auto-incrementing the cluster values for the file data read. Suppose, the clusters allocated a file are not in contiguous, then linked list of clusters are updated in File allocation table.

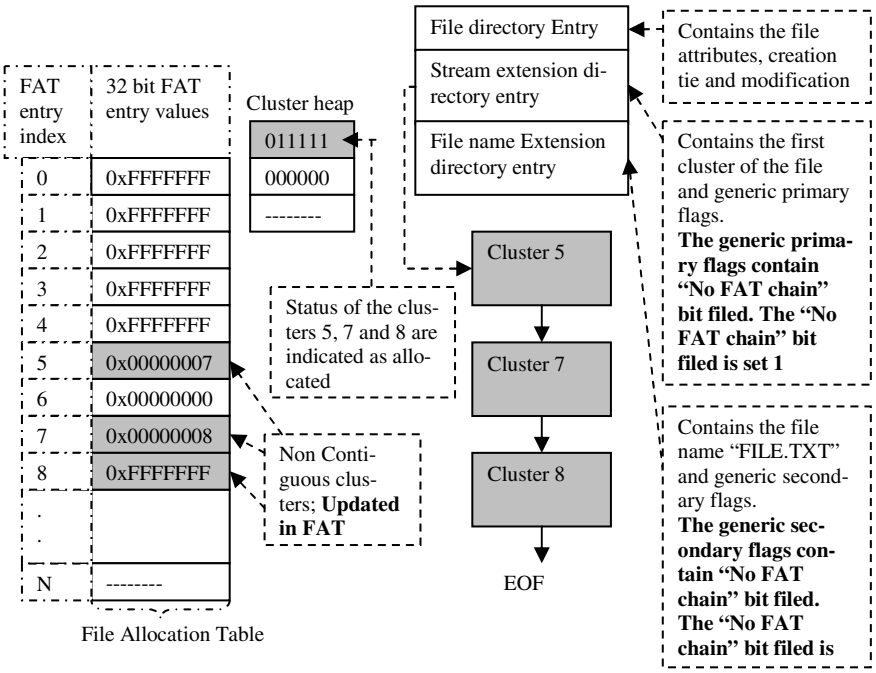


Fig. 3. Non Contiguous cluster allocation indication in ExFAT File system

For example, if the clusters 5, 7, 8 are allocated for the file name "FILE.TXT", then the status of these clusters are updated in cluster heap and in FAT also. This is shown in figure 3, note that the FAT entry 5 contains the value 7 indicating that 7 is the second cluster of the file, similarly the FAT entry 7 contains the value 8 indicating that cluster 8 is next cluster to cluster 7. The End Of File (EOF) value 0xFFFFFFFF at the FAT entry index 8 indicates that cluster 8 is the last cluster of the file. The "No FAT chain" bit of the generic primary flags field of stream extension directory entry and the another instance of "No FAT chain" bit of the generic secondary flags filed of file name extension directory entry are set to value 1 indicating that there are no contiguous clusters allocated to file. In such situations, the file read operations should retrieve the FAT entries values while reading the file data. The cluster group read algorithm [7] uses the most significant 4 bits of the FAT entries to indicate the contiguous cluster allocation for FAT32 file systems. This technique exploits the fact that in FAT32 file system, even though FAT entry size is 32 bits, only the 28 bits are used to store the cluster numbers. The MSB 4 bits are not used. But, this technique still writes the FAT entries even though allocated clusters are contiguous. These FAT entries update are required to maintain the compatibility of the FAT specification [1].

4 Experimental Results

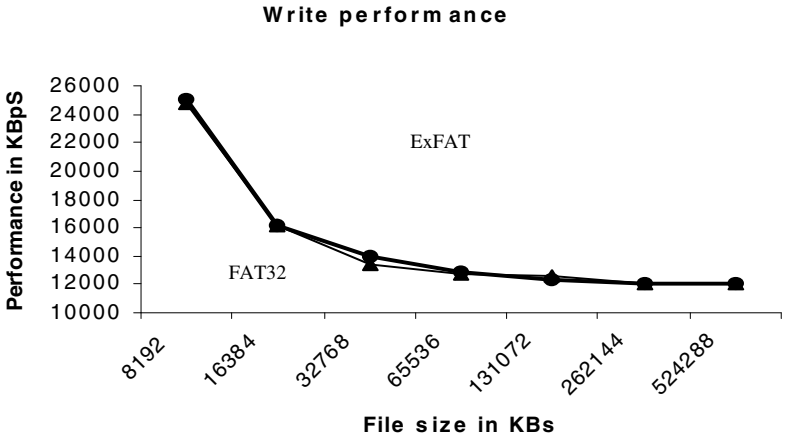


Fig. 4. File Write Performance of ExFAT and FAT32 file systems

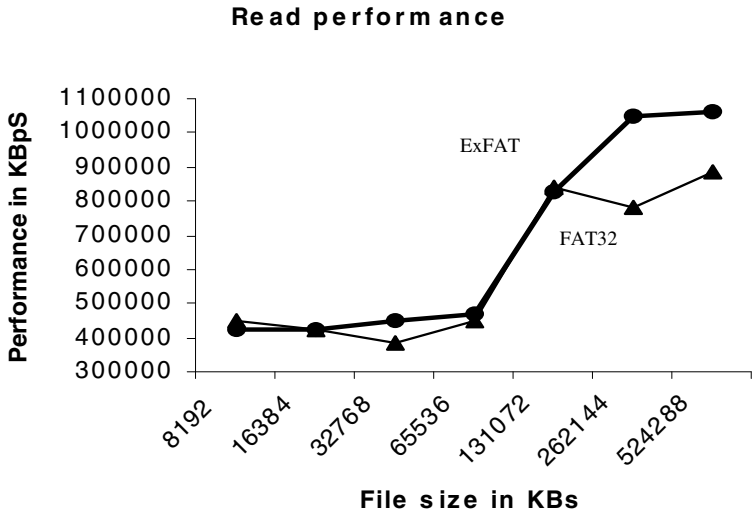


Fig. 5. File Read Performance of ExFAT and FAT32 file systems

The performance benchmarking of file write and read operations are conducted on Dell laptop named latitude D630 containing Intel core 2 duo processors of 2.4 GHz and 1.17 GHz speed. The Transcend 4 GB USB (Universal Serial Bus) drive is used as storage device. The file system benchmarking tool Iozone version 3.3 [12] is used in the Windows XP operating system for the performance measurements. The Iozone software uses the record size of 4MB (Mega Bytes) for file read and write operations. Both ExFAT [13] and FAT32 file systems sets the cluster size as 32KB (Kilo Bytes). The cluster search optimizations by cluster heap and avoiding FAT entry writes yields the file write performance improvement of 90-100 KBps (Kilo Bytes per Second), as shown in figure 4, and the contiguous cluster read file read performance improvement of 40-50 MBps (Mega Bytes per second) as shown in figure 5. Since, cluster writes are time consuming compare to cluster read operations, the file read performance is generally better than file write performance improvements.

5 Conclusion

The cluster search optimizations and contiguous cluster read algorithm of the ExFAT File system are examined. The cluster search optimizations improve file write performance and contiguous cluster read algorithm improves the file read performance. The contiguous cluster read algorithm will not have the effect, if the available free clusters are not contiguous. The scattered data clusters are formed gradually due to consistent file/directory creation and update operations in the file system. For data files, such as word documents, the update operations may lead to FAT entries update if the free clusters are scattered in File Allocation Table. The contiguous cluster read algorithm is optimal to use with the audio and video files, because usually these files are of larger size and user generally does not update/modify these files.

References

1. Microsoft, FAT32 File System Specification, FAT: General Overview of On-Disk Format (2000)
2. Puipeddi, R.V., Ghotge, V.V., Thind, R.S.: Quick file name look up using name hash. USPTO application: 12/389396, filed on February 20 (2009)
3. Munegowda, K., Venkatraman, S., Raju, G.T.: The Extend FAT file system: Differentiating with FAT32 file system. In: Linux Conference, Prague, Czech Republic, Europe (October 2011)
4. Choi, M., Park, H., Jeon, J.: Design and Implementation of a FAT File System for Reduced Cluster Switching Overhead. In: International Conference on Multimedia and Ubiquitous Engineering (2008)
5. Samsung, TFS4 Contiguous Cluster Allocation, version 1.0 (January 2008)
6. Zhang, J.: Research of Embedded FAT file system. In: IEEE International Conference on Uncertainty Reasoning and Knowledge Engineering (2011)
7. Munegowda, K., Raju, G.T., Raju, V.M.: Performance and Space Optimization techniques for FAT File system for embedded storage devices. In: International Conference on Data Engineering and Communications, ICDECS (December 2011)

8. Card, R., Ts'o, T., Tweedie, S.: Design and implementation of the Second Extended File system. In: First Dutch International Symposium on Linux
9. Tweedie, S.C.: Journaling the Linux ext2fs File system. In: Proceedings of the 4th Annual LinuxExpo, Durham, NC (May 1998) (retrieved June 23, 2007)
10. Tweedie, S.C.: Ext3, journaling file system. In: Ottawa Linux Symposium, Ottawa Congress Centre, Ottawa, Ontario, Canada (July 20, 2000)
11. Microsoft Corporation, TexFAT file system, http://msdn.microsoft.com/en-us/library/cc907927.aspx
12. Iozone, File systems Benchmarking tool, http://www.iozone.org
13. ExFAT file system downloadable software package for WindowsXP, http://www.microsoft.com/download/en/details.aspx?id=19364
14. Fuse based ExFAT implementation for Linux, http://code.google.com/p/exfat/

Audio Steganography Used for Secure Data Transmission

Pooja P. Balgurgi[1] and Sonal K. Jagtap[2]

[1] Department of E &TC Engineering, SKNCOE, Pune, (MS)- 411041
pooja4858@gmail.com
[2] Department of Electronics and Telecommunication Engg, SKNCOE, Pune, (MS)-411041
sonalkjagtap@gmail.com

Abstract. Information hiding technique is a new kind of secret communication technology. The majority of today's information hiding systems uses multimedia objects like image, audio, video. Audio Steganography is a technique used to transmit hidden information by modifying an audio signal in an imperceptible manner. It is the science of hiding some secret text or audio information in a host message. The host message before steganography and stego message after steganography have the same characteristics. Embedding secret messages in digital sound is a more difficult process. Varieties of techniques for embedding information in digital audio have been established. This paper presents comprehensive survey of some of the audio steganography techniques for data hiding. Least Significant Bit (LSB) technique is one of the simplest approach for secure data transfer. In this paper different data hiding methods used to protect the information are discussed. Audio data hiding is one of the most effective way to protect the privacy.

Keywords: Steganography, Audio Steganography, Cryptography, Least Significant Bit (LSB) Coding, Information Security, Human Auditory System (HAS).

1 Introduction

By development of computer and the expansion of its use in different areas of life and work, the issue of security of information has gained specific importance. A message is hidden within a cover signal in the block called embed and audio processing block using a stego key, which is same at the transmitter and receiver side. The output of this block is stego audio signal. At the receiver side, the embedded message is retrieved from the cover audio signal using stego key in the block called extract and audio processing.

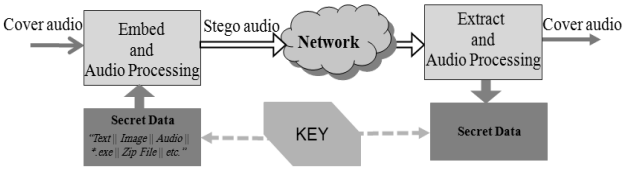


Fig. 1. Block diagram of information hiding and retrieval

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 699–706.
springerlink.com

Fatiha Djebbar_, Beghdad Ayady, Habib Hamamzand Karim Abed-Meraimx has proposed novel and versatile audio steganographic methods. H.B.Kekre, Archana Athawale, Swarnalata Rao, Uttara Athawale has proposed a novel method by considering the parity and XORing operation. Masahiro Wakiyama†, Yasunobu Hidaka†, Koichi Nozaki†† has proposed extended low bit coding. Muhammad Asad, Junaid Gilani, Adnan Khalid has proposed the enhanced LSB technique. Poulami Dutta1, Debnath Bhattacharyya1, and Tai-hoon Kim2 has proposed general principles of audio steganography. Pradeep Kumar Singh, R.K.Aggrawal has proposed image hiding in audio signal.

2 Literature Survey

There are three types of domains & various techniques of audio steganography are present in each domain. They all are having the different embedding methods.

2.1 Temporal Domain

The temporal domain contains LSB coding and echo hiding method.

2.1.1 Least Significant Bit Coding

It consists of embedding each bit from the message in the least significant bit of the cover audio in a specific way. The LSB method gives high embedding capacity for data and is relatively easy to implement and to combine with other hiding techniques. Generally the length of the secret message to be encoded is smaller than the total number of samples in a sound file.

2.1.2 Echo Hiding

By introducing short echo to the host signal, the echo hiding method embeds the data into audio signal. Once the echo has been added, the stego signal retains the same statistical and perceptual characteristics. In this method the data hiding depends on three parameters of the echo signal: initial amplitude, offset (delay) and decay rate so that the echo is not able to be heard. The effect is indistinguishable for a delay up to 1 ms between the original signal and the echo.

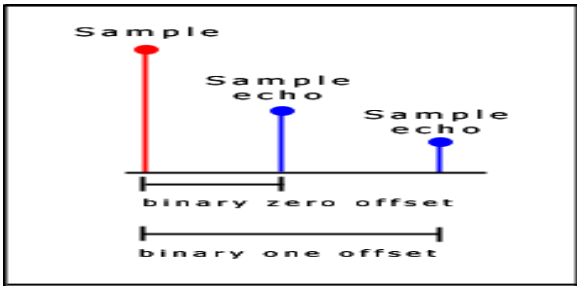


Fig. 2. Echo Hiding Technique

2.2 Frequency Domain

Frequency domain contains phase coding, spread spectrum and tone insertion.

2.2.1 Phase Coding

Phase coding method makes use of the human audio system insensitivity to relative phase of different spectral components. The data is hided in selected phase components of the original speech spectrum. For inaudibility purpose, phase components modification should be kept small. In this method, imperceptible phase modifications are obtained using controlled phase alteration of the host audio. The method has an embedding capability of 20 to 60 bps in 44.1 kHz, is resistant to compression.

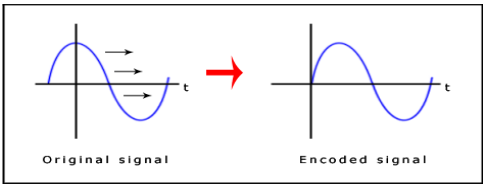


Fig. 3. Phase Coding Technique

2.2.2 Spread Spectrum

Spread spectrum technique spreads hidden signal data through the frequency spectrum. Spread Spectrum (SS) is developed in communications to ensure a proper recovery of a signal sent over a noisy channel by producing redundant copies of the data signal. Basically, Data is multiplied by an M-sequence code known to sender and receiver, and then embedded in the cover audio. Thus, if noise corrupts some values, there will still be copies of each value left to recover the embedded message[1].

2.2.3 Tone Insertion

Tone insertion techniques rely on the inaudibility of lower power tones in the presence of significantly higher ones. The ”masking” effect is a property of HAS which make any weak speech component imperceptible by listeners in presence of a much louder one. To embed one bit in a speech frame, a pair of tones is generated at two frequencies. By inserting tones at known frequencies and at low power level, concealed embedding and correct data extraction are achieved[1].

2.3 Wavelet Domain

Wavelet domain contains wavelet coefficients explained in the present section.

2.3.1 Wavelet Coefficients

This method is based on discrete wavelet transform. Data is embedded in the LSBs of the wavelet coefficients achieving high capacity of 200 kbps in 44.1 kHz audio signal. For improving embedding data imperceptibility, it employed a hearing threshold when embedding data in the integer wavelet coefficients, while avoiding data hiding

in silent parts of the audio signal. Even though data hiding in wavelet domain gives high embedding rate, data extraction at the receiver side might have some errors that it gives lossy data[1].

3 Embedding Methods in LSB Technique

There are several methods of data embedding in LSB technique.

3.1 Lowest Bit Coding

The lowest bit coding is the method that embeds secret data only in the least significant bit (LSB). This method minimizes the transition before and after the audio is embedded. Since the audio data use only the lowest bit, this method gives embedding capacity upto one eighth or 12.5% of wave file. In figure 4 we show the way how the wave data is embedded using the lowest bit method. We express wave data and secret data by binary digits and replace secret data in low bit with the wave data one by one.

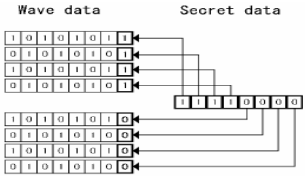


Fig. 4. The lowest bit coding

3.2 The Variable Low Bit Coding

It is the improved version of the lowest bit coding which increases the embedding capacity. Consider the range of audio data is from 0 to 255. The middle range of audio data is 128 and at that range the sound is a silent. Therefore we don't use this range in audio data for embedding. Data embedding in the silent sound gives existence of the secret data. Therefore we don't embed the secret data into the middle range data. The middle range 128 is used to calculate the standard level of sound.

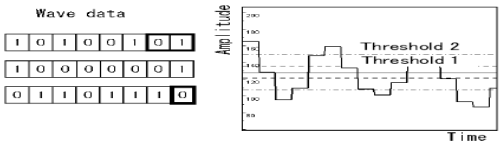


Fig. 5. The variable low bit coding

3.3 Average Amplitude Method

This method uses the average amplitude data of surroundings audio data as threshold. We calculate the average about absolute value of amplitude regarding middle value 128 as a value 0. The average of an amplitude level for audio data is calculated by 10 audio data about before and after 5 audio data except for own audio data about all sample data. If the level of the amplitude is bigger than that of the average value, 2 binary digits are used for the embedding data. If not, binary digits are not used for the embedding data.

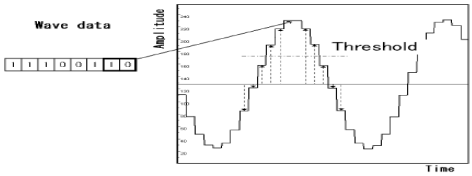


Fig. 6. The principle of average amplitude method

3.4 Parity Coding

In this method, instead of breaking a signal down into individual samples, this method breaks a signal down into separate regions of samples and encodes each bit from the secret message in a sample region's parity bit. If the parity bit of a selected region does not match the secret bit to be encoded, the process flips the LSB of one of the samples in that region. Thus, the sender has more choice in encoding the secret bit, and the signal can be changed more attractively.

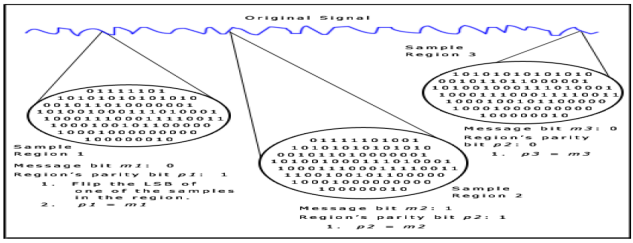


Fig. 7. Parity Coding

3.5 XORing of LSB's Method

This method performs XOR operation on the LSBs and depending on the result of XOR operation and the message bit to be embedded, the LSB of the sample is modified or kept unchanged. The XOR operation on first 2 LSBs is described below. To increase the level of encryption XORing can be further expanded upto16 LSBs. This method is simple to implement, and is computationally inexpensive. The tabular representation of the embedding procedure is as given in Table 1.

Table 1. Procedure of Data Embedding using XORing

LSB	Bit next to LSB	XOR	Action if message bit is 0	Action if message bit is 1
0	0	0	No Change	Flip LSB
0	1	1	Flip LSB	No Change
1	0	1	Flip LSB	No Change
1	1	0	No Change	Flip LSB

4 LSB Method with Parity Coding

Instead of using the single LSB technique, we are using the LSB technique with parity method which is simplest method and easy to implement and due to combining the LSB with the parity coding, it gives high level of security. In this method, the LSBs of digitized samples are not directly replaced with the message bits, it first checks the parity of the samples and then carries out data embedding[2]. The process of data embedding and data retrieval are explained as follows:

Steps for Data embedding:

1. Read the cover audio signal.
2. Read the message to be embedded, its size should be less than the size of the cover audio signal and convert it into binary sequence of message bits.
3. Depending upon the value of the message bit to be embedded (0/1), the LSB of the sample of the cover audio signal is modified or unchanged.
4. If the message bit to be embedded is 0, then the LSB of the sample of the cover audio signal is modified or unchanged such that the parity of the sample after embedding of this message bit is even.
5. If the message bit to be embedded is 1, then the LSB of the sample of the cover audio signal is modified or unchanged such that the parity of the sample after embedding of this message bit is odd.
6. The modified cover audio samples are then written to the file forming the stego audio signal.

Steps for Data Extraction/ Retrieval:

1. The Stego audio file is read.
2. The parity of every sample of the stego is checked.
3. If the parity is even, then the message bit retrieved is 0.
4. If the parity is odd, then the message bit retrieved is 1.
5. After every such 16 message bits are retrieved, they are converted to decimal equivalents.
6. Finally the secret message is reconstructed.

5 Results and Discussion

The below table shows the comparison of various methods of audio steganography in terms of strengths, weakness and hiding rate. By considering the high embedding capacity, the wavelet coefficients and the least significant bit method is suitable for data

hiding, because other techniques gives low embedding capacity. In case of wavelet coefficient, lossy data is retrieved. So we cant detect the original message, it may be changed. In LSB coding, the message is easy to destroy, but we can combine the LSB with another method to give more security, so it cannot be detectable to the 3rd party. LSB is simple method and is easy to implement, so that we are using the LSB method along with the parity coding.

Table 2. Comparison of Audio steganography methods

Hiding domain	Methods	Strengths	Weakness	Hiding rate
Temporal domain	1.Least significant bit coding	Simple and easy way of hiding information with high bit rate	Easy to extract and to destroy	16 kbps
	2.Echo hiding	Resilient to lossy data compression algorithms	Low security and capacity	40-50 bps
Frequency domain	1.Phase coding	Robust against signal processing manipulation and data retrieval needs the original signal	Low capacity	333 bps
	2.Spread spectrum	Provide better robustness	Vulnerable to time scale modification	20 bps
	3.Tone insertion	Imperceptibility and concealment of embedded data	Lack of transparency and security	250 bps
Wavelet domain	Wavelet coefficients	Provide high embedding capacity	Lossy data retrieval	200 kbps

6 Conclusions

- Audio Steganography is more challenging than Image Steganography because the human Auditory System (HAS) has more precision than Human Visual System (HVS).
- Steganography complements rather than replaces encryption by adding another layer of security, it is much more difficult to decrypt a message if it is not known that there is a message.
- The Enhanced LSB is more secure than any other method & improves the conventional LSB method & make it more secure.
- It gives great security & the embedded message cannot be extracted without the knowledge of embedding process.
- There is no difference between stego audio signal & the original audio signal i.e. hidden information is recovered without any error.

7 Future Scope

Future Scope of this paper is to increase the capacity as well as to improve the confidentiality of audio steganography. Enhance the storage capacity of the system in

terms of each method, and the possibilities of improvements in the method of hiding in audio. First area is focus on how much maximum data can me hidden in audio signal and make it robust. Secondly improve the methods by applying mixed approaches means make the system more secure toward detection by using the combination of various techniques of data hiding in audio signals.

References

1. Djebbar, F., Ayady, B., Hamamz, H., Abed-Meraimx, K.: A view on latest audio steganography techniques. In: International Conference on Innovations in Information Technology (2011)
2. Kekre, H.B., Athawale, A., Rao, S., Athawale, U.: Information Hiding in Audio Signals. International Journal of Computer Applications (0975–8887) 7(9) (October 2010)
3. Asad, M., Gilani, J., Khalid, A.: An Enhanced Least Significant Bit Modification Technique for Audio Steganography, IEEE978-1-61284-941-6/111$26.00 (2011)
4. Wakiyama, M., Hidaka, Y., Nozaki, K.: An audio steganography by a low-bit coding method with wave files. In: Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (2010)
5. Singh, P.K., Aggrawal, R.K.: Enhancement of LSB based Steganography for Hiding Image in Audio. International Journal on Computer Science and Engineering 02(05) (2010)
6. Dutta, P., Bhattacharyya, D., Kim, T.-H.: Data Hiding in Au-dio Signal: A Review. International Journal of Database Theory and Application 2(2) (2009)
7. Sridevi, R., Damodaram, A., Narasimham, S.: Efficient method of audio stega-nography by modified lsb algorithm and strong encryption key with enhanced security. Journal of Theoretical and Applied Information Technology (2005-2009)

An XML Parser of Efficient Updates for a Binary String: A Case Study

J. Bhagyashala and S. Shefali

WCE Sangli, India
{bjadhawar123,shefali.sonavane}@gmail.com

Abstract. In many emerging applications, such as XML publishing systems, electronic commerce and intelligent Web searching, ordered XML data are available in query processing. An XML query processing based on labeling schemes has been thoroughly studied in the past several years. However, all these techniques have high update cost, cannot completely avoid re-labeling in XML updates and increase the label size. This paper experiments a labeling scheme, called IBSL (Improved Binary String Labeling), which supports order sensitive updates without relabeling or recalculation. By using IBSL, University Web search has been considered as a separate case study using conventional Google search and applying IBSL algorithm along with the search. This paper reports that the IBSL algorithm is time efficient.

Index Terms: Web search, dynamic XML, order-sensitive update, tree labeling.

1 Introduction

A number of labeling schemes have been designed to facilitate the query of XML based on which the ancestor-descendant relationship between any two nodes [1]. Another important feature of XML is that the elements in XML are intrinsically ordered. However, the label update cost is high based on the present labeling schemes. They have to re-label the existing nodes or re-calculate some values when inserting an order-sensitive element. Thus, it is important to design a scheme that supports order-sensitive queries, yet it has low label update cost. Focusing Intelligent Web searching technique for generalize idea of XML parser by using the Improved Binary String Labeling algorithm.

XML has become a standard to represent and exchange data on the web. In the definition of XML, one Element is allowed to refer to another, therefore theoretically an XML document is a graph. However for simplicity, most of the research works process queries over the XML data that conform to an ordered tree-structured data model. Fig.1 shows an ordered XML tree.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 707–714.
springerlink.com

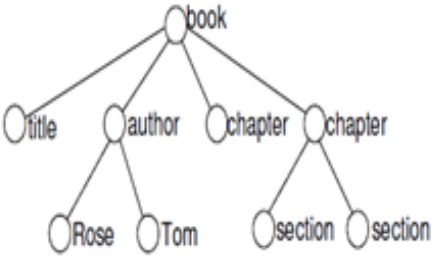


Fig. 1. An ordered XML tree

A binary tree is made of nodes, where each node contains a "left", a "right" node, and a data element. The "root" node (book) points to the topmost node in the tree. The left (title) and right (author) node recursively point to smaller "subtrees" on either side. A null node represents a binary tree with no elements i.e. the empty tree. The formal recursive definition is: a binary tree is either empty (represented by a null node), or is made of a single node, where the left and right nodes (recursive definition ahead) each point to a binary tree.

Elements in XML data can be labeled according to the structure of the document to facilitate query processing. The labeling schemes, such as containment scheme, prefix scheme [2] and prime scheme can determine the ancestor-descendant (A-D), parent-child (P-C) etc. relationships efficiently in XML query processing if XML data are static. The elements in XML are intrinsically ordered, which is referred to as the document order (the element sequence in XML). The relative order of two paragraphs in XML is important because the order may incense the semantics of XML. In addition, the standard XML query languages XPath [3] and XQuery [4] include both ordered and un-ordered queries. Thus, it is very important to maintain the document order when XML is updated.

2 Proposed Work

The main contributions of this paper are summarized as follows:

1. Generation of a labeling scheme for XML documents, named as Improved Binary String Labeling (IBSL), which takes advantage of the lexicographical order of binary strings. IBSL neither relabels any existing nodes, nor recalculates any values when inserting an order-sensitive leaf node and subtree into the XML tree [1].
2. Implementation of algorithms to implement IBSL in which the cost of updating ordered XML is much lower than that of the existing schemes [2].
3. Experimentation of IBSL also for to extract reuses of oder-sensetive queries and leaf node updates.

3 Related Work

XML queries can be expressed as linear paths or twig patterns. To solve this relabeling problem, Amagasa et al. [5] has proposed the use of float-point values for the "start"s and "end"s of the intervals. However, this approach still cannot avoid relabeling in the case of frequent insertions. The DeweyID scheme [6] requires the sibling nodes located after the inserted node and the descendants of theses siblings to be relabeled to maintain the order. In the context of XML twig query processing, the extended Dewey scheme can derive the names of all the elements along the path from the root. OrdPath [7] is tolerant to insertions; however, as pointed out by, it suffers from poor query performance. Wu et al. [8] proposed the Prime number labeling scheme to label XML trees. But, Prime needs to recalculate the SC values based on the new ordering of the nodes. In addition, Li et al. proposed two schemes, QED [9] and CDBS [10], which include novel encoding methods to support code (a binary or quaternary string) insertion into a sequence of existing codes without disturbing the order between the nodes or having to relabel them. However, while CDBS can efficiently process dynamic XML data, it cannot completely avoid relabeling; QED can avoid relabeling; perhaps, its label size is large and its update and query performance is not as good as that of CDBS.

4 Labeling Scheme

Scheme elaborates on IBSL, which is a binary-string-based prefix scheme. The most important feature of IBSL is that it compares the labels based on their lexicographical order rather than their numerical order. With IBSL, labels can be inserted between any two consecutive labels with their order being kept and without relabeling the existing labels.

Definition :-(Lexicographical order <):

Given two consecutive binary strings S_{left} and S_{right}, S_{left} is said to be lexicographically equal to S_{right} if they are exactly the same. To determine whether S_{left} is lexicographically smaller than S_{right}, i.e., $S_{left} < S_{right}$, the following procedure is performed.

1. The lexicographical comparison of S_{left} and S_{right} is performed bit by bit from left to right. If the current bit of S_{left} is 0 and the current bit of S_{right} is 1, then $S_{left} < S_{right}$ and the comparison is stopped, or
2. If $len(S_{left}) < len(S_{right})$, S_{left} is a prefix string of S_{right}, and the remaining bits are 1 except for the prefix string of S_{left} then $S_{left} < S_{right}$ and the comparison is stopped,
3. If $len(S_{left}) > len(S_{right})$, S_{right} is a prefix string of S_{left}, and the remaining bits are 0 except for the prefix string of S_{right}, then $S_{left} < S_{right}$ and the comparison is stopped.

For example, given two binary strings 10 and 110, 10 < 110 lexicographically (condition 1). Given two binary strings 1100 and 11001, 1100 < 11001 (condition 2), while 11000 < 1100 (condition 3).

Algorithm: Assign, Insert, update label (URL result) at memory location.
Input: *left self _label* N_{left} and *right self _label* $N_{right.}$
Output: *new inserted self_ label* $N_{new.}$
Begin
/* *substring (self _label N, 1, P-1)* extracts the *P-1* long bits from 1st position of *self_ label N*, where *P* denotes the position of the leftmost different string between N_{left} and N_{right}. */
Case1. N_{left} is empty but N_{right} is not empty;
$N_{temp} = N_{right} \oplus 0$; // N_{temp} is the temporary binary string
if $N_{temp} < N_{right}$ lexicographically **then** $N_{new} = N_{temp}$; return N_{new}; **end if**
Case 2. N_{left} and N_{right} are not empty;
if $(len(N_{left}) < len(N_{right}))$ **then** // Case 2(a)**if** N_{left} is the prefix string of N_{right} **then** $N_{new} = N_{right} \oplus 0$;
else if N_{left} is not the prefix string of N_{right} **then if** $len(N_{left})$ is not equal to *P* **then** N_{temp} = substring $(N_{right}, 1, P - 1)$; $N_{new} = N_{temp} \oplus 0$;
return N_{new}; **else if** $len(N_{left})$ is equal to *P* then $N_{new} = N_{right} \oplus 0$; return N_{new}; **end if end if**
else if $(len(N_{left}) = len(N_{right}))$ **then** // Case 2(b) **if** all the extracted same bits of N_{left} and N_{right} are **"1" then** N_{temp} =substring $(N_{left}, 1, P - 1)$;$N_{new} = N_{temp} \oplus 0$; return N_{new}; **else**N_{new} = substring $(N_{left}, 1, P - 1)$;**end if else if** $(len(N_{left}) > len(N_{right}))$ **then** // Case 2(c) **if** N_{right} is the prefix string of N_{left} **then**
$N_{new} = N_{left} \oplus 1$; return N_{new}; **else if** N_{right} is not the prefix string of N_{left} **then** N_{temp} = substring $(N_{left}, 1, P - 1)$; $N_{new} = N_{temp}$;return N_{new}; **end if end if**
Case 3. N_{left} is not empty but N_{right} is empty; $N_{temp} = N_{left} \oplus 1$; **if** $N_{temp} < N_{right}$ lexicographically **then**
$N_{new} = N_{temp}$; return N_{new};
getResult:Url
BinConvert: return binary url **if** NewUrl memAlocate: randomLocation **else** referPrevious: retrieve url
if NextStringSearch go to step above **if** SearchSave: memoryAlocate
if garbageCollected: updateLocation **end if End**

The above algorithm inserts the label with the smallest length between two labels in the case of both insertions and deletions. The main idea of algorithm is that it compares N_{left} and N_{right} bit by bit to find N_{new} such that N_{new} has the smallest length of all of the labels between N_{left} and N_{right} lexicographically. In addition to the algorithm described in [1], the additional provision is made to work with URL.

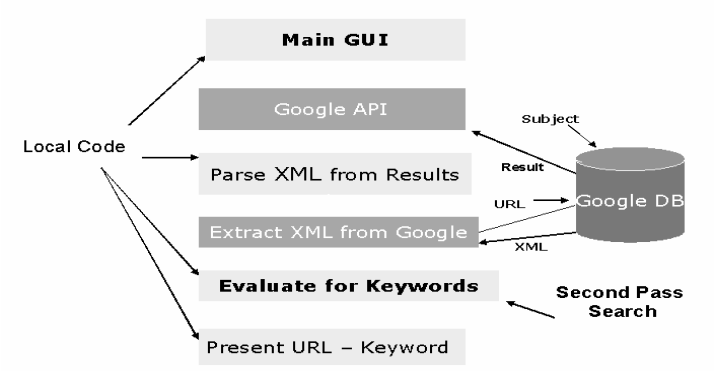


Fig. 2. XML Search

The Figure-2 shows the XML Search diagram that can analyze XML process flow using Google APIs. Local code will interact as an input to system with main GUI, Google API. Also system will provide keyword as an input for further search through Google database.

Scenario of XML: In this project we will use XML DOM/SAX Parser APIs along with user defined APIs to store, retrieve, update, delete nodes XML request will traverse through database and act as a node tuner Reverse way response will come through XML response to GUI as an output.

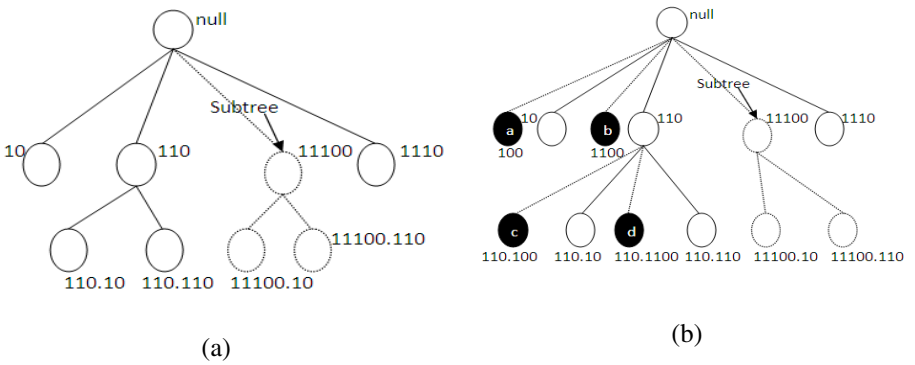


Fig. 3. Example of (a) IBSL and (b) Update operation

In Fig. 3(a) shows an example of IBSL. The prefix label's of the three child nodes (non-shaded and non-dotted circles: 10, 110, 1110) are all empty strings; thus, the self label is exactly the same as the complete label for the three-child nodes. The label of the node concatenates its parent's label (prefix_label) and its own label (self_label).

In Fig. 3(b) the shaded circle denotes the leaf node update and the white dotted circle denotes subtree update. Based on algorithm described a binary string can be inserted between two existing labels without the need for relabeling. For example, when inserting node **'a'** in Fig. 3 (Case 1), the *self_label* of **'a'** is 100 (10 ⊕0 → 100). When inserting node **'b'** (Case 2), since the *left self_label* of **'b'** is 10 with length 2 and the *right self _label* of **'b'** is 110 with length 3, we directly concatenate one more 0 after the *right self_label* (110 ⊕ 0 → 1100), whereupon the *self_ label* of **'b'** is 1100.Same procedure is follows by *self_lable* **'c'** and **'d'**.

5 Experimental Performance

A case study *University* web search considers two types of query searches; one is on the basis of catogery word and the other is catogery.

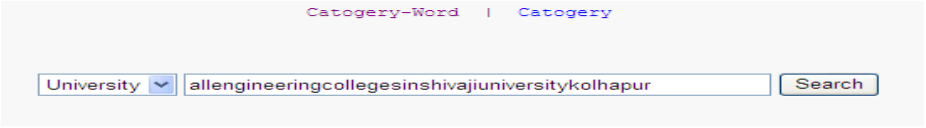


Fig. 4. Search Type

Fig.4 shows **University** is a root node and **allengineeringcollegesinshivaji universitykolhapur** as a leaf node. In catogery type, only can search root node. In a binary string every space is a same binary value if we can put query with space which is considered as end of the string. Therefore in IBSL we can put sentence in continuous.

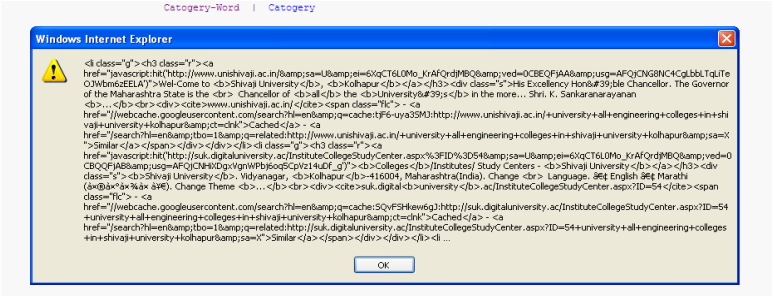


Fig. 5. XML Format

Fig.5 shows an XML Format of the Google database with type catogery word shown in Fig.4

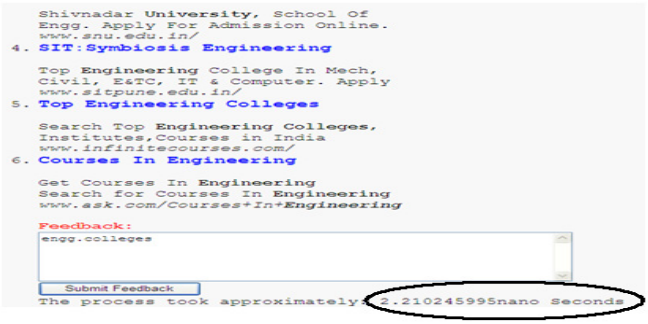


Fig. 6. IBSL Search

Fig.6 shows IBSL Search of similar query carried out in Fig.5. The time requirement of the query processing in Fig.6 is in nano seconds; where the time requirement of the same query by Google search which is shown in Fig.7 is in seconds.

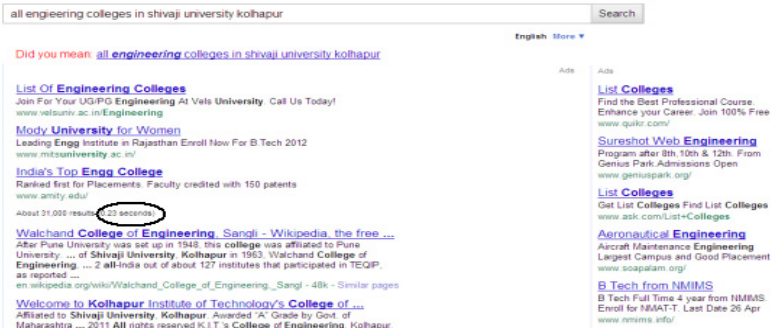


Fig. 7. Google Search

6 Remark

The Experiments were carried out on a Intel(R) Core(TM) I_3 Processoser, 2.86 GB of RAM running Microsoft Windows XP Professional Service pack 2. The XML data sets (and their corresponding labels) were stored in shield SQL, Yog SQL, The setup also requires database connectivity to MYSQL and SQL 2000. Similar observations are obtained with few of the other case datasets as shown in Fig.8 and it has been observed that ISBL algorithm proves more time efficient and the time requirement varies linearly with the no. of nodes in the XML tree.

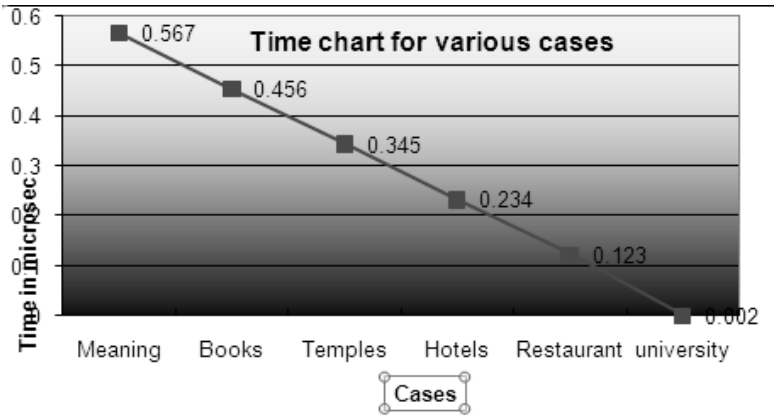


Fig. 8. Results of Datasets in various cases

References

[1] Ko, H.-K., Lee, S.: A Binary String Approach for Updates in Dynamic Ordered XML Data. IEEE Transactions on Knowledge and Data Engineering 22(4), 602–607 (2010)

[2] Li, C., Ling, T.W., Hu, M.: Efficient Updates in Dynamic XML Data: From Binary String to Quaternary String. Very Large Data Bases J. 17(3), 573–601 (2008)

[3] Berglund, A., Boag, S., Chamberlin, D., Fernandez, M.F., Kay, M., Robie, J., Simon, J.: XML path language (XPath) 2.0. W3C working draft 04 (April 2005)

[4] Boag, S., Chamberlin, D., Fernandez, M.F., Florescu, D., Robie, J., Simon, J.: XQuery 1.0: An XML Query Language. W3C working draft 04 (April 2005)

[5] Amagasa, T., Yoshikawa, M., Uemura, S.: QRS: A Robust Numbering Scheme for XML Documents. In: Proc. Int'l Conf. Data Eng. (ICDE), pp. 705–707 (2003)

[6] Lu, J., Ling, T.W., Chan, C.-Y., Chen, T.: From Region Encoding to Extended Dewey: On Efficient Processing of XML Twig Pattern Matching. In: Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 193–204 (2005)

[7] O'Neil, P., O'Neil, E., Pal, S., Cseri, I., Schaller, G.: ORDPATHs: Insert-Friendly XML Node Labels. In: Proc. ACM SIGMOD, pp. 903–908 (2004)

[8] Wu, X., Lee, M.-L., Hsu, W.: A Prime Number Labeling Scheme for Dynamic Ordered XML Trees. In: Proc. Int'l Conf. Data Eng. (ICDE), pp. 66–78 (2004)

[9] Li, C., Ling, T.W.: QED: A Novel Quaternary Encoding to Completely Avoid Re-Labeling in XML Updates. In: Proc. Int'l Conf. Information and Knowledge Management (CIKM), pp. 501–508 (2005)

[10] Li, C., Ling, T.W., Hu, M.: Efficient Processing of Updates in Dynamic XML Data. In: Proc. Int'l Conf. Data Eng. (ICDE), pp. 13–22 (2006)

SVM-DSD: SVM Based Diagnostic System for the Detection of Pomegranate Leaf Diseases

Sanjeev S. Sannakki[1], Vijay S. Rajpurohit[1], and V.B. Nargund[2]

[1] Gogte Institute of Technology, Belgaum, Karnataka, India
[2] University of Agricultural Sciences, Dharwad, Karnataka, India
sannakkisanjeev@yahoo.co.in, vijaysr2k@yahoo.com, nargund56@gmail.com

Abstract. This work proposes a methodology for detecting pomegranate leaf diseases early and accurately using image processing techniques and Support Vector Machine (SVM). Color image segmentation using K-means clustering technique is performed to extract the region of interest from the pomegranate leaf image. Further significant texture and color features are extracted from the region of interest for the purpose of training SVM classifier. Classification is performed by considering two different feature sets viz. i) entropy and saturation ii) hue and energy. Experimental results show that SVM classification is highly accurate with entropy and saturation feature set compared to that of energy and hue set. This automated system assists farmers to detect the healthy & diseased leaves without human intervention.

Keywords: Plant Pathology, Color Transformation, K-means Clustering, Feature Extraction, Support Vector Machine (SVM).

1 Introduction

Agriculture plays a key role in the development of human civilization. Plant disease is one of the crucial causes that reduces quantity and degrades quality of the agricultural products. Excessive use of pesticides for plant disease treatment raises the danger of toxic residue levels as well as leads to groundwater contamination.

Pomegranate *(Punica granatum),* the so called *"fruit of paradise"* is one of the major fruit crops of arid region. In India it is cultivated over the area of about 63,000 ha, and its production is about 5 lakh tons/annum. There are various diseases that can affect pomegranate plant such as Bacterial Blight affected by bacteria, Anthracnose, Alterneria and Cercosphora affected by fungi. Hence, the current work is aimed to develop an automated system that detects the plant diseases accurately without human intervention by using various image processing and machine learning techniques.

2 Methodology

As diseases are inevitable in plants, early detection and diagnosis of diseases is a crucial aspect in the field of agriculture. This can be achieved using an automated image processing system in which the following steps have to be undertaken(Figure 1).

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 715–720.
springerlink.com

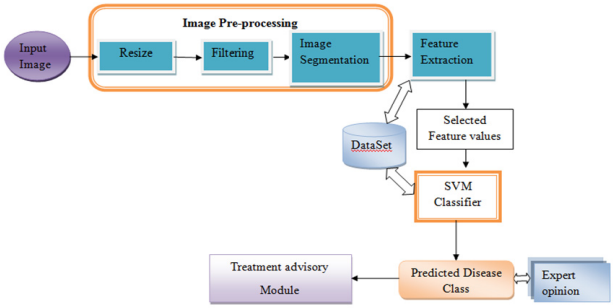


Fig. 1. Block Diagram

2.1 Image Acquisition

In the current work, a total of 143 images are captured(88 healthy & 55 diseased samples) using Nikon Coolpix L20 digital camera having 10 megapixels of resolution and 3.6x optical zoom, maintaining an equal distance of 16 cm to the leaf object. All the images are saved in the common JPEG format.

2.2 Image Pre-processing

Image pre-processing is the improvement of digital image quality without knowledge about source of degradation.

In the present work, the captured images are resized to a fixed resolution so as to reduce the computational burden in the later processing. Resized images are then filtered using Gaussian filter to get noise free images.

Image segmentation refers to the process of partitioning the digital image into multiple segments. The level to which the partitioning is carried depends on the problem being solved i.e. segmentation should stop when the objects of interest in an application have been isolated [3]. There are various techniques for image segmentation such as clustering, compression-based methods, histogram-based methods, region growing etc.

In the present work K-means clustering algorithm is employed for color image segmentation. K-Means Clustering is a method of cluster analysis which aims to partition n observations into k mutually exclusive clusters in which each observation belongs to the cluster with the nearest mean and it returns the index of the cluster to which it has assigned each observation.

In color based image segmentation, initially, original image is converted from RGB color space to L*a*b* color space. *L*a*b** color space enables to quantify the visual differences. It aspires to perceptual uniformity, and it's *L** component closely matches human perception of lightness. Classification of colors takes place in a*b* space using K-means clustering. It labels every pixel in the image using the results from k-means to form the cluster indexed image. Finally the original image is clustered into multiple segments based on color.

2.3 Feature Extraction

For an image, the desirable property for a feature detector is *repeatability*; i.e., whether or not the same feature will be detected in different images. In image processing, image features usually include color, shape and texture features. In the present work, texture and color features are considered. These features are used as characteristic values to identify whether the leaf is healthy or diseased sample.

Following are the three major texture features extracted from 143 images for the purpose of classification.

Contrast: A measure of the intensity contrast between a pixel and its neighbor over the whole image.

Energy: Returns the sum of squared elements in the Gray Level Co-occurrence Matrix [GLCM]. Energy is 1 for a constant image.

Entropy: A statistical measure of randomness that can be used to characterize the texture of the input image.

Following are the three color features which are considered for all 143 samples used in the present work.

Hue: A color attribute that describes a pure color (pure yellow, orange or red). As hue varies from 0 to 1.0, the corresponding colors vary from red through yellow, green, cyan, blue, magenta, and back to red, so that there are actually red values both at 0 and 1.0.

Saturation: A measure of the degree to which a pure color is diluted by white light. As saturation varies from 0 to 1.0, the corresponding colors (hues) vary from unsaturated (shades of gray) to fully saturated (no white component).

Value: Brightness value is one of the key factors in describing color sensation. It embodies the achromatic notion of intensity. As value varies from 0 to 1.0, the corresponding colors become increasingly brighter.

From the comparison of texture features for both diseased and healthy image samples (Figure 2a), we can observe that except contrast, other two features have distinction between diseased and healthy samples. Hence, we can retain these two texture features for the purpose of classification.

From the comparison of color features for both diseased and healthy image samples (Figure 2b), we can observe that all color features have a distinction line between diseased and healthy samples. Hence, we can retain all the 3 color features for the purpose of comparison.

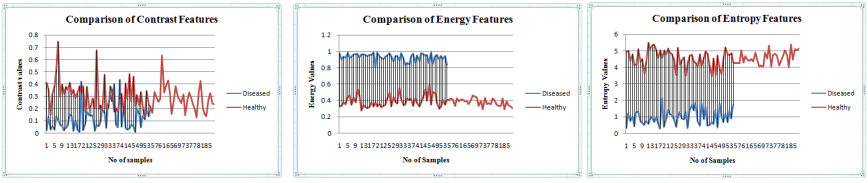


Fig. 2a. Comparison of Texture features for Diseased and Healthy samples

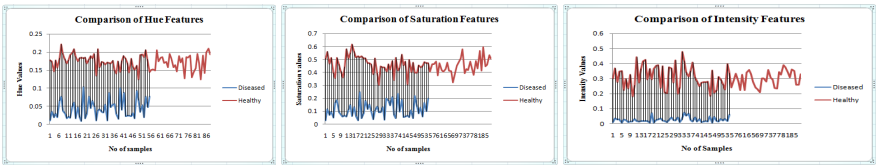


Fig. 2b. Comparison of Color features for Diseased and Healthy samples

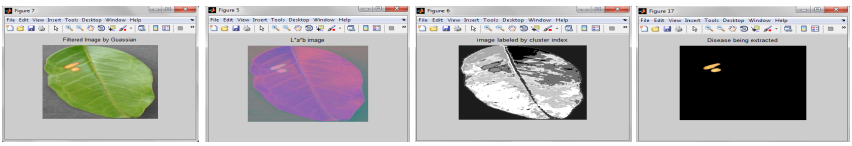
2.4 Classification

In the present work, Support Vector machines (SVM) are used for the purpose of classification as they have the very advantages of high-dimensionality and non-linear capabilities. Support vector machines, developed by Vapnik, are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other.

In the present work, SVM is trained using texture and color features of 88 healthy and 55 diseased leaf samples. SVM chooses an optimal hyperplane based on training data and classifies the query image as either healthy or diseased. In this work, two feature sets are considered i) entropy and saturation feature set and ii) energy and hue feature set. Then classification is performed based on each of these feature sets. Finally performance properties are found and compared to get which feature set provides more classification accuracy.

3 Experimental Results and Analysis

The work begins with capturing the images of healthy and diseased leaves. Query image undergoes several pre-processing steps- image resize, filtering (Figure 3a, 3b, 3c) and segmentation. Selection of a particular segment/region of interest (Figure 3d) from the segmentation result depends on the mean_clustr_value which is the output parameter of k-means.



a. b. c. d.

Fig. 3a. Resized & Filtered Image b. LAB Image c. Cluster Index Image d. Diseased portion

Once the infected region/diseased spots are extracted, next step is to extract the color and texture features out of it (Figure 4).

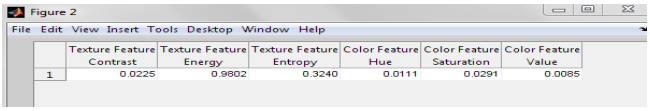


Fig. 4. Color and texture features extracted from the region of interest of the query image

Experimental results yield that, the classification accuracy of SVM is 95.95% for energy and hue feature set (Figure 5a), and 97.30% for entropy and saturation feature set (Figure 5b). Hence entropy and saturation feature set is retained for classification of diseased and healthy leaves (Figure 6).

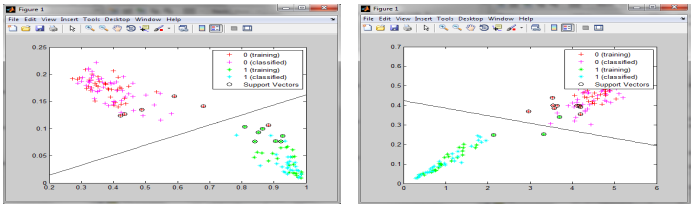


Fig. 5a. SVM classifier for energy & hue feature set Fig. 5b. SVM classifier for entropy & saturation feature set.

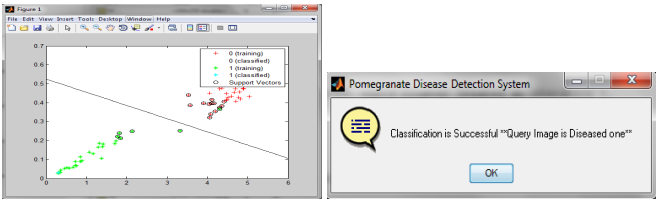


Fig. 6. Result of SVM Classifier for query Image

4 Conclusion

The current work succeeded in developing a quick, automatic and accurate system for disease identification for pomegranate leaves. The potency of the system is the capability to extract the diseased portion in the query images. The system employs diverse image processing techniques and Support vector to categorize the query leaf image as either healthy or diseased sample. Implementation is carried out using MATLAB's Image Processing and Bioinformatics Tool Boxes. The results showed that SVM could effectively detect the disease spots and classify the given leaf image appropriately to an accuracy of 97.30%. Thus the system can be satisfactorily used for plant disease classification which ultimately helps agriculturists/farmers. Further, treatment advisory module can be prepared for different diseases by seeking advice from the agricultural experts.

Acknowledgments. We thank Visvesvaraya Technological University, Belgaum, Karnataka, India for funding this project and providing a platform for Research and Development.

References

1. Camargo, A., Smith, J.S.: An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosystems Engineering, 9–21 (2009)
2. Dheeb, A.B., Braik, M., Sulieman, B.-A.: Detection and Classification of Leaf Diseases using K-means-based Segmentation and Neural-Networks-based Classification. Information Technology Journal 10(2), 267–275 (2011)
3. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing, 2nd edn. Pearson Education
4. Image Processing Toolbox™7 User's Guide, ©Copyright by The MathWorks Inc. (1993–2010)
5. Meunkaewjinda, A., Kumsawat, P., Srikaew, A.: Grape leaf disease detection from color imagery using hybrid intelligent system. In: IEEE 5th International Conference ECTICON, vol. 1, pp. 513–516 (2008)
6. Mustafa, N.B.A., Syed, K.A., Zaipatimah, A., Yit, W.B., Aidil, A.Z.A., Zainul, A.M.S.: Agricultural Produce Sorting and Grading using Support Vector Machines and Fuzzy Logic. In: IEEE International Conference on Signal and Image Processing Applications, pp. 391–396 (2009)
7. Singh, A.K.: Precision Farming. Water Technology Centre, I. A. R. I., New Delhi
8. Tellaeche, A., Xavier, P., Burgos, A., Pajares, G., Ribeiro, A.: A Vision-based Classifier in Precision Agriculture Combining Bayes and Support Vector Machines
9. Tian, Y., Tianlai, L., Niu, Y.: The Recognition of Cucumber Disease Based on Image Processing and Support Vector Machine. In: CISP, vol. 2 (2008)
10. Qing, Y., Zexin, G., Yingfeng, Z., Jian, T., Yang, H., Baojun, Y.: Application of Support Vector Machine for Detecting Rice Diseases Using Shape and Color Texture Features. In: International Conference on Engineering Computation, pp. 79–83 (2009)

Encrypted Traffic and IPsec Challenges for Intrusion Detection System

Manish Kumar[1], M. Hanumanthappa[2], and T.V. Suresh Kumar[1]

[1] Dept. of MCA,
M S Ramaiah Institute of Technology,
MSRIT Post, Bangalore-560 054, India
manishkumarjsr@yahoo.com
hod_mca@msrit.edu
[2] Dept. of Computer Science and Applications,
Jnana Bharathi Campus,
Bangalore University,
Bangalore -560 056, India
hanu6572@hotmail.com

Abstract. Now a day IPsec has now become a standard information security technology throughout the Network and Internet society. It provides confidentiality, authentication, integrity, secure key exchange and protection mechanism though encrypting a packet. The use of IPsec, which encrypts network traffic, renders network intrusion detection, virtually useless, unless traffic is decrypted at network layer. In this paper we are discussing that how a IPSec or other encryption techniques create challenges for Intrusion Detection System.

1 Introduction to Internet Protocol Security (IPsec)

The use of IPsec, which encrypts network traffic, renders network intrusion detection, virtually useless, unless traffic is decrypted at network layer. The alternative to NIDSs, host-based intrusion detection systems (HIDSs), provide some of the functionality of NIDSs but with some limitations. HIDSs cannot perform a network-wide analysis and can be subverted if a host is compromised.

IPsec can be used in either of two modes: transport or tunnel as shown in Fig 1. With transport mode IPsec, the IP payload is encrypted and the IP header is left unencrypted. If transport mode IPsec is used, other parties on the network can see the source and destination addresses of packets but no other information. With tunnel mode IPsec, the entire packet is encrypted and a new IP header is added to the packet. If tunnel mode IPsec is used, the original source and destination addresses of the packet are hidden as well, providing network monitors with even less information. The secrecy provided by encrypting traffic with IPsec acts as a dual-edged sword. Malicious parties can no longer eavesdrop on network traffic. However, encryption hides the majority of traffic content from any intrusion detection system monitoring traffic in the network.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 721–727.
springerlink.com

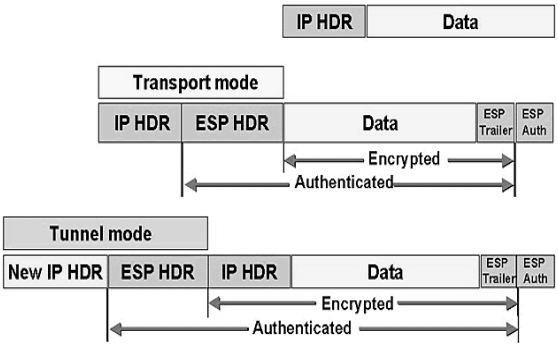


Fig. 1. ESP in Transport and Tunnel Mode

2 IPSec Challenges for IDS

IPSec configuration was shown to be exposed to variety of "initialization vector attacks"[5]. The authors of [5] have analyzed the security risks posed on encryption-only IPSec when it is used as intermediate layer of the protocol stack. But their analysis has a limited scope - just the implication on the next (possible) upper-layer protocol (e.g. TCP, UDP, IP). The authors of [7] have exposed even more serious weakness of the encryption-only configuration on IPSec. It has been shown that the "encryption-only" configuration is subject to a variety of ciphertext-only attacks. The attack consists of two phases: in the first phase the attacker modifies certain fields in the first few blocks of the ciphertext The second phase proceeds with further recovery of the plaintext. The attacks presented in [7] are efficient and have been proven practical by the authors, i.e. against an implementation of the IPSec stack in the Linux kernel.

3 DoS Attack Scenario on IPsec

3.1 Internet Checksum

The (internet) checksum [3,2,11] is used to discover errors while datagrams are transmitted. The purpose of the internet checksum is to provide an efficient protection against transmission errors but not to provide cryptographic integrity protection of the content.

The way this checksum is computed allows the fast incremental update of the checksum when the data over which it is computed is changed. For example, to update the checksum C in the IP header m to the new IP header m', but without recomputing it over the entire new IP header, the updated checksum C' can be easily computed as $C' = C + (m - m')$, as shown in [1,15]. Note that, to compute the new checksum, one needs to know $m - m'$, but neither m, m' nor the old checksum C are to be known.

3.2 Fields Manipulation

The choice made by some implementers or users for the "encryption-only" ESP configuration (i.e. without the "costly" authenticated integrity) is based on the false belief that confidentiality protection together with the structure of IP, TCP (or IP, UDP, RTP) datagrams is enough to detect any data change (malicious or not). This amounts at making the hypothesis that any modification to the encrypted datagrams will be detected during the parsing of the decrypted datagrams (i.e. they are supposedly malformed) or during the verification of the checksums.

This is not true; an attacker can as claimed by Ventzislav Nikov [16] , in fact very easily, modify the encrypted datagrams in such a way that they are still acceptable for the embedded protocols such as IP, TCP, UDP.

Consider a confidentiality protection with a block cipher with length n = 128 bits (Fig. 2). In this case, P_1, the first block of data, contains the IP header checksum, source IP address and destination IP address. Now an attacker can change the source IP address to whatever she/he wants, then using the difference between the old and the new source IP addresses she/he can compute the new correct checksum. When a gateway receives such a datagram, it accepts it as a valid datagram (i.e. the IP header is valid) and forwards it further to the corresponding destination address.

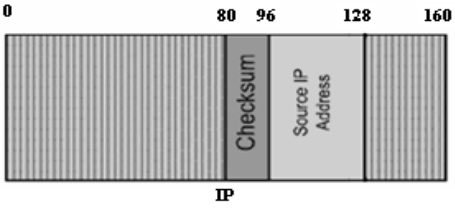


Fig. 2. Block Cipher Lebgth of 128 bits

3.3 Attack Scenario and Preventive Measures

Assume a source device 'X' communicates with a destination device 'Y' through IPSec gateways using "encryption-only" IPSec (ESP). Consider the n = 128 bits case (Fig 2), an attacker can intercept a legitimate ESP datagram, modify the source address, such that the altered ESP datagram and the embedded IP header are still valid and re-inject the new ESP datagram in the network.

Gateway 'Y' will accept this modified ESP datagram, extract the IP header and embedded payload and forward them to the destination device. Notice that, for any legitimate ESP datagram, the attacker can generate up to 2^{32} false source addresses and hence so many false ESP datagrams, which will have the same content as the legitimate one but will claim to come from different sources. All of them will be accepted by the gateway 'Y' and forwarded to the destination device. In this way, the gateway as well as the destination device are overloaded with junk datagrams. The attacker can mount this attack on the fly in real-time. If we assume (as in [7]) that the attacker knows several source IP addresses of machines which legitimately communicate through the IPSec gateway, then even stateful firewall could not defeat the attack

Several researchers have proposed selective encryption schemes for providing network-based services, such as a NIDS, access to data encrypted by IPsec [9,14,17,19]. This approach uses selective encryption on an IP packet by breaking the packet into multiple encryption zones. This technique was proposed in 1999, largely for improving network performance, independently as Multilayer IPsec (ML-IPsec [16,19]) and Layered Encryption Security (LES [9]). Each encryption zone is assigned a cryptographic key and a portion.

Cynthia McLain et.al.[6] has suggested "Two-Key IPsec", which restrict the number of encryption zones to two: one for the header information at the beginning of the packet and the other for the remainder of the packet, containing the application data. End hosts have access to the entire packet.

Figure 3 illustrates the difference between traditional and Two-Key IPsec [6]. In the top two packet diagrams, the hashed area represents the data that traditional IPsec encrypts using a single key, K_1. Only the end points of the connection know K_1. In the bottom two packet diagrams of Figure 3, the two hashed areas show the effect of using Two-Key IPsec to encrypt the bytes of the packet with two separate keys, K_1 and K_2. The first N bytes of the ESP payload are encrypted with the second key, K_2. Note that for IPsec tunnel mode, N′ bytes are encrypted using K_2, where $N' = N + L_{IP\_Hdr}$ and L_{IP_Hdr} is the length of the encapsulated IP header. K_2 is shared among the end hosts and the network-based service. K_1 is used to encrypt the remainder of the ESP payload and, as with traditional IPsec, is known only to the end points. In both cases, IPsec adds the ESP header and appends an integrity check value, or authentication data (ESP Auth).

The Two-key IPsec can reduce the DoS attack up to some extent but since the partial packet need to be opened at network layer poses the security threats.

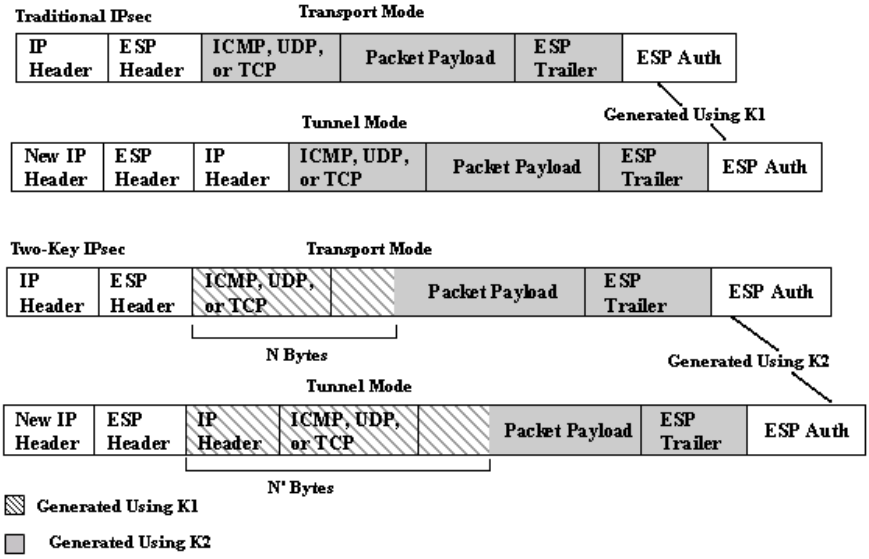


Fig. 3. Difference of Traditonal and Two-Key IPsec in Transport & Tunnel Modes

4 Other DoS Attacks on IPsec

In this work we considerd DoS attacks, that can eavesdrop and spoof packets, yet even weaker, blind spoofing, attacker can mount a DoS on IPsec. For instance, it is known that fragmentation can expose IPsec to DoS attacks, e.g., IPsec cannot prevent attacks on fragments' buffer at the recipient if fragmentation is allowed. Specifically, since authentication is performed prior to fragmentation, spoofing attacker could launch a DoS attack by swamping the receiving gateway with (maliciously crafted) IP fragments, which could not be reassembled, thus legitimate packets could not be accepted, e.g., in [2,5,15,18]. This attack is made possible due to the fact that IPsec reassembles the fragments prior to authenticating them, and the attack can be prevented by defining minimal fragment size and not allowing fragmentation; another solution is to only allow prefragmentation, i.e., fragmentation by IPsec gateway prior to applying IPsec processing on the outgoing packet.

DoS attacks can also be launched on IKE (key establishment protocol of IPsec), which was designed to run over UDP in order to avoid DoS attacks on TCP. In [5], the authors show an attack on IKE, by exploiting fragmentation.

A vulnerability of IPsec to DoS when using Explicit Congestion Notification (ECN) is investigated in [7]. If the IPsec gateway at the exit of the tunnel does not copy the ECN bit, then it ruins the ECN mechanism; on the other hand, if the gateway copies the ECN bit, then an attacker can degrade performance. The attack can be launched since the authentication that IPsec performs does not protect the ECN bit. However, there is no analysis of this attack; such analysis is rather similar to the analysis presented, of similar attacks.

Tunneling using IPSec protocols creates a new risk: It allows the encapsulation of IPv6 traffic in an IPv4 data stream for routing through non-complaint devices.It includes leading threats of IPv4 DoS and DDoS attack.

5 Conclusions

In this paper we discussed about the DoS attack against the IPSec. Secure channel protocols, with IPsec being the predominant one, are used to securely connect virtual private networks (VPN), i.e., authenticate data and origin, ensure confidentiality, and performance. IPsec is designed to protect against man-in-the-middle (MITM) adversaries that can eavesdrop on the communication and inject spoofed segments into the message stream. It is widely believed, and also specified e.g., in [12], that IPsec also defends higher layer traffic from DoS attacks when attacker has limited resources (e.g., can only block, inject or reorder a limited number of packets). Defense against DoS attacks is often an important consideration in adopting IPsec for protecting a VPN. We saw that this belief is not precise and that IPsec does not deliver on its performance guarantees, by presenting several DoS attacks on TCP when used over IPsec.

System administrators commonly use NIDSs to protect their networks. A common concern is that more widespread use of IPsec for encrypting network traffic will render NIDSs virtually useless. Common solutions rely on decrypting the entire packet at the NIDS or placing IDSs on the end hosts. The first technique can have a negative impact on privacy while the second technique can have a negative impact on security.

Under Two-Key IPsec [6], an attacker who compromises the NIDS can read only the IP header,ICMP, TCP, or UDP header, and the small portion of the packet encrypted under the shared key.Note that the ability of the attacker to modify or create new packets is dependent on how the encrypted portions of the packet are authenticated. If the encrypted portions of the packet are authenticated from end-to-end, it will not be possible for the attacker to successfully modify the packet or create new packets. If the separately encrypted portions of the packet are authenticated separately, it may be possible for the attacker to modify a portion of the packet, or even create new packets without being detected.

In this paper we discussed about a DoS attack against encrypted traffic using IPsec. We belief that this attack will serve as a strong argument to improve the existing standard.

Acknowledgments. I would like to thank MSRIT Management, my colleagues and Dept. of Computer Science and Applications, Bangalore University, for their valuable suggestion, constant support and encouragement.

References

1. A DoS Attack Against the Integrity-Less ESP (IPSec):- A DoS Attack Against the Integrity-Less ESP (IPSec), Ventzislav Nikov
2. Studer, A., McLain, C., Lippmann, R.: Tuning Intrusion Detection to Work with a Two Encryption Key Version of IPsec. MIT Lincoln Laboratory, Carnegie Mellon University, Lexington, Pittsburgh, PA
3. Herzberg, A., Bar, H.S.: Stealth DoS Attacks on Secure Channels, Ilan University Department of Computer Science, Ramat Gan, 52900, Israel
4. Kaufman, C., Perlman, R., Sommerfeld, B.: DoS protection for UDP-based protocols. In: Proceedings of the 10th ACM Conference on Computer and Communications Security, p. 7. ACM (2003)
5. McCubbin, C., Selcuk, A., Sidhu, D.: Initialization vector attacks on the IPSec protocol suite. In: WETICE 2000, pp. 171–175. IEEE Computer Society (2000)
6. McLain, C., Studer, A., Lippmann, R.: Making Network Intrusion Detection Work with IPsec, March 9 (2007)
7. Paterson, K.G., Yau, A.K.L.: Cryptography in Theory and Practice: The Case of Encryption in IPsec. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 12–29. Springer, Heidelberg (2006)
8. Ramakrishnan, K., Floyd, S., Black, D.: The addition of explicit congestion notification (ECN) to IP (2001)
9. Roesch, M.: Snort: Lightweight intrusion detection for networks. In: Proceedings of the 13th Conference on Computer and Communication Security (LISA 1999), pp. 229–238 (November 1999)
10. Karir, M.: IPSEC and the internet. Master's Dissertation, University of Maryland, Department of Electrical Engineering (1999)
11. Braden, R., Borman, D., Partridge, C.: Computing the Internet Checksum, RFC 1071 (September 1988)
12. Kent, S., Seo, K.: Security architecture for the Internet Protocol. Internet Engineering Task Force, RFC 4301 (December 2005), `http://www.rfc-editor.org/rfc/rfc4301.txt`

13. Kent, S., Seo, K.: Security Architecture for the Internet Protocol. RFC 4301 (Proposed Standard) (December 2005)
14. Kasera, S.K., Mizikovsky, S., Sundaram, G.S., Woo, T.Y.C.: On securely enabling intermediary-based services and performance enhancements for wireless mobile users. In: Workshop on Wireless Security, pp. 61–68 (2003)
15. Mallory, T., Kullberg, A.: Incremental Updating of the Internet Checksum, RFC 1141 (January 1990)
16. Nikov, V.: A DoS Attack Against the Integrity-Less ESP (IPSec). Philips TASS and App-Tech, Leuven, Belgium
17. Cheswick, W.R., Bellovin, S.M., Rubin, A.D.: Firewalls and Internet Security, 2nd edn. Repelling the Wily Hacker, pages 10,281. Addison-Wesley (2003)
18. Gilad, Y., Herzberg, A.: Lightweight Opportunistic Tunneling (LOT). In: Backes, M., Ning, P. (eds.) ESORICS 2009. LNCS, vol. 5789, pp. 104–119. Springer, Heidelberg (2009)
19. Zhang, Y.: Multi-layer protection scheme for IPsec. IETF Internet Draft, IETF (1999), http://tools.ietf.org/html/draft-zhang-ipsec-mlipsec-00

An Effective User Interface Tool for Retrieval of Heart Sound and Murmurs

Kiran Kumari Patil, B.S. Nagabhushana, and Vijaya Kumar B.P.

Reva Institute of Technology and Management
Kattegenhalli Yelahanka, Bangalore
kirankumari@revainstitution.org

Abstract. The phonocardiogram (PCG) is an important biomedical signal of the heart of audio nature (heart sounds and murmurs) related to the contractile activity of the cardiohemic system and represents a recording of the heart sound signal. The audio retrieval problems are studied in audio information retrieval (AIR) or music information retrieval (MIR) systems and are modeled as feature vectors and employ the similarity measures for speech or music retrieval. We extend these content-based retrieval techniques exclusively for heart sounds and murmurs. In this paper, we propose a framework for audio modeling of heart sounds and murmurs using feature vectors (spectral, and perceptual) and implementation of content based heart sound and murmurs retrieval algorithms and auditory user interfaces for cardiologist, in which he/she can directly audio query and obtain the ranked heart and murmur audio files using similarity measures. The query results are displayed in a heart sound and murmur browser, where cardiologists not only visualize (temporal and frequency domain) the phonocardiography signals, but also listen and make effective clinical decisions. The preliminary results of the research work show 80% precision and good retrieval efficiency.

1 Introduction

The heart sound signal is the most traditional biomedical signal and stethoscope is used by physicians for clinical investigations. Heart sounds and murmurs are acoustic phenomenon caused by mechanical events of the heart. Auscultation or hearing of heart sounds and its interpretation by conventional stethoscope or electronic stethoscope is not purely a mechanical phenomenon of sound wave propagation but also auditory, sensory, cognitive and perceptual event for the cardiologists. The salient characteristics or features of typical heart sounds and murmurs are described in Table 1. The phonocardiography (PCG) – the art and science of recording, listening and interpreting heart a sound using latest digital technology has significantly helped in understanding and interpreting complex heart sounds and murmurs. The PCG techniques are used for the effective clinical investigations and corrective diagnostic heart related diseases and in particular valvular heart diseases. In the early days of phonocardiography, the heart sounds were printed on a graph paper and supported visual inspections. It mainly focused on timing, amplitude and turbulent sounds (murmurs), systole/diastole duration etc. characteristics [3] and not amenable to auditory perception. With latest development biomedical instrumentation, it is

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 729–737.
springerlink.com

possible to store heart sounds in digital format and perform analysis using digital processing techniques. A majority of work mainly focused on applying digital processing techniques – time and frequency techniques such as FFT, spectral analysis, wavelet analysis etc. for analysis, segmentation, classification and interpretations of heart sounds and murmurs [3]. From digital signal processing (DSP) perspective, the heart sounds and murmurs are complex, dynamic, non-stationary, time – series physiological data and more amenable biomedical signal analysis and DSP techniques (time and frequency domain) and extensive work has been carried out [1]and significantly contributed to the hearts sound analysis and murmur studies. Computer-aided auscultation techniques provide accurate and objective interpretation of the heart sounds for clinical diagnosis. Several algorithms for automatic analysis of heart sounds [1, 2] and detection of coronary heart diseases from murmur analysis are effective in coronary surgery management. In our research, we model and look from a different perspective – the heart sounds and murmurs are audio signals and amenable to audio perception during auscultation. We are more interested in audio retrieval the hearts sounds that matches the user interface requirements (perceptual model) based on psychoacoustic principles for the cardiologists. The task is further complicated by the subjective nature of the interpretation of the heart sound and murmurs and we also provide quantitative and qualitative reasoning framework with visual inspection The traditional database consisting of textual data is relatively easy to process when compared with multimedia content and real-time biomedical signals. The text processing techniques and query processing algorithms are inadequate due to the inherent nature of the multimedia content (text, images, video, and audio) and complex data types. With the recent developments in digital signal processing (DSP) and multimedia databases and it is now feasible to integrate and store heart sounds and murmurs in digital form and audio streams along with textual and semistructured (XML) formats in clinical information systems. Recent developments Audio Information Retrieval (AIR) [4] and Music Information Retrieval (MIR) [5] are successfully implemented in research prototypes [12] and commercial applications [13]. These are content-based retrieval techniques based on spectral or audio properties are exploit and apply pattern recognition techniques using feature vectors and are used for speech recognition, music understanding [12], instrument classification, genre classification [12], MIDI and instrument recognition and animal sounds [12]. Our work [17, 18] and in this paper, mainly deals with hearts sounds and murmurs and focus on content-based retrieval of heart sounds and murmurs are characterized complex psychoacoustics (e.g., pitch, loudness etc.), spectral content (frequency, spectrum, spectral centroid etc.) and at the signal level, they are highly complex real time, non-stationary and fall below the threshold of the human audio perception. In comparison with speech and music, the heart sounds and murmurs are difficult from content-based retrieval perspective and we are extending content-based retrieval algorithms and similarity measures. In this paper, we also focus on visualization in time and frequency domain, visual rendering of hearts sounds and murmurs of their MFCC coefficients and auditory user interfaces in which, the cardiologists can submit an audio query and system retrieve the ranked list of best matched heart and murmurs files. It also support qualitative and quantitative assessment for cardiologists to avoid subjective interpretation during auscultation.

The paper is arranged as follows. In section 2, we discuss the audio modeling framework for heart sounds and murmurs and research related to AIR and MIR. In section 3, we propose and show the use of this framework in modeling and analysis and its use in user interface or audio browser design. In section 4, we model the heart sound and murmurs as a audio retrieval problem and discuss the salient features audio retrieval algorithms with clinical heart sound and murmur database. In section 5, we discuss the implementation architecture and experimental setup and audio information retrieval results and derive performance results. The section 6 concludes with preliminary research results and future works.

2 Content Based Audio Retrieval and Similarity Measures

In this section, we discuss research developments in music information retrieval (MIR) and Content-Based Audio Indexing and Retrieval (CBAIR) and techniques such a Query By Example (QBE) and Query by Humming (QBH) in MIR and CBAIR systems. In general, the research in MIR and CBAIR discusses about the different methods to represent musical objects, such as feature-based representation, musical parameter-based representation; similarly based retrieval strategies, melody or theme based retrieval of musical objects. Content-Based retrieval systems accept data type queries i.e. drawing sketch for an image retrieval; or clip of a song for a song retrieval; and a video clip or set of images from some video short for a video retrieval. Content-Based retrieval allows users to describe the query as what they want, so it makes query formulation more comprehensive and easier than key word based retrieval [7, 14] and similarity search based on approximate matching produce batter results compare to exact matching. For such systems, compact and more comprehensive music representation along with more efficient indexing structures and retrieval strategies would be main consideration. Music is represented written score or music notations, recorded performance and MIDI (Musical Instrument Digital Interface) format and are stored in the databases. The listener can query using perceived parameter viz., melody and melody extraction algorithms are used on the intensity for each note and chord that represent melody. In general, the content-based retrieval techniques for music are successfully in many MIR [6] and CBAIR systems [4]. In recent years many retrieval systems [5, 15] implemented content-based retrieval for music and instruments. The representation of an audio object by a template that characterizes the object using feature vectors and uses a template in which an audio signal is first divided into overlapping frames of constant length then using digital signal processing techniques, for each frame a13-dimensional feature vector is extracted (12 Mel- Frequency Cepstral Coefficients (MFCC) plus Energy) at a 500Hz, and then these feature vectors are used to generate templates using tree based Vector Quantizer trained to maximize mutual information (MMI) [11]. For retrieval, query is first converted in to template in the same way described earlier then for its similarity search template matching is applied which uses distance measure, and finally a ranked list is generated based on minimum distance. In this system performance of the system with Euclidean distance as well as Cosine distance, is also compared, and experimental results show that cosine distance performs slightly better than Euclidean distance. The internal

indexing structure is B-tree and uses well known clustering techniques and similarity search algorithms [12]. Another approach [12] is to represent an audio object is characterized by its frame level and global acoustical and perceptual parameters. These features are extracted at frame level using signal processing techniques and globally using statistical analysis based on frame level features and using musical analysis. Frame level features consist of loudness, pitch and MFCCs and analyzed by using histogram modeling techniques [22]. For musical objects, musical features (i.e. rhythm, events and distance (interval)) are extracted using simple signal processing techniques like pitch tracking, voiced and unvoiced segmentation and note formation are used for indexing and similarity measures.

We use the similar techniques and concepts; however, we differ in the following ways:

1. The audio objects are heart sounds and murmurs which are characterized by time- series data, band-limited (10-400 Hz) and highly non-stationary.
2. We develop the psychoacoustic model of the heart sounds and murmurs for audio query processing which are more intuitive for cardiologists.
3. The subjective assessment of the doctors by auscultation is supported by both qualitative and quantitative reasoning by providing visual inspection of the PCG, visual rendering of MFCC coefficients, frequency domain techniques (spectrogram, magnitude spectrum etc.) and perceptual features (pitch, intensity, rhythm, loudness, etc.) integrated into auditory user interfaces. We address these specific retrieval challenges in which cardiologist will submit an audio query – unknown heart sound and murmur in audio browser and the system will retrieve ranked audio files with visualization with high degree of similarity search and measures. The scope of audio and it retrieval is broad and includes speech, music, instruments, natural and man-made sounds and in this paper, it refers to the heart sounds and murmurs only and excluding other biomedical signals and sounds (e.g., lung sounds).

3 Psychoacoustic Modeling of Heart Sounds and Murmurs

Psychoacoustics is the study of the subjective human perception of sounds and correlates psychological of the physical parameters of acoustics [1]. Auscultation is a highly subjective task and solely depends on the experience and listening skills of the cardiologists. The psychoacoustic research model parameters of auditory sensation in terms of physical signal parameters and provide a framework and the subjective bias is minimized and helps in visual inspection.

4 Feature Extractions and System Architecture

The block diagram of feature extraction and query processing is described in fig 1.

The audio processor and feature extraction is performed by using feature extractor tool JAudio and built-in functions of MATLAB and stored in XML format for each heart sounds and murmurs in cardio database. The heart sounds and murmurs are

acquired using commercial electronic stethoscope (e.g. Littman 400x) and stored in the raw audio format. The audio signals are of duration of 10 to 60 seconds and sampled at the 11 KHz frequency.

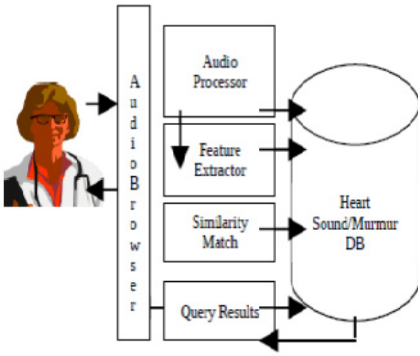


Fig. 1. Feature extraction and query processing

Table 2. Feature vectors (Temporal, spectral perceptual, harmaonic and statistical) of heart soun and murmurs

Feature	Dimension	Remarks
FFT Bin Frequency Labels	256	The bin label, in Hz, of each power spectrum or magnitude spectrum bin useful for spectrum and power spectrum.
Magnitude Spectrum	256	A measure of the strength of different frequency components and mainly concerted in the 10-40 Hz and useful for spectrum analysis.
Power Spectrum	256	A measure of the power of different frequency components and useful for spectral visualization.
Root Mean Square	1	A measure of the power of a signal and is related to the pitch and sound intensity.
Zero Crossings Overall Standard Deviation	1	The number of times the waveform changed sign. An indication of frequency as well as noisiness. Useful for murmur detection.
MFCC	13	MFCC calculations for perceptual features.
LPC Overall Standard Deviation	10	Linear Prediction Coefficients calculated using

The acquired heart sounds and murmurs are processed using feature extractor in various features vectors as shown in Table 2. The database contains the frame level feature vectors and their values as well as average features at the complete audio file. Each audio frame is overlapped with 10ms Hamming window to derive stationary signal features and values. Some of the feature vectors have 256 dimension (e.g., magnitude spectrum) or 13 dimensions (e.g., MFCC) and may lead the dimension reduction problems. The heart and murmurs sound databases is populated with about 300 heart sounds and murmurs recording from the various internet resources as well as clinical settings. The heart and murmur database consist of 30 types of hearts disease and cardiovascular pathologies. The system support temporal, spectral, perceptual, harmonic and statistical features and highly configurable and cardiologists can customize as per the clinical investigation needs. The similarity search can performed on the feature vectors and in particular we focus on MFCC features and can represent the perceptual features. It is also second order statistics and derivatives are useful for segmentation and categorization. The selected features are also used for indexing the audio database and helps in ranking to derive better performance.

5 Content Based Audio Retrieval Algorithm

Firstly, the feature vectors are extracted from the test PCG selected by the cardiologist and extract its feature vectors. On a conventional PC, it takes about 5-10 seconds for raw audio data of size (1K – 10 Kbytes) of sample of 5-10 second recordings. When the raw audio data is large (100-200 Kbytes) and takes about 15-30 seconds and

depend on the selection of the feature vector of interest. The similarity measures and correlations studies can be performed on the selected feature vectors. In particular, we can compare and contract various feature vectors of interest. For example, zero crossing ratio (ZCR) is an important feature vector that distinguishes the murmurs with normal sounds due to the high frequency noise and sign change. Feature extraction is based on a variant of the Melfrequency Ceptral coefficient (MFCC) representation. MFCCs are commonly used in speech recognition systems because they provide a concise representation of spectral and perceptual characteristics. The MFCCs provide spectral as well as perceptual features and used for song similarity search [9]. Each coefficient has a value for each frame of the heart sound and murmurs. The changes within each coefficient across the range of various PCG signal are examined here. The derivation of MFCCs involves analyzing and processing following steps.

1. Separate the audio signal into frames: The acquired audio signal is divided into small frames of duration 20- 40 ms with or without overlap. We recommend an overlap of 10% to minimize the transient frame boundaries.
2. Calculate the amplitude spectrum: By applying the standards FFT techniques derive the magnitude spectrum of each frame and identify the major frequency components.
3. Derive log spectrum: Once, we obtain the amplitude spectrum, derive the log of the of the magnitude spectrum for each frame.
4. Obtain Mel Scale: Convert the log spectrum to the Mel scale to model the human auditory perception.
5. Apply Discrete Cosine Transform (DCT): Apply the DCT on the Mel scale and derive the MFCC coefficients.

$$sim_{ab} = \cos^{-1}\left(\frac{A \cdot B}{\|A\| \, \|B\|}\right) \tag{1}$$

$$sim_{ab} = \frac{1}{2}\sum_{i=1}^{k} \frac{(A_i - B_i)^2}{A_i + B_i} \tag{2}$$

$$sim_{ab} = \sqrt{\sum_{i=1}^{k}\left(\frac{A_i}{|A|} - \frac{B_i}{|B|}\right)^2} \tag{3}$$

$$\delta = \frac{\text{number of correctly retrieved audio objec}}{\text{the number of audio objects should be retrieved}} \tag{4}$$

$$\xi = \frac{\text{the of correctly retrieved audio objects}}{\text{The number of all retrieved objects}} \tag{5}$$

$$\eta = \delta + \xi \tag{6}$$

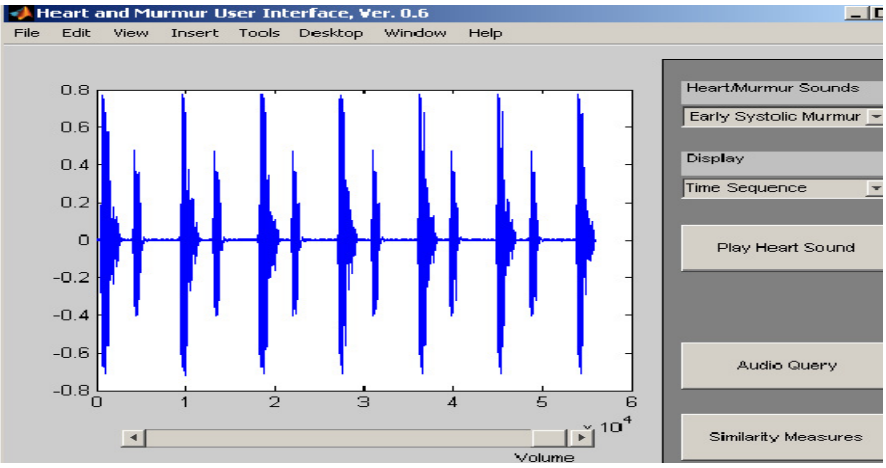


Fig. 2. Auditory user interface

The template matching is carried out to detect and locate the reference PCG from the input PCG. Retrieval strategies assign a measure of similarity between an audio query and audio file in the hearts and murmur databases. These strategies are based on simple notion that more often terms found in both the audio object and the audio query, the more relevant the audio document is seem to be the query. Mathematically, a retrieval strategy is an algorithm that takes query Q and set of documents D1, D2…Dn, and evaluates the similarity coefficient (SC) for each document in the database. Both the audio query and each audio object are represented as vectors in terms of n-dimensional space. A measure of similarity between the query and each object in the database is computed on the basic notion of the vector space model. This model shows that query is transformed into a feature vector **q**, now search engine finds the most relevant audio file for the entire database that contains documents (objects) as features vectors d1, d2…**d**n. This model involves the construction of vectors representing the salient features of objects, so similarity between two objects is determined by measuring the closeness of two vectors in space; this is done by computing distance between two vectors. Now relevance between two audio objects is to know how closely they are located in n-dimensional space and measures Euclidean distance or Cosine distance or both using equations 1,2 and 3. Histogram modeling of MFCC is also very useful parameter audio retrieval modeling. There is one limitation of this technique that is,

its retrieval complexity is linear O (n) where in the number of objects in the database; linear complexity works efficiently for small databases but for very large databases linear complexity is not good. To get rid of this problem k-mean clustering techniques can be used for better retrieval performance.

5.1 Auditory User Interface

The cardiologists prefer to interact with the system in an intuitive manner and exploit their domain knowledge of phonocardiography and cardiovascular disease. The user interface prototype (fig. 2) is implemented in Java and Matlab with SQL database to store heart sounds and murmurs in the backend database. The cardiologists can select and plot the heart sounds and murmurs as function of time i.e., time series data and can identify start of S1, end of S2, duration of diastole or systole, amplitude and related parameters. The cardiologists can perform visual inspection and easily detect the murmurs and derive murmurs characteristics. The user interface prototype also supports the visual rendering of the frequency domain parameters. For example, the spectrogram of diastolic rumble and early systolic murmur is rendered and clearly distinguish two spectral components of the two murmurs. It also characterizes each signal in frequency and pitch of the heart sounds. In particular the heart sounds and murmurs which are difficult only with auscultation. Another novel feature supported in the user interface is visual rendering of MFCC coefficients (13 dimension) and beat frequencies which clearly shows marked differences that characterize the critical bands and more useful for cardiologists. Another unique feature of the user interface is that the cardiologists can submit the audio query i.e. the unknown heart sound and murmur and system will recommend a list of ranked heart sounds and murmurs with similarity measures. The cardiologists can listen and play the heart sounds and murmurs and refine the query for resubmit them.

6 Conclusions and Future Work

In this paper has discussed a content based retrieval of heart sounds and murmurs based on psychoacoustic principles. We showed that MFCC and similarity based search algorithms are adequate for heart sound and murmur retrievals. MFCC feature vectors and various techniques of similarity measures are discussed for pattern matching of PCG signals. Search accuracy of histogram matching is tested and useful and initial results encouraging. We also observed that MFCC based histogram for content based heart sound retrieval is more efficient and accurate. The future work is refine the content based algorithms in particular ranking terms and validate the experimental results with cardiologist.

Acknowledgements. We thank Dr. S. Ravi and Dr. Cyrilraj of Dr. MGR Educational and Research Institute, Chennai for the encouragement and research guidance.

References

[1] Rangayyan, R.M., Lehner, R.J.: Phonocardiogram Signal Analysis: A Review. Critical Reviews in Biomedical Engineering 15(3), 211–236 (1988)

[2] Rangayyan, R.M.: Biomedical Signal Analysis: A Case- Study Approach. Wiley India Pvt. Ltd., New Delhi (2007)

[3] Luisada, A.A., Portuluppi, F.: The Heart Sounds – New Facts and Their Clinical Implications. Praeger, New York (1982)

[4] Foote, J.T.: Content-Based Retrieval of Music and Audio. In: Proc. SPIE, vol. 3229, pp. 138–147(1977)

[5] Foote, J.: An overview of audio information retrieval. Multimedia Systems 7, 2–10 (1999)

[6] Foote, J.: Visualizing Music and Audio using Self- Similarity. In: Proc. ACM Multimedia 1999, pp. 77–80 (1999)

[7] Ghias, A., Logan, J., Chamberlin, D., Smith, B.C.: Query By Humming. In: Proc. ACM Multimedia 1995, pp. 231–236 (1995)

[8] Lu, L., You, H., Zhang, H.J.: A New Approach to Query by Humming in Music Retrieval. In: ICME 2001, Tokyo (August 2001)

[9] Kosugi, N., Nishihara, Y., Kon'ya, S., Yamamuro, M., Kushima, K.: Let's Search for Songs by Humming! In: Proc. ACM Multimedia 1999 (Part 2), p. 194 (1999)

[10] Uitdenbogerd, A., Zobel, J.: Melodic Matching Techniques for Large Music Database. In: Proc. ACM Multimedia 1999, pp. 57–66 (1999)

[11] Yoshitaka, A., Ichikawa, T.: A Survey on Content-Based Retrieval for Multimedia Databases. IEEE Trans. Knowledge and Data Engineering 11(1), 81–93 (1999)

[12] Wold, E., Blum, T., Keislar, D., Wheaton, J.: Content- Based Classification, Search and Retrieval of Audio. IEEE Multimedia 3(3), 27–36 (1996)

[13] Blum, T., Keislar, D., Wheaton, J., Wold, E.: Audio Databases with Content-Based Retrieval. In: Intelligent Multimedia Information Retrieval, pp. 113–135. AAAI Press, Menlo Park (1997)

[14] Tzanetakis, G., Cook, P.: Audio Information Retrieval (AIR) Tools. In: International Symposium on Music Information Retrieval (2000)

[15] Veltkamp, R.C., Tanase, M., Sent, D.: Features in content-based image retrieval systems: a survey. In: State-of-the-Art in Content-Based Image and Video Retrieval, pp. 97–124 (1999)

[16] Yang, C.: Music Database Retrieval Based on Spectral Similarity. Stanford University Database Group Technical Report 2001-14 (2001)

[17] Patil, K.K., Nagabhushan, B.S., Vijay Kumar, B.P.: Psychoacoustic Models for Heart Sounds. In: Meghanathan, N., Kaushik, B.K., Nagamalai, D. (eds.) CCSIT 2011 Part II. CCIS, vol. 132, pp. 556–563. Springer, Heidelberg (2011)

[18] Patil, K.K., et al.: An Efficient Retrieval Technique for heart sounds using psychoacoustic similarity. In: IJEST (December 2010) (issue)

[19] Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Trans. Speech Audio Process 10(5), 293–302 (2002)

[20] Foote, J., Cooper, M., Nam, U.: Audio retrieval by rhythmic similarity. In: Proc. Int. Symposium on Music Information Retrieval (ISMIR), pp. 265–266 (2002)

[21] Zwicker, E., Fastl, H.: Psychoacoustics: Facts and Models. Springer, Berlin (1999)

[22] Logan, B.: Mel Frequency Cepstral Coefficients for Music Modelling. In: ISMIR (2000)

DCell-IP: DCell Emboldened with IP Address Hierarchy for Efficient Routing

A.R. Ashok Kumar, S.V. Rao, Diganta Goswami, and Ganesh Sahukari

Computer Science and Engineering Department, IIT Guwahati
{r.ashok,svrao,dgoswami,sahukari}@iitg.ac.in

Abstract. The increasing complexity and sophistication of data center applications demands new features in the data center designs. With the deployment of large range of applications, there is a demand for low latency and high throughput from the underlying network infrastructure. This resulted in many Data Center Network (DCN) designs. There are class of designs for DCN called server-centric-networks, where routing is done by server rather than intermediate switches. BCube and DCell are two such designs. These designs use source routing for communication. The symmetry and hierarchy in their design enable source to compute the entire path to the destination. The current implementation of DCell uses addresses other than IP address for addressing the servers and switches. In this paper, we provide a new implementation for DCell called DCell-IP using IP address hierarchy. Since IP addresses are also symmetric and hierarchical, we eliminate the need for using new scheme for addressing servers and intermediate switches. This also eliminates the need for translating design specific addresses to IP addresses during the routing.

1 Introduction

Applications deployed on DCN are often bandwidth intensive. This resulted in many designs providing sufficient bandwidth at all the levels [1, 2, 3, 4]. These architectures are either variant of multi-rooted tree or hyper cubes. DCell [5] and BCube [6] use multiport servers along with COTS mini switches to support large number of servers. In these designs routing is performed by servers. Fat-tree [7] is an example for multi-rooted tree-like design. Large DCN are build with these designs using only commodity switches. Further, these designs address shortcoming of conventional tree-like design [8]. Conventional tree- like design introduces oversubscription at higher levels in order to reduce the total cost.

DCell is a server-centric design where routing is done by servers rather than switches. In DCell, the source constructs the path to the destination and uses this path for source routing. Problem with the current implementation of DCell is – it uses the addresses other than IP address for addressing the servers and the intermediate switches. Further, during the actual transfer, these addresses are changed to IP address as data transfer uses Internet standards for communication.

In this paper, we propose new implementation for DCell called DCell-IP using IP address hierarchy to eliminate the need for new addressing scheme without changing the original design. Our design DCell-IP eliminates the mapping of DCell address to

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 739–746.
springerlink.com

IP address, which in turn eliminates the need for changing the packet content. This resulted in improving the performance of routing. Moreover, DCell-IP reduces number of intermediate switches and connections.

We demonstrate using our new implementation, construction of DCell-IP using only IP addresses and directly using source routing feature of IP for routing. Also, we are not modifying the any packet content during packet transfer. The rest of the paper is organized as follows. The section 2, original design of DCell is explained briefly. Proposed implementation for DCell is given in the section 3. Comparison of proposed implementation with the existing one is done in the section 4. Details of experiments conducted and results obtained are given in the section 5. Finally, we conclude with the section 6.

2 Original Implementation

In this section, we give brief outline of the design DCell which solves inherent problem with the IP and Ethernet with respect to scaling. DCell is a recursive defined structure. $DCell_0$ is the basic building block for constructing larger $DCell_k$.

$DCell_0$ is consisting of n servers and a mini switch. All the servers of $DCell_0$ are connected to a mini switch. $DCell_1$ is constructed from $n+1$ $DCell_0$. In $DCell_1$, each $DCell_0$ is connected to all other $DCell_0$ with one link for each $DCell_0$. Construction of $DCell_1$ from $n+1=5$ $DCell_0$s is showed in the Fig-2(A). $DCell_1$ connects 5 $DCell_0$s as follows. Each server is assigned a tuple $[a_1, a_0]$, where a_1 and a_0 are the $level_1$ and $level_0$ IDs respectively in the range $[0,5]$ and $[0,4]$. Two servers with addresses $[i, j-1]$ and $[j, i]$ are connected with the link for every i and j > i. In a similar fashion, $DCell_k$ are constructed from lower $DCell_{k-1}$s. Hierarchy of DCell is defined in terms of g_k, defining number of $DCell_{k-1}$s in $DCell_k$ and t_k, defining number of servers in $DCell_k$, where $g_k = t_{k-1} + 1$ and $t_k = g_k * t_{k-1}$

DCell uses a simple and efficient source routing for data transfer. For two nodes *src* and *dst* in the same $DCell_k$ but on different $DCell_{k-1}$, intermediate link (n_1, n_2) connecting these two $DCell_{k-1}$ is determined. Thus the path will be from *src* to n_1 and n_2 to *dst*. Again, the paths from *src* to n_1 and n_2 to *dst* are determined recursively in the similar fashion.

3 DCell Using IP Address

The traffic analysis for DCN suggest ON/OFF behavior pattern [9,10]. An application running on a machine may need to refer many other applications that are spread around the DCN. Thus, efficiency of the routing greatly influences the overall performance of DCN. DCell incurs additional overhead for address mapping and packet modification. This reduces the routing performance. In this section, we propose DCell-IP which eliminates the above overhead and improves the routing performance.

Classless Inter Domain Routing (CIDR) and Variable Length Subnet Mask (VLSM) help to create different size networks using hierarchy in IP address. CIDR and VLSM helps in better utilization of IP address space as allocation of IP addresses often follows topological significance [11]. We use this feature of IP address for implementation of DCell-IP.

3.1 Defining Different Levels

As in original DCell implementation, we number each server in *DCell-IP*$_k$ with <$a_k a_{k-1} \ldots a_0$>. But, instead of using different address format as in original implementation, we embed the address information inside IP address. We explain in this section, the mechanism for assigning addresses to servers in different levels of DCell-IP. For this we denote X to denote an octet of IP address and *x* to denote a bit in octet. Servers and the interfaces connecting to switches at different levels are assigned IP address of the form *10.2.xxxxxxxx.X*. The last octet of IP address is used for addressing the servers and the intermediate connections. We choose remaining three octets to define each level. This is illustrated in the Fig-1.

Fig. 1. Defining levels using recusrive format of IP address

3.2 Constructing Different Levels

In DCell-IP, we connect servers of *DCell-IP*$_0$ to TOR switch and TOR switches to other switches at different levels instead of connected servers to switches as in original implementation. We made this change because conventional networks favors multi-port switches rather than multi-port servers. This arrangement also reduces latency by one hop compare to original implementation. The major advantage we gain from this modification is, we are able to check the double exponential growth of DCell construction. This is turn enables us to encode DCell address inside IP address which is of 32 bits in size. We illustrate this with example. Original implementation of *DCell*$_3$ will have 1,76,820 servers. Next level, *DCell*$_4$ will have 1,76,821 *DCell*$_3$s and total of 1,76,820 * 1,76,821 servers. Not just any data center need these many servers, but it is also difficult to encode addresses above *DCell*$_3$.

To overcome this problem, we used *DCell-IP*$_0$ as basic building blocks, instead of servers of *DCell-IP*$_0$. For *n=4*, *DCell-IP*$_1$ will have *4+1=5 DCell-IP*$_0$s. The Fig-2 better illustrates the differences between original implementation and our proposed implementation using IP address hierarchy. We give detail comparison for two implementations in the section 4.

Recursive structure of *DCell-IP*$_k$ is defined similar to original implementation of DCell using g_k and t_k . But g_k and t_k are defined in terms of *DCell-IP*$_0$ instead of servers in *DCell-IP*$_0$. More specifically, our new implementation considers TOR switches as the components for defining g_k and t_k. We explain this with an example. In original implementation, *DCell*$_2$ have 21 *DCell*$_1$s since each *DCell*$_1$ have 20 servers. Whereas *DCell-IP*$_2$ will have 6 *DCell-IP*$_1$s since *DCell-IP*$_1$ have 5 TOR switches. Thus, *DCell-IP*$_3$ will have 421 *DCell-IP*$_2$. These 421 *DCell-IP*$_2$s contain 1,76,821 switches supporting 7,07,284 servers. Modern data centers will not be having these many servers in present scenario. Also, at present 48-port and 64-port switches are often used in data centers. In general, with *n-port* switch we can support *(n-4) * 1,76,821* servers. For n=48 or n=64, this is going to be a very large number. Thus, through our new

implementation, we are restricting the design of DCell-IP to only 4 levels. This enables us to encode the DCell addresses inside IP address. The last octet of IP address is reserved for addressing interfaces of TOR switches of *DCell-IP*$_0$s. These interfaces are used for connecting servers and TOR switches of higher *DCell-IP*s. We use remaining three bytes of IP address for defining levels.

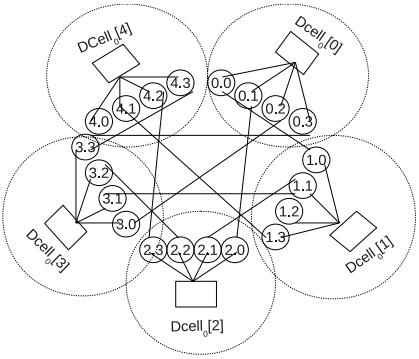


Fig. 2(A). Original Implementation of DCell

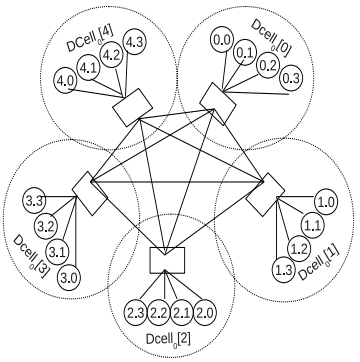


Fig. 2(B). DCell uing IP address Hierarchy

For better understanding we use *n=4*, but our new implementation works for any value for *n*. Only factor the *n* decides is, number of bits to be used for encoding each level. For *n=4*, as *DCell-IP*$_1$ have 5 *DCell-IP*$_0$s, we use *3* (2^3*=8)* bits to encode *level*$_1$. The last three bits of third octet is used for encoding level1. Next, as *DCell-IP*$_2$ have 21 *DCell-IP*$_1$, we use next *5* (2^5*=32)* bits for defining *level*$_2$. Finally, as *DCell-IP*$_3$ have 421 *DCell-IP*$_2$s, we use next *9* (2^9*=512)* bits for defining the *level*$_3$. The arrangement of encoding DCell address inside IP address is shown in the Fig-1.

3.4 Routing in DCell-IP

We use similar algorithm to determine the path from source to destination as in original implementation. But after determining the path, we embed the path information inside IP address and directly use source routing feature of IP for data transfer. The new routing algorithm is given in the Algorithm-1. For the source 10.0.17.1 (10.0.00010001.1) and destination 10.0.25.2 (10.0.00011001.2), respective <a_2, a_1> and <b_2, b_1> are <*2, 1*> and <*3, 1*>. Original DCellRouting for <*2, 1*> and <*3, 1*> gives the path (<*2, 1*>,<*2, 2*>,<*3, 2*>,<*3, 1*>). Encoding the path with IP address gives the path (10.0.17.1, 10.0.18.0, 10.0.26.0, 10.0.25.2).

We use this path directly in source routing for data transfer. Thus without using DCell addresses we perform the source routing from source to destination. This not just eliminates the need for mapping DCell address to IP addresses during data transfer but also eliminate the need for modifying the packet to include DCell header.

Next, to enable TOR switch to select the next hop during the data transfer we followed the symmetry in assigning the addresses for interfaces of TOR switch.

Algorithm 1: New source routing for DCell
DCellRoutingIP (*Saddr*, *Daddr*)
Input: *Saddr* (Source Address), *Daddr* (Destination Address)
Output: *Path* from *Saddr* to *Daddr*
1. Mask last eight bits of Saddr, Daddr
2. Extract bits 24-16, 15-11 and 10-8 of Saddr into A $<a_3, a_2, a_1>$
3. Extract bits 24-16, 15-11 and 10-8 of Daddr into A $<b_3, b_2, b_1>$
4. Call DCellRouting (A, B) /* Call to original DCell Routing */
5. Encode the Path returned from DCellRouting in IP address.
6. Return encoded path in IP address format.

This is given in the Table-1. If first four bits of IP address are used for assigning addresses to servers connected to $DCell\text{-}IP_0$, then fifth bit set along with mask up to fifth bit of fourth octet is used for addressing the interface connecting the switch at $level_1$. Similarly, sixth bit set along with the mask up to sixth bit of fourth octet is used for interfaces connecting switch at *level2*. Similar pattern is continued for assigning addresses to the interfaces connecting further levels. The intermediate switch is selected through the interfaces through which TOR switch of DCell is connected, which in turn determined through the bits corrected.

Table 1. IP/Mask for interfaces of TOR switch

IP/Mask	Address Range	Assigned To
10.0.X.0/29	10.0.X.1 – 10.0.X.6	For $DCell_0$ servers
10.0.X.0/29	10.0.X.17 – 10.0.X.30	Switches at $DCell_1$
10.0.X.0/29	10.0.X.33 – 10.0.X.62	Switches at $DCell_2$
10.0.X.0/29	10.0.X.65 – 10.0.X.126	Switches at $DCell_3$
10.0.X.0/29	10.0.X.129 – 10.0.X.254	Switches at $DCell_4$

4 Comparisons

We compare our implementation of DCell with the original implementation. The Table-2 summarizes the advantages of our new implementation against the original implementation.

Not just our new implementation eliminates the need for maintaining and mapping DCell address to IP address but also reduces total number of switches and interconnections at intermediate levels. Our implementation also eliminates the need for modifying the packet header to include the DCell header. Finally, as servers connected to TOR switch at $DCell\text{-}IP_0$ are assigned IP addresses in the same broadcast domain, we reduce total path length by one hop. This is because, once the packet reached the destination $DCell\text{-}IP_0$, it is broadcasted to reach all the servers connected to $DCell\text{-}IP_0$. In the original implementation, if server connecting to different $DCell_k$s and destination servers are different and when the packet reaches the first, it will be forwarded to the switch connecting them and from there to destination.

Table 2. System Comparison

	Addressing Scheme	Routing	Address Translation	Switches at Level k	Interconnection at Level k
DCell	DCell Addresses	DCell Source Routing	$O(t_k)^1$	$O(g_k)^1$	$O(t_k * g_k)^1$
DCell-IP	IP Addresses	IP Source Routing	$O(t_k)^2$	$O(g_k)^2$	$O(t_k * g_k)^2$

1 – Defined in terms of number of servers
2 – Defined in terms of number of TOR switches

5 Experiments

Experiments are conducted using NS3 simulator. For better comparison, three implementation of DCell is done. For this DCells of different levels with *n=4* are constructed. In the first implementation, OSPF is used as routing protocol. In the second implementation, original DCell Source Routing (DSR) is used for routing. In the third implementation, our proposed source routing using IP address hierarchy is used for routing. For simulation, we used Intel® Core2™ Duo CPU E8400 @3.00 GHz processor with 512 KB cache and 4 GB RAM, running Laughlin GNU/Linux 2.6.35.

5.1 Memory Usage

For measuring memory usage profiling is done using valgrind. For this, we constructed $DCell_3$ for n=4 supporting 256 servers. Total memory used by different implementation is given in Fig-3. Analysis of the snapshot indicated the peak usage of memory during the data transfer and it is the routing algorithms that use large memory. The graph in the figure indicates that OSPF is using large memory and memory usage is considerably reduced with the DSR. We achieved still less memory usage through our new implementation.

This improvement is a result of our design avoiding necessity of mapping DCell addresses to IP addresses during every packet transfer. Next, we recorded peak memory usage for each implementation for different levels of DCells. The graph in Fig-4 gives indication for increase in memory requirements for different DCell implementation supporting large number of servers.

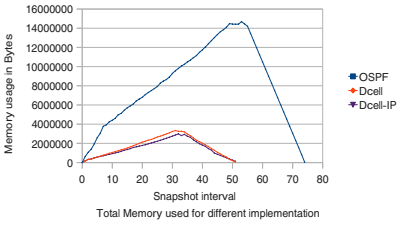


Fig. 3. Memory usage

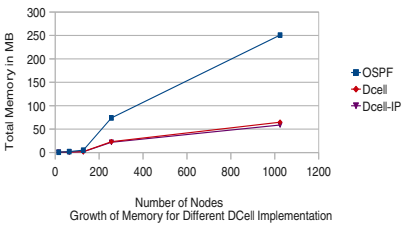


Fig. 4. Growth of memory

5.2 Route Delay

To demonstrate the efficiency of different routing algorithms, we constructed a large topology consisting of 1024 servers. We used flow monitor feature NS-3 to measure the delay. We used bulk-send application as traffic generator and each node sending data of 500 Mbps. Following is the benchmark suite used for measuring performance.

- *All-to-all*: Every node is sending traffic to every other node.
- *Stride(i)*: A host at x DCell_0 of y $DCell_k$ is sending traffic to host at $x+i$ $DCell_0$ of $y+i$ $DCell_k$.
- *StaggeredProb (Level3P, Level4P)*: A host at x $DCell_0$ is sending traffic to host at y $DCell_0$, where x and y belongs to same $Level_3$ with the probability of *Level3P*. Next, host at x $DCell_0$ is sending traffic to host at y $DCell_0$, where x and y belongs to same $Level_4$ with the probability of *Level4P*. Finally, host $DCell_0$ is sending traffic to host at arbitrary $DCell_0$ with the probability *(1 – Level3P + Level4P).*

The Table-3 gives the average delay for all the flows in respective scenario for three implementations. Our proposed routing algorithm works better compare to original implementation as it does not need translation of DCell address to IP address at every hop thereby avoiding the modification of packet to include DCell header. To substantiate our claim we measured the time for selecting path from source to destination for all three implementations. The result is given in the Table-4.

We simulated all-to-all traffic for different size DCells and measured the time taken for each flow to compute the path to the destination. We sum them all to obtain the result. For OSPF, we measured the time taken for routing tables to converge.

Table 3. Average in seconds for each flow

Traffic	DCell-IP	DCell	OSPF
All-to-All	0.0628	0.0631	0.1031149
Stag-Prob(3,2)	0.02113	0.02006	0.0331219
Stag-Prob(5,3)	0.02242	0.02186	0.0362820
Stride(4)	0.02198	0.02067	0.0304013
Stride(8)	0.02360	0.02140	0.0310824
Stride(16)	0.02406	0.02253	0.0332413

Table 4. Delay in seconds for computing route

No of Nodes	DCell-IP	DCell	OSPF
16	0.82	0.93	1.05
64	1.05	1.26	3.23
128	1.26	1.68	12.21
256	2.09	2.62	73.15
1024	10.57	11.74	3967.91

6 Conclusions

In this paper, we proposed a new implementation for DCell. Our proposed implementation used IP address hierarchy to encode the DCell addresses. We all together eliminated the need for new addressing scheme as in original implementation. We achieved this using the hierarchical nature of IP address. As construction of DCell is hierarchical and symmetrical and also is construction of subnet and supernet addresses, we embedded the DCell addresses inside the IP address. This not only eliminated the need for new addressing scheme but also resulted in simpler source routing.

Through our new implementation we transformed DCell from general Hypercube to pure networking entity. We used only IP address for servers and intermediate switches and use source routing of IPV4 for communication. The results obtained showed improvement in performance of routing with respect to original implementation of DCell as our proposed design eliminates the need for mapping DCell address to IP address during every packet transfer.

References

[1] Krishna, K.: Data center evolution: A tutorial on state of art, issues, and challenges. Computer Networks 53(17), 2939–2965 (2009)

[2] Vishwanath, K.V., Greenberg, A., Reed, D.A.: Modular data centers: how to design them? In: Proceedings of the 1st ACM Workshop on Large-Scale System and Application Performance, LSAP 2009, pp. 3–10. ACM, New York (2009)

[3] Popa, L., Ratnasamy, S., Iannaccone, G., Krishnamurthy, A., Stoica, I.: A cost comparison of datacenter network architectures. In: Proceedings of the 6th International Conference, Co-NEXT 2010, pp. 16:1–16:12. ACM, New York (2010)

[4] Greenberg, A., Hamilton, J., Maltz, D.A., Patel, P.: The cost of a cloud: Research problems in data center networks. SIGCOMM Comput. Commun. Rev. 39, 68–73 (2008)

[5] Guo, C., Wu, H., Tan, K., Shi, L., Zhang, Y., Lu, S.: Dcell: a scalable and fault-tolerant network structure for data centers. In: Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication, SIGCOMM 2008, pp. 75–86. ACM, New York (2008)

[6] Guo, C., Lu, G., Li, D., Wu, H., Zhang, X., Shi, Y., Tian, C., Zhang, Y., Lu, S.: Bcube: a high performance, server-centric network architecture for modular data centers. In: Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM 2009, pp. 63–74. ACM, New York (2009)

[7] Al-Fares, M., Loukissas, A., Vahdat, A.: A scalable, commodity data center network architecture. SIGCOMM Comput. Commun. Rev. 38, 63–74 (2008)

[8] Greenberg, A., Lahiri, P., Maltz, D.A., Patel, P., Sengupta, S.: Towards a next generation data center architecture: scalability and commoditization. In: Proceedings of the ACM Workshop on Programmable Routers for Extensible Services of Tomorrow, PRESTO 2008, pp. 57–62. ACM, New York (2008)

[9] Benson, T., Akella, A., Maltz, D.A.: Network traffic characteristics of data centers in the wild. In: Proceedings of the 10th Annual Conference on Internet Measurement, IMC 2010, pp. 267–280. ACM, New York (2010)

[10] Benson, T., Anand, A., Akella, A., Zhang, M.: Understanding data center traffic characteristics. In: Proceedings of the 1st ACM Workshop on Research on Enterprise Networking, WREN 2009, pp. 65–72. ACM, New York (2009)

[11] Ruiz-Sanchez, M.A., Biersack, E.W., Dabbous, W.: Survey and taxonomy of ip address lookup algorithms. IEEE Network 15, 8–23 (2001)

An Approach to Securing Data in Hosted CRM Applications

Siddharth M. Pandya, Abhishek Srikumar, and Chandrika T.

Infosys Limited, Survey No.210,
Rajendranagar Mandal,
RR Dist., Hyderabad – 500019
{siddharthm_pandya,abhishek_srikumar,chandrika_t}@infosys.com

Abstract. Organizations which look forward to expand its customer base focus heavily on customer relations. To retain customers, it is important to 'know the customers' and this paves the way for Customer Relationship Management (CRM). All such organization needs the privacy of customer information that is being handled by various stakeholders. There are numerous statutory and regulatory requirements that mandate the CRM vendors, who provide hosted CRM applications, to ensure data security and data privacy.

Despite all these numerous regulations by various enforcement bodies, an assuring solution addressing the concern of various organizations to adapt the hosted CRM solution prevails. Few of the concerns are:

a. Organizations outsource security management to a third party that hosts their IT assets (loss of control).[2]
b. With multi tenancy as one of the major characteristics of cloud model, the organizations are unsure of the strength of security controls exercised. [2]
c. The lack of documented security assurance processes and procedures by the cloud providers. [2]
d. Hosting of relatively valuable assets on publicly available infrastructure increases the probability of attacks. [2]

For business organizations to move their data on the cloud, it is critical for the end users to have a reassurance on the security of the data. An effective means to achieve this is to encrypt or mask the data even before it is streamed to the cloud.

This implies that the vendors offering the hosted CRM system could also provide the technology to encrypt or mask the critical fields as an Out-Of-The-Box (OOTB) feature. While the Cloud Service Providers assure that the data in their databases comply with various statutory regulations to protect the raw data, this feature further offers to the end users the necessary empowerment to protect the data that is critical for them.

1 Introduction

For every organization, the prevailing market scenario cannot over emphasize the need to manage the customers in the best possible manner. Customer Relationship

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 747–751.
springerlink.com

Management refers to methodologies and technologies that enable the business organization for customer acquisitions, retention and extension. [3]

Data security and data privacy are critical areas where the organizations want to take adequate measures. Though data security and data privacy may be related, they have different meanings. Data security is the ability of a system to protect information and system resources with respect to confidentiality and integrity. Data privacy relates to digital collection, storage and sharing while adhering to the statutory regulations involved subject to the industry domain and the geographical location.

Currently, security and privacy parameters are evaluated in an ad hoc basis according to the requirements of the customers. For example, a customer may be interested in knowing [4]:

a) How customers' data are protected from an unauthorized access?
b) How is data privacy ensured?
c) What are the security policies and how often have they changed in the past?

As the security and privacy policies differ from each cloud service provider and are not standard, the risk of the cloud provider's not adhering to the statutory norms to secure the privacy of their customers' data prevails.

The lack of a standardized approach for establishing security and privacy metrics and enforcing them on the cloud adds to the apprehension and therefore, business organizations hold the plans of moving to cloud. [4]

2 Various Prevailing Options Available to CRM Vendors

To enable the organizations know their customers, IT plays a pivotal role in CRM. CRM application vendor giants like Oracle®, SAP®, SFDC®, Microsoft® and Amdocs® have developed software that enables an organization to capture and analyze the marketing trends, purchasing behavior of customers and potential opportunities and threats, by providing an effective means of maintaining CRM database. While CRM applications can be deployed within the premise, the impelling trend is to opt for a hosted solution because of the cost-benefit analysis ratios.

2.1 Prevailing Scenario

Despite efforts from various cloud service provider, and the organization using their service to ensure data security and privacy, the data still continues to remain raw and therefore prone to be misused. There are very few third party vendors and cloud providers who offer data encryptions or masking feature; and most of these vendors offer it as a premium feature that is completely absent in the OOTB software.

2.2 Implications with the Prevailing Approach

The ownership of data encryption or masking on the cloud or via a third party vendor is yet to be proven. The reason being, that the data is still available in raw format and

an entity with malefic intent may gain access to it. This only gives data security a new garb instead of eliminating the threat. In addition to this, the possibility of data theft by use of crawlers is to be studied too. Finally, these features are inherent with additional overheads as 'premium features', that too from third party vendors and not through the package vendors.

2.3 Alternative Approach to Data Security for Hosted CRM Applications

A potential approach would be for the CRM application vendors to provide data encryption or masking technique at field level as an OOTB feature. The field level encryption eliminates the complexity involved in encrypting and decrypting the complete dataset in the application. With this feature, the organizations would not only reduce their investment on IT, but also have a reliable means to achieve data security.

3 Recommendation

At the outset a consortium needs to be established to define the relevant standards and protocols, as-well-as to assist the CRM application vendors to understand the evolving need and trends in the data management by the end users. This consortium would also be responsible to make recommendations around the framework to be used/ implemented by the vendors. This would not only offer standardization, but also ensure basic level of acceptable security being offered by the application vendor.

3.1 Approach

Based on the consortium's guidelines, a standard has to be evolved detailing the various levels of criticality for fields of interest, guidelines around encrypting or masking these fields (their dos and don'ts) and an approach to implement this in an efficient and effective manner. The consortium's guidelines should help in directing the organizations for regulatory compliance as it is the responsibility of the organizations for security and integrity of the data and not that of the cloud service provider. [1]

Based on these recommendations from the consortium, CRM vendors should offer the user the empowerment to not only choose the fields that may be critical to the organization but also define the degree of protection for these fields. This should then invoke a set of encryption or masking algorithms - depending on the severity and criticality of the data. The CRM applications should give the end users the option to select the fields, either during installation and configuration or through administrative privileges, to be encrypted or masked.

3.2 Implementation Approach

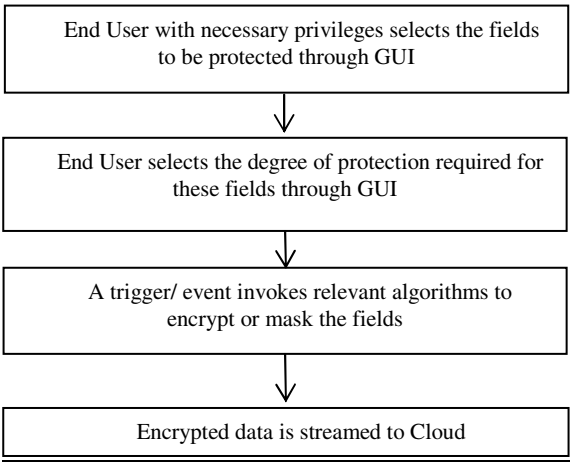


Fig. 1. End User's Perspective

The end user defines the field sets for encryption or masking based on the criticality of the data. The CRM application should select appropriate encryption or masking algorithms based on the desired level of protection of the fields and thus, encrypt the data before streaming it onto the cloud.

3.3 Benefits

With data encryption or masking being offered as an OOTB feature, the end user can potentially expect that,

a. The CRM applications to be configurable in a manner that the fields to be encrypted or masked can be identified a priori - as this would be highly end user driven.
b. The field level encryption eliminates the complexity involved in encrypting and decrypting the complete dataset in the application.
c. There can be reasonable assurance on the data streamed to cloud via the internet to be secure.
d. The end user's investment on IT infrastructure and private cloud could be avoided.

4 Conclusion and Future Scope of Work

While this paper suggests an approach, the CRM vendors will have to come up with the implementation methodology as-well-as technology, that best suits the end user requirements to encrypt or mask the data at field level, and provide a simple Graphical User Interface (GUI) for end users to comprehend the feature.

Data confidentiality is one of the key concerns for end users to move towards cloud. With this approach, the organizations can host the CRM application on the cloud with reinforced degree of confidence.

References

1. Kandukuri, B.R., Ramakrishna Paturi, V., Rakshit, A.: Advanced Software Technologies. Cloud Security Issues (2009) IEEE ©, doi:10.1109/SCC.2009.84
2. Morsy, M.A., Grundy, J., Müller, I.: An Analysis of The Cloud Computing Security Problem. In: Proceedings of APSEC 2010 Cloud Workshop, Sydney, Australia, November 30 (2010)
3. Chowhan, S.S., Saxena, R.: Customer Relationship Management from the Business Strategy Perspective with the Application of Cloud Computing. The Proceedings of DYNAA 2(1) (2011)
4. Nourian, A., Maheswaran, M.: Privacy and Security Requirements of Data Intensive Computing in Clouds. In: Furht, B., Escalante, A. (eds.) Handbook of Data Intensive Computing, ch. 19, ©Springer Science+Business Media, LLC (2011), doi:10.1007/978-1-4614-1415-5_19
5. Carlin, S., Adams, M., Curran, K.: Security issues in cloud computing (2011), http://www.elixirjournal.org/user_articles/1314858615_38%20%204069-4072.pdf (accessed on March 9, 2012)
6. ENISA, Cloud computing: benefits, risks and recommendations for information security (2009), http://www.enisa.europa.eu/act/rm/files/deliverables/cloud-computingrisk-assessment (accessed March 9, 2012)
7. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. Journal of Network and Computer Applications, Corrected Proof (in press)

Alzheimer's Disease Detection Using Minimal Morphometric Features with an Extreme Learning Machine Classifier

M. Aswatha Kumar[1] and B.S. Mahanand[2]

[1] Department of Information Science and Engineering,
M.S. Ramaiah Institute of Technology,
Bangalore, India
maswatha@yahoo.com
[2] Department of Information Science and Engineering,
Sri Jayachamarajendra College of Engineering, Mysore, India
bsmahanand@sjce.ac.in

Abstract. In this paper, we present an accurate method of detection of Alzheimer's disease using a minimal number of voxel-based morphometry features obtained from the brain MRI scans. The problem of early detection of AD is formulated as a binary classification problem and solved using an extreme learning machine classifier. The functional relationship between the voxel-based morphometry features extracted from magnetic resonance images and Alzheimer's disease is approximated closely using the extreme learning machine classifier. Since, the extreme learning machine is computationally efficient and provides a better generalization ability, Principal Component Analysis along with the Extreme Learning Machine classifier (referred to here as the PCA-ELM classifier) is used to select the minimal set of morphometric features from the brain MRI images for Alzheimer's disease detection. Performance of the PCA-ELM classifier is evaluated using the Open Access Series of Imaging Studies (OASIS) data set. The results are also compared with the well-known support vector machine classifier. The study results clearly show that the PCA-ELM classifier produces a better generalization performance with a minimal set of features.

1 Introduction

Dementia is a clinical syndrome characterized by a significant loss or decline in memory and other cognitive abilities. Alzheimer's Disease (AD) is the most common cause for dementia in elderly persons and is a progressive, neuro-degenerative disorder that leads to memory loss, problems in learning, confusion and poor judgment. Early detection of AD will help in slowing down its progression. One can detect AD by performing a brain autopsy which is an invasive technique. Another way of detecting AD is by performing brain imaging. Magnetic Resonance Imaging (MRI) is one of the most important brain imaging technique that provides an accurate information about the shape and volume of the AD brain.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 753–762.
springerlink.com

Typically, two major ways of extracting the features from the MRI scans are: the Regions-of-Interest (ROI) approach and the whole brain morphometric approach. The morphometric approach focusses on an automated whole brain analysis, whereas ROI approach concentrates on specific brain regions identified by manual tracing. Hence, its performance is influenced by human error in the tracing. Among different morphometric approaches, Voxel-Based Morphometry (VBM) is well-known [1, 2].

In this study, we use the VBM approach which identifies those significant areas with an increase in gray matter density in the normal persons relative to the AD patients by performing the following operations, viz., unified segmentation, smoothing and statistical testing. From the VBM identified significant areas (voxel locations), the gray matter tissue probability values are extracted as AD features.

Application of machine learning methods is becoming popular for detection of AD using MRI. There has been a growing interest in using Support Vector Machines (SVM) for AD classification [4, 9]. The application of ANNs for AD detection have been reported in [5, 14]. In [5], models such as Multi-layer Perceptron (MLP) and k-nearest neighbor (k-NN) were used. Another related study is reported in [14], where Radial Basis Function Networks (RBF), Learning Vector Quantization Networks (LVQ) and Probabilistic Neural Networks (PNN) are used. The computational efforts for training high dimensional features and sample imbalances influence the accuracy of AD detection using the above mentioned methods. In addition, the learning process is computationally intensive.

AD detection using the method of Principal Component Analysis (PCA) has been reported in [10, 17]. In [10], high dimensional features extracted using VBM are reduced using the PCA and the reduced features were then used for a sequential detection scheme of AD using a Self-adaptive Resource Allocation Network (SRAN) classifier [15]. To approach the problems associated with high dimensional features, a computationally lesser intensive, Extreme Learning Machine (ELM) classifier for AD detection is used in this paper.

ELM is a fast learning algorithm which makes use of the Single Hidden Layer Feed-forward Neural Network (SLFN) [8]. In ELM, the hidden node's parameters are randomly generated and the output weights are analytically determined. Recently, various improved versions of the basic ELM have been developed for solving classification problems. A complete review on ELM with its different variants along with their different applications can be found in [7].

In this study, the MRI volumes of 30 mild AD to moderate AD patients and 30 normal persons from the publicly available Open Access Series of Imaging Studies (OASIS) data set have been used [12]. The gray matter tissue probability values extracted from the voxel locations of significant areas identified from the VBM approach are used as features. First, we investigate the performance of the ELM classifier with all the features. Next, an approach of employing ELM classifier with a reduced set of morphometric features obtained by the PCA method is presented. Performance comparison studies of PCA-ELM and SVM classifiers are also presented and the study results indicate that the PCA-ELM classifier achieves better generalization with lower number of features.

The paper is organized as follows: Section 2 describes the AD detection problem. Section 3 presents the feature reduction using the PCA method and classifications based on ELM/SVM. Finally section 4 summarizes the conclusions from this study.

2 Alzheimer's Disease Detection Problem

This section presents a brief description of the data used, the feature extraction using the VBM approach and also a brief review of the ELM classifier.

2.1 Materials

In our study, the publicly available OASIS data set has been used [12]. OASIS data set has a cross-sectional collection of 416 persons covering the adult life span, aged between 18 to 96 including individuals with AD in an early-stage. The data includes 218 persons aged between 18 to 59 years and 198 persons aged between 60 to 96 years. Of the 198 older persons, 98 had no AD i.e with Clinical Dementia Rating (CDR) of 0, 70 persons have been diagnosed with a very mild AD (CDR=0.5), 28 persons are diagnosed with mild AD (CDR=1) and 2 persons with moderate AD (CDR=2). In our study, we have not considered the 70 *very mild* AD patients and have concentrated only on 30 *mild and moderate* AD patients. For normal persons we have used 30 (out of 98) persons to maintain a sample balance in performing group analysis using VBM.

2.2 Feature Extraction

A feature extraction approach based on the VBM method is employed in this work [11]. VBM is a fast and fully automated approach to identify the regional gray matter differences between the brains of normal persons and AD patients [1]. The steps involved in our VBM approach are: unified segmentation, smoothing and statistical testing. The unified segmentation step is a generative modeling approach, in which tissue segmentation, bias correction and image registration are combined in a single model [2]. The segmented and registered gray matter images are then smoothed by convolving with an isotropic Gaussian kernel. Here, a 10 mm full width at half-maximum kernel was employed. After performing the unified segmentation and smoothing steps, statistical tests were conducted finally. Statistical testing uses a general linear model which is based on the random Gaussian field theory [6]. In our statistical testing, estimated total intracranial volume is used as the covariate in the design matrix of the general linear model. Also a two-sample t-test is performed on the smoothed images of normal persons and AD patients and a multiple comparison correction method, namely, family wise error with a $P<0▷05$ has been applied.

From this unified VBM approach, significant areas with an increase in gray matter density in the normal persons relative to the AD patients are obtained. From the voxel locations of these significant areas, gray matter tissue probability values are extracted as features.

The problem of AD detection can be formulated as a binary classification problem and the objective of the classification problem is to accurately identify the functional relationship between the extracted features from the significant voxel locations and the class labels. In this paper, we employ the ELM classifier to approximate this functional relationship.

2.3 A Brief Review of the Extreme Learning Machine Classifier

ELM is a single hidden layer feed forward neural network where the input weights are chosen randomly and the output weights are calculated analytically. ELM employs any continuous/discontinuous non-linear function (possible functions are sigmoidal or Gaussian or hard-limiting function) as an activation function in the hidden layer and a linear activation function in the output layer. In this paper, we use the Gaussian function as an activation function in the hidden layer. The architecture of ELM network is shown in Fig. 1.

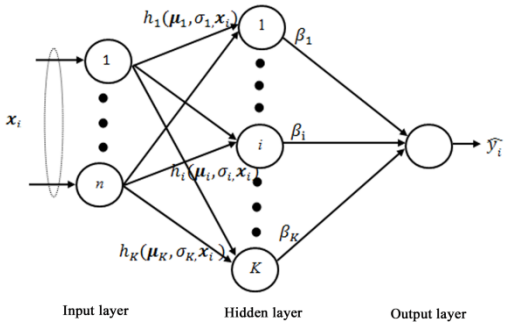


Fig. 1. ELM Network

Let $\mu_j \in \Re^n$ be the center and $\sigma_j \in \Re^n$ be the width of j^{th} Guassian neuron. For a given input $\mathbf{x}_i$, the output of ELM classifier $\hat{y}_i$ with K hidden neurons is given by

$$\hat{y}_i = \sum_{j=1}^{K} \beta_i \, h_j(\mu_j, \sigma_j, \mathbf{x}_i), \qquad i = 1,2,\ldots,N \tag{1}$$

where β_i is the output is the output weight connecting j^{th} hidden neuron and an output neuron $h_j(\mu_j, \sigma_j, \mathbf{x}_i)$ is the output of j^{th} Guassian neuron and is defined as

$$h_j(\mu_j, \sigma_j, \mathbf{x}_i) = \exp\left(\frac{-\|\mathbf{x}_i - \mu_j\|^2}{2\sigma_j^2}\right), \qquad j = 1,2,\ldots,K \tag{2}$$

The output of ELM classifier given in Eq. (1) can be written in matrix form as

$$\hat{Y} = \beta \, Y_h \tag{3}$$

where Y_h is an K x N output matrix, which is defined as

$$Y_h = \begin{bmatrix} h_1(\mu_1,\sigma_1,\mathbf{x}_1) & h_1(\mu_1,\sigma_1,\mathbf{x}_2) & \ldots h_1(\mu_1,\sigma_1,\mathbf{x}_N) \\ \vdots & \vdots & \vdots \\ h_K(\mu_K,\sigma_K,\mathbf{x}_1) & h_K(\mu_K,\sigma_K,\mathbf{x}_2) & \ldots h_K(\mu_K,\sigma_K,\mathbf{x}_N) \end{bmatrix}$$

In the ELM algorithm, the centers (μ) and widths (σ) of the Gaussian function parameters are chosen randomly for a given number of hidden neuron (K). By assuming that the predicted output $\hat{y}$ is equal to the coded labels y, the output weights β are estimated analytically as

(4)

where is the Moore-Penrose generalized pseudo-inverse of the hidden layer output matrix ($\mathbf{H_h}$).

In summary, the steps in developing the ELM classifier are:

- For a given set of training samples $(\mathbf{x}_i, y_i)$, select the number of hidden neurons;
- Select the centers μ and widths σ for the Gaussian functions randomly. Then, calculate the output weights β analytically as in :
- If the predicted output ($\hat{y}$) from ELM is greater than zero then the sample is considered to belong to an AD patient. If it is below zero then it belongs to a normal person.

In the ELM algorithm, the number of hidden neurons are chosen arbitrarily. In this paper, we use a constructive and destructive procedure described in [16] to select an appropriate number of hidden neurons required to approximate the decision function. Next, we evaluate the performance of the ELM classifier and compare the results with a standard Support Vector Machine (SVM) classifier based on the complete set of features obtained from VBM method.

2.4 Performance Evaluation Based on the Complete Set of Features

In our study using both the ELM and SVM classifiers, 75% and 25% samples are randomly chosen for training and testing for each trial. All the implementations and simulations for the ELM and SVM classifiers are carried out in a MATLAB 7.9 environment running in an Intel Xeon, 2.33 GHz processor. SVM experiments are carried out using the libSVM software package [3]. The input features are normalized in the range of [-1, +1] to avoid the dominance of some of the features. For SVM classifier, the cost parameter (c) and the kernel width (γ) of the Gaussian kernel are chosen using the grid search method. The features extracted using the VBM approach are then used as an input to the classifier. Performances of both the ELM and SVM classifiers are studied using 20 different random combinations of the training and testing data sets to obtain a meaningful mean and standard deviation of training and testing efficiencies. The mean, standard deviation (STD), and the best training/testing efficiencies for ELM and SVM classifier are presented in Table 1.

Table 1. Classification performances of ELM and SVM on all features

Classifier	Hidden neurons	Training efficiency			Testing efficiency		
		Mean	STD	Best	Mean	STD	Best
SVM	31	100	0	100	89	6.9	100
ELM	15	98.37	1.01	100	94.63	5.54	100

From Table 1, we can see that the ELM classifier produced a mean training efficiency of 98.37% and a mean testing efficiency of 94.63% with only 15 hidden neurons, whereas, SVM requires twice the number of hidden neurons as support vectors to produce a mean training efficiency of 100% and mean testing efficiency of 89%. The results clearly indicates that ELM classifier with all the features produces a mean of 5% improvement in generalization performance with minimal number of hidden neurons.

In our study, the sensitivity and specificity values are calculated to find the miss-classification rate of the classifier. The ELM classifier with all the 5788 morphometric features produced a mean sensitivity of 0.95 and mean specificity 0.95 whereas SVM classifier produced a mean sensitivity of 0.90 and mean specificity 0.89. This clearly indicates that the ELM classifier achieves better classification performance along with a reduced miss-classification rate.

3 AD Detection Using PCA Reduced Features with an ELM Classifier

All the 5788 morphometric features obtained from VBM approach may not be significant for AD detection. Hence, a study was conducted by employing PCA on the complete set of features to find whether a lower number of features are sufficient for an accurate detection of AD.

3.1 PCA-ELM Classifier

We propose a PCA-ELM classifier which does feature reduction using the PCA and then use these reduced features as an input to the ELM network for further classification of persons into AD patients or normal persons. The schematic diagram of the PCA-ELM classifier is shown in Fig. 2.

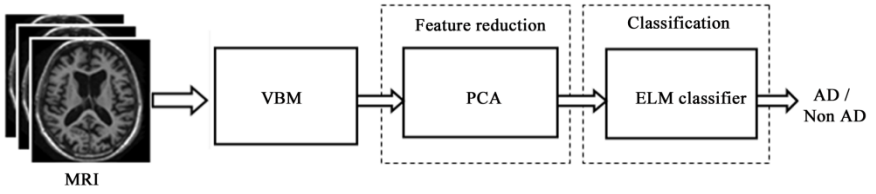


Fig. 2. PCA-ELM Classifier

Principal Component Analysis (PCA) is a statistical method of the mapping of the original high-dimensional data onto a lower-dimensional space [13]. PCA calculates the eigenvectors and eigenvalues of the sample covariance matrix of the data set. Principal components are the eigenvectors with the highest eigenvalues which are selected to represent the data in a new lower-dimensional space.

In our study, the training data consists of 45 samples with 5788 features, so the dimensionality of that subspace using PCA can maximally be 45. Hence, we use a

maximum of 45 significant features in the ELM and SVM classifiers. The 5788 morphometric features obtained using VBM approach were reduced to different combinations of 10, 20, 30 and 45 features using PCA.

3.2 Performance Evaluation of PCA-ELM Classifier

The classification performance of ELM classifier on PCA reduced features is shown in Table 2.

Table 2. Classification performance of ELM on PCA reduced features

No. of Features	Classifier	Hidden neurons	Training efficiency			Testing efficiency		
			Mean	STD	Best	Mean	STD	Best
10	SVM	16	100	0	100	89.93	5.92	100
	SRAN	12	99.37	1	100	90.76	4.29	100
	ELM	12	98.94	1.12	100	89.56	7.19	100
20	SVM	19	100	0	100	89.54	6.26	100
	SRAN	13	99.76	0.70	100	91.18	5.27	100
	ELM	14	98.43	1.62	100	91	6.26	100
30	SVM	21	100	0	100	90.24	6.28	100
	SRAN	11	99.11	1.09	100	90.83	5.22	100
	ELM	14	98.58	1.09	100	89.87	6.80	100
45	SVM	30	100	0	100	90.57	6.29	100
	SRAN	15	99.24	1.07	100	90.42	3.21	100
	ELM	15	98.94	1.12	100	90.57	5.90	100

From Table 2, we can see that PCA reduced 20 features using ELM produced a mean testing efficiency of 91% while SVM produced 89.54%. A mean testing efficiency of 90.57% was obtained using the SVM on PCA reduced 45 features while ELM also produced the same 90.57%. The mean training efficiency of 100% was obtained by SVM with 45 features, where as the ELM classifier achieved the mean training accuracy of 98.34% with 20 features. Also, the Table 2 indicates that for the case of 20 features the best training and testing efficiencies are 100% for the 20 random trials.

Another related study in [10], was conducted using VBM along with PCA and a SRAN classifier on the same OASIS data set. From Table 2, it is observed that SRAN classifier produced a mean training efficiency of 99.76% and a mean testing efficiency of 91.18% using 20 PCA reduced morphometric features. From this study, it can be inferred that both ELM and SRAN classifiers achieve similar generalization performances with PCA reduced features.

Fig. 3 shows the sensitivity performance of both the ELM and SVM classifiers with the PCA reduced features. The figure shows the number of features along the X axis and the sensitivity is shown along the Y axis.

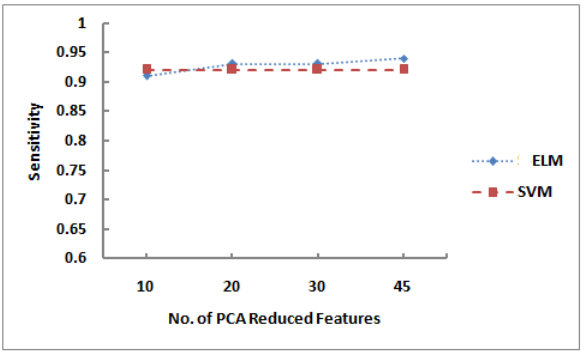


Fig. 3. Sensitivity performance of ELM and SVM classifiers on PCA reduced features

Fig. 4 shows the specificity performance of ELM and SVM classifiers with PCA reduced features, where the number of features are shown along the X axis and the specificity is shown along the Y axis.

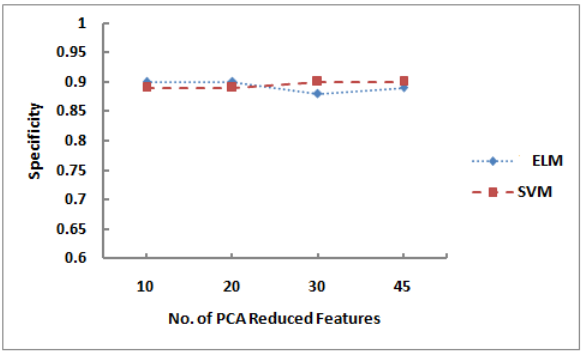


Fig. 4. Specificity performance of ELM and SVM classifiers on PCA reduced features

From Fig. 3 and Fig. 4 it is observed, that using the set of PCA reduced 45 features, the ELM classifier produced a mean sensitivity of 0.94 and mean specificity 0.89 whereas the SVM classifier produced a mean sensitivity of 0.92 and mean specificity 0.90. Also, the mean sensitivity of the ELM classifier is marginally higher as compared to the SVM classifier for all combinations of the features sets. This clearly indicate that ELM classifier reduces the mis-classification rate and has better classification accuracy with PCA reduced features.

From this study, it can be inferred that lower number of features are sufficient for the accurate classification of AD patients and normal persons using the ELM classifier.

4 Conclusions

This paper has presented a new approach of using the PCA with an ELM classifier to select a minimal number of morphometric features for accurate AD detection. The performance of the PCA-ELM classifier have been evaluated using OASIS MR images of 30 normal persons and 30 AD patients. With complete morphometric features (5788), ELM classifier produced a mean testing efficiency of 94.63% whereas SVM produced 89%. With 20 PCA reduced features, ELM produced a mean testing efficiency of 91% compared to SVM's of 90.57% using 45 reduced features. Based on the study results, it can be concluded that accurate detection of AD can be performed using VBM-PCA-ELM classifier with a minimal number of features. Also the above approach produces better generalization and lower misclassification rates. Further, ELM classifier may be used with features extracted from other morphometric methods namely the tensor-based morphometry for AD detection.

Acknowledgments. The authors are indebted to Professor N. Sundararajan and Dr. S. Suresh, School of Computer Engineering, Nanyang Technological University, Singapore, for their valuable inputs and contribution to this work. We thank the Washington University Alzheimer's Disease Research Center for making the MRI data available for this study.

References

1. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. NeuroImage 11(6), 805–821 (2000)
2. Ashburner, J., Friston, K.J.: Unified segmentation. NeuroImage 26, 839–851 (2005)
3. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Trans. Intelligent Systems and Technology 2, 27:1–27:27 (2011)
4. Davatzikos, C., Fan, Y., Wu, X., Shen, D., Resnick, S.M.: Detection of prodromal Alzheimer's disease via pattern classification of MRI. Neurobiology of Aging 29, 514–523 (2008)
5. El-Dahshan, E.S.A., Hosny, T., Salem, A.B.M.: Hybrid intelligent techniques for MRI brain images classification. Digital Signal Processing 20(2), 433–441 (2010)
6. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189–210 (1994)
7. Huang, G.B., Wang, D.H., Lan, Y.: Extreme learning machines: A survey. Int. J. Machine Leaning and Cybernetics 2(2), 107–122 (2011)
8. Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: Theory and applications. Neurocomputing 70, 489–501 (2006)
9. Kloppel, S., Stonnington, C.M., Chu, C., Draganski, B., Scahill, R.I., Rohrer, J.D., Fox, N.C., Jack Jr., C.R., Ashburner, J., Frackowiak, R.S.J.: Automatic classification of MR scans in Alzheimer's disease. Brain 131(3), 681–689 (2008)
10. Mahanand, B.S., Suresh, S., Sundararajan, N., Aswatha Kumar, M.: Alzheimer's disease detection using a self-adaptive resource allocation network classifier. In: Proceedings of International Joint Conference on Neural Networks (IJCNN 2011), San Jose, USA, pp. 1930–1934 (2011)

11. Mahanand, B.S., Suresh, S., Sundararajan, N., Aswatha Kumar, M.: Identification of brain regions responsible for Alzheimer's disease using a self-adaptive resource allocation network. Neural Networks (2012), doi:10.1016/j.neunet.2012.02.035
12. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J. Cognitive Neuroscience 19(9), 1498–1507 (2007)
13. Oja, E.: Neural networks, principal components and subspaces. Int. J. Neural Systems 1, 61–68 (1989)
14. Savio, A., García-Sebastián, M., Hernández, C., Graña, M., Villanúa, J.: Classification Results of Artificial Neural Networks for Alzheimer's Disease Detection. In: Corchado, E., Yin, H. (eds.) IDEAL 2009. LNCS, vol. 5788, pp. 641–648. Springer, Heidelberg (2009)
15. Suresh, S., Dong, K., Kim, H.J.: A sequential learning algorithm for self-adaptive resource allocation network classifier. Neurocomputing 73(16-18), 3012–3019 (2010)
16. Suresh, S., Mani, V., Omkar, S.N., Kim, H.J.: Divisible load scheduling in tree network with limited memory: A genetic algorithm and linear programming approach. Int. J. Parallel Emergent and Distributed System 21(5), 303–321 (2006)
17. Zhang, J., Yan, B., Huang, X., Yang, P., Huang, C.: The diagnosis of Alzheimer's disease based on voxel-based morphometry and support vector machine. In: Proceedings of Fourth International Conference on Natural Computation (ICNC 2008), Jinan, Shandong province, China, October 18-20, vol. 2, pp. 197–201 (2008)

Bidding Strategy in Simultaneous English Auctions Using Game Theory

Nirupama Pavanje

Dept of PG Studies, Jnanasangama,
VTU, Belgaum
nirupama.pavanje@gmail.com

Abstract. With more and more people using the internet for a wide range of purposes, internet use has become an absolute necessity for businesses to survive and grow. Online auction have expanded rapidly over the last decade and have become a fascinating new type of business or commercial transaction in this digital era. The online auction is an important e-commerce application which enables the buying and selling of goods through a dynamic pricing strategy. Users can access the auction system through the Web, WAP-enabled devices and agents. The paper assumes that the auction system supports only English auction. Predicting bidding strategy is not easy, since it is dependent on many factors such as the behavior of each bidder, the number of bidders participating in that auction as well as each bidder's reservation price. Here, simultaneous English auctions for the same item are considered. This paper uses the concept of Game Theory, to predict the bidding strategy in an auction and helps the user to decide whether to proceed with the auction or to back off from the auction so as to maximize the bidder's profit. This paper considers the user, bidding for an item simultaneously in more than one auction site.

Keywords: English auctions, bidding and Game Theory.

1 Introduction

In general, e-commerce is about the buying and selling of goods using network. Nowadays, online auctions have become a popular e-commerce business model. Auction is a market institution with an explicit set of rules determining resource allocation and price on the basis of bid from the market participants. Auction mechanisms have become very popular within electronic commerce and have been implemented in many domains with assorted environments. Unlike traditional auction houses, online auction websites offers a better place for people to purchase and publicize merchandise through a bidding process [5] [6]. Online auction have given consumers a "virtual" flea market with all the new and used merchandises from around the world. They also give sellers a global storefront where they are able to market their goods. By means of online auction; sellers can find suitable buyers effectively over the Internet using a dynamic pricing strategy.

In recent years, many commercial auction services have been launched, mostly based on the English auction method. Agents are well suited to support auction

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 763–771.
springerlink.com

services because of their autonomous and communicative functions. Concurrently, with the advancement in mobile computing technologies, handheld computing devices such as palm computers are becoming increasingly economical and popular. This presents opportunities for developing more advanced auction services. For example, a consumer can dispatch a bidding agent to the network from anywhere and control the agent activities through his handheld device.

Online auctions provide many benefits compared to the traditional auctions. One disadvantage of having traditional auction is that it requires simultaneous participation of all bidders or agents at the same location [18] [1]. In online auction, this does not exist as online auction allows clients to make their purchases anywhere anytime. Online auction also provide bidders more flexibility on when to submit their bids since online auctions usually lasts for days or even weeks [13]. Online auctions can be more effective as the target audiences will be in a mass amount where there is no geographical limitation since both sellers and buyers do their trading in a "virtual" environment any payment transaction can be made through the online banking. Having a low price and a wider market in products and services, it had made the online auction a success where it attracts many bidders and also sellers as well. Online auctions also allow sellers to sell their goods efficiently [19].

2 System Architecture

Fig 1 shows the general architecture of an auction system. It consists of the following subsystems.

- Access: This is for users to access the system through the client/server-based and agent-based approaches.
- Auction Logic: This is for handling types of auction techniques.
- Database: This is for storing various information such as auction-related information, customer records etc.
- Management System: This is for the system administrator to manage the system through a single interface.
- Other Support System: This provides different supporting functions, e.g. payment functions for the system.

In each auction process, an "auction space" is created and three types of agents are involved, namely:

- Buyer coordinator: for coordinating the bidding process.
- Buyer agent: for bidding the item.
- Seller agent: for selling the item.

To explain the English auction protocol, the following assumptions and notations are used:

- Consider there are N buyer agents in an auction space, i.e. A_1, $A_2 \ldots A_N$, managed by a seller/auctioneer agent.
- Currently the latest bid is the K^{th} bid, the maximum bidding price is M_k and the winner is W_k.
- The latest bid of buyer n is R_n.
- The current time is t.
- The deadline of the auction is d.

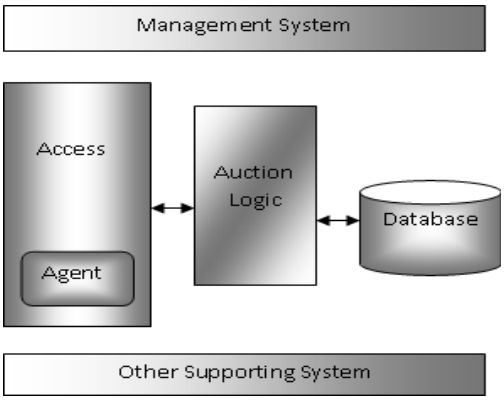


Fig. 1. Auction System Architecture

The implementation steps are:

- Step1: Before the deadline (t<d), the seller multicasts M_k to all buyers and waits for the response.
- Step2: For a buyer n, it may respond with a new bid R_n.
- Step3: Whenever a valid bid is received (i.e. $R_n > M_k$), the seller updates M_k, provided that t<d, and then repeats step1.
- Step4: The process stops when the deadline is reached (t<d).

There are a lot of prediction models in the market such as Neural Network, Fuzzy Logic, Time Series, Evolutionary Computations, Probability Function, Genetic Algorithm and Game Theory [10]. For example, to predict the bid price for an item, one has to consider the number of auctions selling the same item at the same time, the number of bidders participating, the behavior of each bidder and individual bidder's reservation price. The development of agents for bidding in multiple auctions involves three aspects [5]:

- Auction Tracking: Discovering and monitoring auctions for the required item.
- Bid Management: Placing bids and monitoring the outcome of these bids.
- Strategic Bid Planning: Deciding where to bid, when and how much.

This paper assumes that there are multiple English auctions. The auctions sell the same item. The required bidding information (previous and current bid prices) can be obtained. The aim is to formulate the bidding strategy for getting one item from the auctions.

Thus a bidder needs to determine how much to bid in each auction to maximize its payoff across the complete set of auctions.

Game Theory

It is a very useful tool for studying interactive decision-making, where the outcome for each participant or player depends on the actions of others [20]. Each decision maker is a player in the game of business; hence, when making a decision or choosing

a strategy one must take into account the potential choices of others, keeping in mind that while, making their choices, other players are likely to think about and take into account your strategy as well.

Game theory is [21]:

- A collection of tools for predicting outcomes of a group of interacting agents where an action of a single agent directly affects the payoff of other participating agents.
- The study of multi-person decision problems.
- A bag of analytical tools designed to help us understand the phenomena that we observe when decision-makers interact.
- The study of mathematical models of conflict and cooperation between intelligent rational decision-makers [4].

Game theory assumes:

- Each player in the market acts on self interest [9]. They pursue well defined exogenous objectives; i.e., they are rational. They understand and seek to maximize their own payoff functions.
- In choosing a plan of action (strategy), a player considers the potential responses/reactions of other players. She takes into account her knowledge and expectations of other decision makers' behavior; i.e., she reasons strategically.

[22] A game describes a strategic interaction between the players, where the outcome for each player depends upon the collective actions of all players involved. In order to describe a situation of strategic interaction, we need to know:

- The players who are involved.
- The rules of the game that specify the sequences of moves as well as the possible actions and information available to each player whenever they move.
- The outcome of the game for each possible set of actions.
- The (expected) payoffs based on the outcome.

An important issue is whether all participants have the same information about the game, and any factors that might affect the payoffs or outcomes of the game. We assume that the rules of the game are common knowledge [3]. That is each player knows the rules of the game, say X, and that all players know X, that all players know that all players know X, that all players know that all players know that all players know that all players know X and so on… and infinitum. In a game of complete information the player's payoff functions are also common knowledge. In a game of incomplete information at least one player is uncertain about another player's payoff function. For example, a sealed bid auction is a game of incomplete information because a bidder does not know how much other bidders value the item on sale, i.e., the bidder's payoff functions are not common knowledge. Bids are submitted in sealed envelopes and the participants do not have any information about their competitors' choices while they make their own.

Similarly Game theory could be used in online English auction so as to guide the customers in selecting the appropriate auction site and helps in decision making, i.e., whether the customers need to bid or back off from the auction by considering strategies of other bidders.

3 Design

An English auction is also known as an open ascending price auction [8]. Participants bid openly against one another, with each subsequent bid higher than the previous bid. An auctioneer may announce prices, bidders may call out their bids themselves (or have a proxy call out a bid on their behalf), or bids may be submitted electronically with the highest current bid publicly displayed. The auction ends when no participant is willing to bid further, at which point the highest bidder pays their bid. Alternatively, if the seller has set a minimum sale price in advance (the 'reserve' price) and the final bid does not reach that price the item remains unsold. Some-times the auctioneer sets a minimum amount by which the next bid must exceed the current highest bid. The English auction is commonly used for selling goods, most prominently antiques and artwork, but also secondhand goods and real estate.

The English auction protocol is as follows [17]:

- The buyer agent sends requests to different sellers agents selling the same product.
- The buyer uses a bid construction model which depends on the bidding status of the auction and uses the Game Theory to make decision, so that it maximizes his/her profit. Then bids are submitted to different seller agents simultaneously.
- The seller agent evaluates the various bids and he/she chooses the winning bid.
- The seller agent accepts the highest bid i.e. winning bid.

So, in the first phase, the bidder bids for an item in English auctions say 1, 2, 3… n, simultaneously and he will be setting a maximum payable price/budget for that item beyond which he is not ready to pay. Now, the concept of game theory comes into picture. Here, the game theory is used so that, the bidder can decide whether he should continue bidding in an auction say '1' or should he quit, depending on the values of bids submitted by other bidders in the same auction '1' [2].

[15] Here, the concept of game theory is applied for the bidder 'A', i.e., Game Theory helps the bidder 'A' to make decisions. Bidder 'A' will compare his payoff value with the other bidder's payoff value for the item being bid [16].

- If the payoff value of bidder 'A' is greater than the payoff value of the other bidders, then bidder 'A' will continue bidding until his maximum payable price reaches out. If it reaches the maximum price then bidder 'A' will quit or else bidder 'A' will continue bidding [7].
- If the payoff value of bidder 'A' is lesser than the payoff value of the other bidders, then bidder 'A' will quit if his next calculated bidding value is higher than his assumed maximum payable price.
- If the payoff value of bidder 'A' is lesser than the payoff value of the other bidders, then bidder 'A' will calculate the next bidding value. If the calculated next bidding value is lesser than the assumed maximum payable price, and if the next calculated bidding price is greater than the other bidder's last bid, then bidder 'A' may bid again [12].

And care should be taken so that, the bidder will not win in more than "one" auction. So, using game theory, bidder A will decide either to go ahead in bidding or to quit using the above three strategies. The payoff table of game theory would be only of two rows and two columns with entries of payoff values.

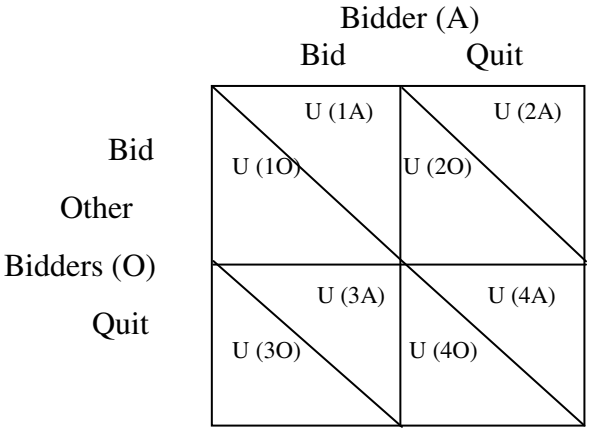


Fig. 2. Payoff Table

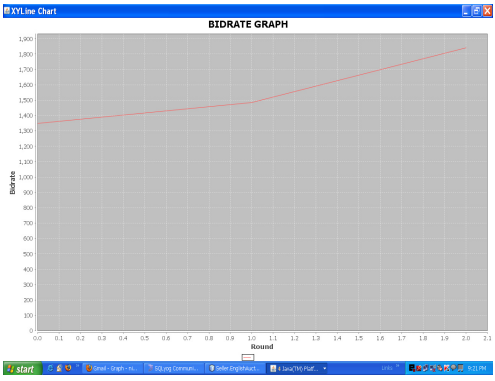
The players of the auction game are bidder A and the other bidders O [11], and both have two actions either to bid or to quit. What action to be performed is determined by the payoff values in the payoff table.

So, depending on the entries in the payoff table, Bidder A may either choose to continue bidding so that it maximizes his chances of winning or may quit so that it saves time and network resources. And this payoff table should be constructed for all the auctions where the Bidder A is bidding.

And also he needs to consider several factors such as the market price of the item being bid so that he will not be under loss, the maximum price that he is ready to pay, seller's feedback rating i.e., if the seller's feedback rating is not good (not a dependable seller) then he may not bid in that seller's auction, the cost at which the seller is ready to sell.

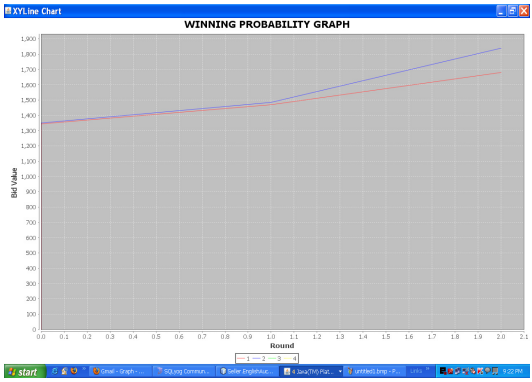
4 Results

BID RATE GRAPH



The values in the payoff table are calculated using Bid Rate Graph. The following snapshot depicts the bid rate graph of the bidder in concern (e.g., bidder A). Here the X axis corresponds to the round number and Y axis corresponds to the bid value in rupees. This graph depicts the pace in which the bidder is bidding.

WINNING PROBABILITY GRAPH



[14] The following snapshot depicts the winning probability (P(W)) of the bidder in concern (e.g., bidder A). Here the X axis corresponds to the round number and Y axis corresponds to the bid value in rupees. In this graph, each line shows how many times the bidder was the highest and in how many rounds.

SELLER'S STRATEGY GRAPH

The following snapshot depicts the seller's strategy. Here the X axis corresponds to the round number and Y axis corresponds to the bid value. In this graph, each line represents different sellers and also depicts when the bidder A has QUIT from which auction and in which round. It also depicts which seller has won the negotiation.

5 Conclusion

This paper has analyzed how game theory can be used in a simultaneous English auction and has proposed the design of it. This paper has also explained the auction architecture and protocol for English auction. Since game theory is used, this project reads the human rational thinking and hence helps the bidder to profit in an auction.

References

[1] Bagchi, A., Saroop, A.: Indian Institute of Management Calcutta. Internet Auctions: Some Issues and Problems
[2] ter Hofstede, A., Governatori, G., Dumas, M., Russell, N.: An Architecture for Assembling Agents that Participate in Alternative Heterogeneous Auctions. In: Proceedings of the 12th International Workshop on Research Issues in Data Engineering e-Commerce/e-Business Systems (RIDE 2002). IEEE (2002)
[3] Aumann, R.: Backward Induction and Common Knowledge of Rationality. Games and Economic Behavior 8, 6–19 (1995)
[4] Axelrod, R.: Effective choice in the Prisoner's Dilemma. Journal of Conflict Resolution 24, 3–25 (1980)
[5] Bapna, R., Goes, P., Gupta, A.: Insights and Analyses of Online Auctions. Communications of the ACM 44(11), 43–50 (2001)
[6] Cassady, R.J.R.: Auctions and Auctioneering 58(4), 959–963 (1968)
[7] Camerer, C.F.: Behavioral Game Theory: Predicting Human Behavior in Strategic Situations
[8] David, E., Rogers, A., Schiff, J., Kraus, S., Jennings, N.R.: Optimal design for English auctions with discrete bid level. In: Proceedings of Sixth ACM Conference on Electronic Commerce (EC 2005), Vancouver, Canada, pp. 98–107 (2005)
[9] Fudenberg, D., Tirole, J.: Game Theory
[10] He, M., Leung, H., Jennings, N.R.: A fuzzy logic based bidding strategy for autonomous agents in continuous double auctions. IEEE Trans. on Knowledge and Data Engineering 15(6), 1345–11363 (2003)
[11] http://www.authorstream.com/presentation/Darshak89-264364-game-theory-economics-eco-mix-straregy-strategy-competition-education-ppt-powerpoint/
[12] Green, K.C.: Forecasting Decision. In: Conflicts: Analogy, Game Theory, Unaided Judgment and Simulation Compared
[13] Lee, H.G.: Electronic Brokerage and Electronic Auction: The Impact of IT on Market Structures. In: Proceedings of the 29th HICSS, Los Alamitos, CA. Information Systems – Organizational Systems and Technology, vol. IV, pp. 397–406 (1996)
[14] Singh, M., Cassaigne, N., Bussey, P.: Bid Price – Calculating the Possibility of Winning
[15] Wang, M.-T., Wu, T.-S.: A Bidding Model Combined Game Theory and AI Paradigms
[16] Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press (1997)
[17] Sidnal, N.S., Manvi, S.S.: Bidding in English Auctions using Cognitive Agents in Mobile e-commerce (June 2011)
[18] Vragov, R.: Implicit Consumer Collusion in Auctions on the Internet. In: Proceedings of the 38th Hawaii International Conference on System Sciences (2005)

[19] Klein, S., O'Keefe, R.M.: The Impact of the Web on Auctions: Some Empirical Evidence and Theoretical Considerations. International Journal of Electronic Commerce 3(3), 7–20 (1999)
[20] http://www.cse.iitk.ac.in/users/arnabb/GameTheory_popular.ppt
[21] http://www.secowinet.epfl.ch/slides/AppB-GameTheory.ppt
[22] http://www.snn.ru.nl/bnaic/tutorial_kearns.ppt

Web Personalization Based on Short Term Navigational Behaviour and Meta Keywords

Siddu P. Algur[1], Nitin P. Jadhav[1], and N.H. Ayachit[2]

[1] Department of Information Science and Engg.,
B.V.B. College of Engg. and Tech.,
Hubli, Karnataka, India
{Siddu_p_algur,nit1jadhav.nitin}@gmail.com
[2] Department of Physics,
B.V.B College of Engg. and Tech.,
Hubli, Karnataka, India
nhayachit@gmail.com

Abstract. The amounts of information residing on web sites make users' navigation a hard task. To address this problem, web Personalization concept came into existence, which is based on similar users' navigational patterns mined from past visits. Vast techniques are proposed for Web Personalization using Web usage mining but it lacks in recommendation of Web pages which are relative to the user's interest and recommendation of Web pages when the new Web pages which are not in the Web log files are accessed. To solve this problem a novel approach for Web personalization is proposed in this paper which consists of two phases, offline phase and online phase. In offline phase the aggregate usage profile is created by processing the web logs and clustering the sessions obtained from web logs using 'Unweighted Pair Group Using Arithmetic averages' method. And meta keywords of all the URLs present in Web logs are generated which will be used in recommendation process. In online phase the short term navigational behaviour of the user's is used to recommend the related Web pages to the user even though unvisited or new URL is accessed for the first time. The experiment is performed on real data which proved that the system is performing well in recommendation.

Keywords: Web personalization, recommendations, meta keywords, Web usage mining, Web log.

1 Introduction

The World Wide Web contains huge amount of data, more number of web pages and connection between these web pages. There are more than 150 million web sites are online at present. The World Wide Web has become a platform to retrieve as well as mine useful knowledge. Due to the presence of huge amount of data, dynamic content and unstructured nature of the Web, the Web users always getting into an ocean of information and facing the problem of information overload or they may get unwanted information. As a result the Web data research has been facing a lot of

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 773–786.
springerlink.com

challenges. One of the ways to deal with this kind of challenging is to analyze the navigational patterns of the user interacting with the one or more websites. Analysis of browsing pattern of user can help to predict some web pages based on user needs. This is called 'web personalization' [9]. Historically, the conception of discovering useful data has been given a variety of names like data mining, data archaeology and data pattern processing. It was Etzioni who first invented the term Web mining which is concerned with extracting knowledge from web data.

Web mining is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, 1) Web usage mining 2) Web content mining and 3) Web structure mining.

Web Usage Mining: Web usage mining [9]-[15] is a process of extracting useful information from server logs i.e. user's history. The web server logs are taken as input to the system which pre-processes it. Then data mining techniques are applied to the pre-processed data to find out the different groups of users with their area of interest which is called as 'Aggregate usage profile'. It is produced as input to recommendation process. The recommendation engine recommends the related web pages by comparing the short term navigational behaviour [6] called as 'Active user session' with 'Aggregate usage profile'. So finally, It is the process of finding out what users are looking for on the internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data, etc.

Web usage mining consists of applications like 1) Personalization 2) System Improvements 3) Site Modification 4) Business Intelligence. In this research we are concentrating on Personalization using Web Usage Mining.

2 Related Work

The Recommendation systems have been implemented by using different methods. The Collaborative Filtering (CF) method [2] is used in e-commerce systems. The customer provides the system with Preference ratings of products that may be used to build a customer profile or his or her likes and dislikes. Then, these systems apply statistical techniques or machine learning techniques to find a set of customers, known as neighbours, which in the past have exhibited similar behaviour. Usually, a neighbourhood is formed by the degree of similarity between the customers. Once a neighbourhood of similar customer is formed, these systems predict whether the target customer would like a particular product by calculating a weighted composite of the neighbour's rating of that product or generate a set of products that the target customers is most likely to purchase by analyzing the products the neighbours purchased. This system is also known as the nearest neighbour CF-based recommender system. The disadvantage of the system is, it requires explicit feedback provided by the user or user ratings on products. To overcome these limitations, recent research has focused on Web Usage Mining approach for Web Personalization [9]. This type of solution generates patterns or usage profile based on implicit information provided by the user such as page visits of the user, duration of page visit etc. Various data mining techniques are applied offline to find out aggregate usage profile which can be used to provide recommendation.

Recent researchers proposed and implemented various recommender systems for Web Personalization using Web Usage Mining. In [1] the system has been divided into two components, online components and offline components. The offline component is used to produce aggregate usage profile based on importance of the web pages visited by the different users. Here importance of a web page is calculated by considering only the frequency of page visit for all type of web pages which fails to find out the exact user's interest. The System [8] relies on the application of statistical and data mining methods to the Web log data, resulting in a set of useful patterns that indicate users' navigational behaviour. The Weighted Association Rules technique [4] is used for web personalization where more work is done to understand the user's interest. The weights are assigned to different Web pages in server logs to find out the importance of Web pages based on some parameters. In the on-line phase the interest degree of a web page to a user in the session is calculated. Matching score is calculated by comparing active user session and the history of that user and pages are recommended based on matching score of the session. These systems did not perform well in recommending web pages. To solve this problem hybrid approach to Web Personalization is proposed in [3] which is combination of two approaches, CF [2] and Content Based Filtering (CBF) [5]. The system [7] heavily uses data mining techniques, thus making the personalization process both automatic and dynamic. Specifically, they discussed the necessary data preparation tasks for pre-processing of Web usage logs and grouping URL references into units of semantic activity called user transactions. Then the technique for extracting aggregated usage knowledge is described which is based on transaction clustering, which would be suitable for the purpose of Web personalization.

By considering all these work, there exist gaps in identifying the importance of Web pages to find out the user's interest and also in recommendation of related Web pages to the user. To fill all these gaps, we have designed a system which understands the importance of Web pages based on the type of page and Meta keyword are used to filter the unwanted Web pages and finally recommends the filtered or related Web pages.

3 Proposed Method

To recommend web pages without using the explicit feedback from the user, the system must understand the user's interest dynamically, which can be done by calculating significance of the web pages accessed by the user. The architecture of the proposed system is shown in Fig. 1.

The system is divided into two parts, 1) offline and 2) online. In offline part the raw weblogs are pre-processed to eliminate the useless records. Then the processed log file is divided into different sessions based on IP address, date and time. "Unweighted Pair-group Method using Arithmetic averages" clustering method is used to cluster the sessions, where each cluster consists of session of similar access patterns. After clustering aggregate usage profile is created. Finally in the offline part, meta keywords of web pages present in the web logs are generated by using the Algorithm 1 which is used in recommendation process. In online part, the active user session is matched with aggregate usage profile to recommend the web pages.

The results are filtered to recommend the relevant web pages by processing the meta keywords. A novel approach is proposed for recommendation when the previously unvisited web page is visited.

3.1 Data Cleaning

Data pre-processing technique has been discussed in detail in [16, 17]. The Web server logs of BVBCET organization is taken as input to the system which is in 'squid' log format. The file contains the date and time in UNIX timestamp format which will be converted to normal date format for later analysis. The log file contains a single record for each access of Web page visited by the user along with the bytes transferred, IP address, Time spent on each page, URL, status of access and other useful information. It also stores a record if the web page is not successfully accessed. Such records will not be used to identify the user's interest. Hence these records will be removed from the log file. We can identify such record by considering status value of the record. If the value is from 200 to 206, 207 and 226 etc then it is visited successfully. After removing unwanted records, the component produces a file with only valid records which will be given as input to the next component.

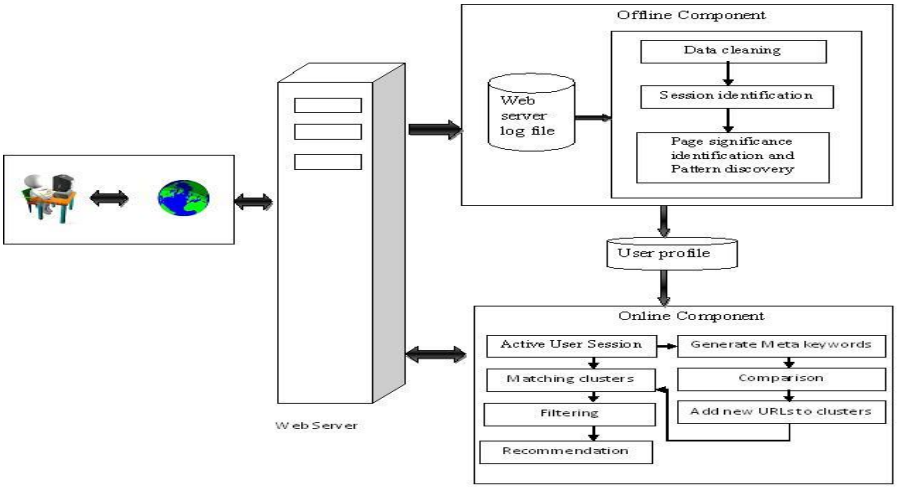


Fig. 1. Architecture diagram of proposed system for Web Personalization

3.2 Session Identification

This component takes the pre-processed Web log and creates session files based on some rules. A session file consists of a sequence of user's request for pages $P = \{p_1, p_2, p_{3,\ldots,} p_n\}$ assigned with weights $W = \{w_1, w_2, w_3, \ldots, w_n\}$ and a set of m sessions $S = \{s_1, s_2, s_3, \ldots, s_m\}$ where each $s_i \in S$ is a subset of P.

The sessions are created based on IP address, date and time of page access. Every 60 minute page access on a single computer is considered as one session. Each session contains URLs visited in that session and corresponding weights.

3.3 Identifying the Significance of a Web Page

We propose a weighting schema which is used to calculate weight for Web page present in Web log to extract the user's interest. We consider frequency of a page to extract the user's interest. Frequency is the number of times the page is visited. It is common that the page with highest frequency is of stronger interest to the user.

The following equation is used to find out the importance of a Web page.

$$\text{weight(p)} = \frac{\text{Number of visit(p)}}{\sum_{Q \epsilon T} \text{Number of visit(p)}} \quad (1)$$

Let us, consider there are five pages in a Web log and their frequency is shown in Table1. First row represents the Web pages in a session and second row represents the frequency of occurrence of corresponding Web page.

Table 1. Page frequency

1	2	3	4	5
2	3	0	5	3

Table 2 shows the weight of each Web page present in a session which is shown in Table1 using Equation (1). First row indicates the Web pages and second row indicates the weight/importance of the corresponding Web pages.

Table 2. Page weights

1	2		4	5
0.1 54	0.2 31		0.3 85	0.2 31

3.4 Pattern Discovery

The next component in the offline part is to determine sessions with "similar" navigational patterns from user session file. To cluster the user sessions more effectively, we used 'Unweighted Pair Group Using Arithmetic averages' method. Equation (2) is used to find out the distance between two sessions.

$$d(S_i, S_j) = \frac{common(S_i, S_j)}{\min(S_i, S_j)} \quad (2)$$

Where, $common(S_i, S_j)$ is the total number of URLs which are common between session S_i andS_j. $\min(S_i, S_j)$ is the minimum of total number of URLs present in session S_i and S_j.

Each cluster represents the several sessions with similar usage pattern. The user's interest is determined by creating aggregate usage profile using Equation (3).

$$wt(p, c_i) = \frac{1}{nc} \sum_{s \in c} W_p^s \tag{3}$$

Where W_p^s represents the weight of the page in sessions $\in c$, nc represents the number of sessions in cluster c and c_i represents the i^{th} cluster.

Finally, Aggregate usage profile consists of r clusters$C = \{c_1, c_2, c_3, \ldots, c_r\}$ where each $c_i \in C$ is subset S.

3.5 Meta Keywords Generation

Each page p_i present in P is fetched from database and the content of the page is accessed to gain knowledge. Instead of fetching the meta keywords of the web pagep_i, here the meta keywords are generated in order to learn the concept present in that page. There are two reasons to generate the meta keywords instead of fetching the already present meta keywords.

1. All web pages present on internet does not have the meta keywords. Some of the pages may have and some of the pages may not have. So we cannot rely on this information.
2. It is observed that meta keywords of the Web page which are generated by the Web page developer are not correct most of the times or irrelevant to the content of the page, which can be done to achieve Search Engine Optimization (SEO).

To address these two reasons an algorithm is proposed to generate the meta keywords of the web page. The algorithm1 is generating the meta keywords for a given web page which are very relative to the content of the web page. This concept is used in recommendation process which helps to recommend the relevant web pages to the user based on his/her interest.

3.6 Recommendation

The recommendation process is an online phase. In order to recommend web pages to the user, we need user's interest. The short history of navigational behaviour of user is considered as user's interest. The short navigational behaviour is nothing but last 'n' page visited by the user in the session is called as active user session. The active user session is used to recommend the web pages to the user. The active user session is also called as Sliding Window because the window slides forward when user visits the next page, so that it contains only last 'n' pages visited by the user. There are two different situations for recommendation process.

Algorithm 1. Meta keyword generation

```
Input: A variable url containing URL of a web page
Output: An array final_keywords[0…n] containing generated meta keywords of the web page
frequency ← 0
content←fileGetContents(url)
content←removeJavaScript(content)
content←convertToPlainText(content)
keywords[]←split the content by space
for i←0 to count(keywords) do
    keywords[i]←strToLower(keywords[i])
end for
new_keywords[]←uniqueKeywords(keywords[])
new_keywords[]←removeStopWords(new_keywords[])
for i←0 to count(new_keywords) do
    for j←0 to count(keywords) do
        if new_keywords[i]=keywords[j] then
            frequency ← frequency + 1
        end if
    end for
    new_keywords_value[i]←frequency
    frequency←0
end for
for i←0 to count(new_keywords) do
    for j←1 to count(new_k eywords) do
        if(new_keywords[i]<new_keywords[j] then
            swap(new_keywords[i],new_keywords[j])
            swap(new_keywords_value[i], new_keywords_value[j])
        end if
    end for
end for
j←0
for i←0 to count(new_keywords) do
    val← new_keywords_value[i]/new_keywords_value[0]
    if val>threshold then
        final_keywords[j]←new_keywords[i]
        j←j+1
    end if
end for
return final_keywords[]
```

First case is when the web page contained in the active user session, that is the web page visited recently is present in the database or which is already visited in past, then the similarity of active user session with aggregate user profile is calculated using similarity measure called cosine similarity. The active user session is represented as S_i and cluster is represented as C_k then similarity is calculated by the following equation.

$$sim(S_j, C_k) = \frac{\sum_{i=1}^{n} W_{i,j} \cdot W_{i,k}}{\sqrt{\sum_{i=1}^{n} W_{i,j}^2} \cdot \sqrt{\sum_{i=1}^{n} W_{j,k}^2}} \quad (4)$$

Where $W_{i,j}$ represents the weight of page i in active user session j and $W_{i,k}$ represents the weight of page i in cluster k.

When the above formula is applied to match the active user session and aggregate user profile, the outcome is the clusters with the value which indicates the similarity measure. The clusters which are greater than the threshold value u_c are selected as the matching clusters. The URLs in these clusters are used to recommend the web pages to the user.

Second case is when the URL in active user session is not yet visited in the past or the URL is newly added to website then our aim is to find out the concept of the content of the new web page and if it is relevant to the any of the clusters then we should add this URL to those matching clusters and other URLs of those clusters will take part in recommendation process. To achieve this task the meta keywords of the new URL is generated using the algorithm1 and these keywords are compared with the meta keywords of each clusters to calculate the similarity measure. The cluster similarity measure which is greater than the threshold value are considered as matching clusters. Now the new URL is added to the database of matching clusters. The active user session is represented as S_i and cluster is represented as C_k. Then similarity measure is calculated as follows:

$$Sim(S_i, C_k) = \frac{\sum_{l=1}^{k} W_m}{\sum_{j=1}^{n} W_j} \quad (5)$$

Where W_m represents the weight of matching meta keyword and W_j represents the weight of jth meta keyword. 'n' is the total number of meta keywords in new URL.

The identification of matching cluster using the above two Equations would help to determine the significant clusters. But to identify the significant pages, still we need to filter the results. To remove the irrelevant pages, the meta keywords of URLs of matching clusters is compared with the meta keywords of URLs in active user session. Now, only the matching URLs are recommended which may be very useful to the user.

4 Experimental Results and Discussion

The Web server logs of BVBCET College are used for this research. After Data Pre-processing and analysis, the log contains 114 different user sessions which are clustered into 47 different clusters and it contains 1531 unique URLs or Web pages.

For example consider the sample set of URLs which are processed from Web server log which is shown in Table 3. Here sliding window is used to store the last 'n' visited pages. Table 5 represents the page visits in the sliding window based on which recommendations will be done. In active user session already visited pages are accessed as well as new pages are accessed by the user. New pages which are not in the web logs are shown in Table 4.

First case is Recommendation of Web pages when the URL present in a Web log is accessed by the user. For example, consider the user visits the URL 1 and 2 which is

shown in Table 5. Then these two URLs are compared with the aggregate usage profile using the equation (4). The clusters which are above the threshold value are used for recommendation.

Second case is Recommendation of Web pages when the new URLs which are not in the Web logs are accessed. For example, consider the user visits URLS 1, 2, n1, n2 and 3, Where URLs 1, 2 and 3 are already present in the Web log. URLs n1 and n2 are visited first time. So now our aim is to include these two URLs in any of the clusters to which it matches. In order to do this, the system must understand the contents of new Web page. The meta keywords of the new Web page n1 are generated using the Algorithm 1. These meta keywords are compared with the meta keywords of the clusters present in the Web log using equation (5).The matching clusters for n1 are cluster 1,3,8,13 and 14 which is shown in Table 6. Now n1 will be added to the matching clusters. Similar procedure will be applied for URL n2. So URL n2 will be added to cluster 1,3,8 and 14. Now the active user session which contains URLs 1, 2, n1, n2 and 3 will be compared with aggregate usage profile using equation (4). The clusters which are above the threshold value are cluster 1, 3, 8, 11, and 13 which is shown in Fig. 2. The aggregate usage profile of cluster1, 3, 8, 11 and 13 is shown in Fig. 3. From Fig. 3, we can come to know that which pages can be recommended to the user. The URLs which are greater than the threshold value will be further filtered to recommend to the user. The final set of URLs after comparing Meta keywords of URL present in the database with the Meta keywords of the URLs present in the active user session is shown in Fig. 4. The URLs present in Fig. 4 are very relative to the needs of user. Our method proves that it recommends the Web pages which are relative to the user's interest The Table 8 shows the recommended set of pages for active user session1.

Table 3. Sample set of web log URLs

No.	URL	Frequency	Weight
1.	http://www.cprogramming.com/	2	0.1
2.	http://www.cprogramming.com/begin.html	2	0.1
3	http://home.java.net	2	0.1
4	http://www.android.com	12	0.6
5	http://www.cplusplus.com/doc/tutorial	15	0.75
6	http://www.indiabix.com/engineering	20	1
7	http://www.cprogramming.com/advtutorial.html	10	0.5
8	http://en.wikipedia.org/wiki/C_(programming_language)	12	0.6
9	http://www.microcontroller.com	16	0.8
10	http://www.topsite.com/best/microcontroller	10	0.5
11	http://en.wikipedia.org/wiki/Data_structure	12	0.6
12	http://www.cpp-home.com	3	0.15
13	http://www.cplusplus.com/	4	0.2
14	http://www.cplusplus.com/info/description/	3	0.15
15	http://www.cpp-home.com/archives/category/surveys	15	0.75

Table 3. (*Continued*)

16	http://c-faq.com/~scs/cclass/notes/top.html	11	0.55
17	http://www.microchip.com	5	0.25
18	http://electronicdesign.com	9	0.45
19	http://www.topsite.com/best/microcontroller	12	0.6
20	http://www.projectsof8051.com	6	0.3
21	http://in.answers.yahoo.com/question/index?qid=20110904052050AAdgIPN	18	0.9
22	http://www.cpp-home.com	12	0.6
23	http://www.webmator.com/microcontroller/microcontroller	7	0.35
24	http://www.oracle.com/technetwork/java/javase/downloads/jdk6-jsp-136632.html	10	0.5
25	http://www.youtube.com/watch?v=QKGPOB5BHmI	12	0.6
26	http://download.cnet.com/Java-Development-Kit/3000-2218_4-12091.html	5	0.25
27	http://en.wikipedia.org/wiki/Operating_system	19	0.95
28	http://www.apnaghar.co.in	12	0.6
29	http://developer.android.com/guide/basics/what-is-android.html	10	0.5
30	http://www.cse.iitd.ernet.in/~sak/courses/cdp/slides.pdfe.flv	8	0.4
31	http://www.indiastudychannel.com/resources/116105-Best-site-for-Java-related-all-Tutorials.aspxe.flv	7	0.35
32	http://www.architecturaldesigns.com/	5	0.25
33	https://www.engr.usask.ca/classes/EE/331/list_of_proj_web.pdf0/Basha%20DvdRip%20Telugu%20Movie/Basha%20DVDrip%20X264.mp4	5	0.25
34	http://www.similarsitesearch.com/alternatives-to/pjrc.com0/Basha%20DvdRip%20Telugu%20Movie/Basha%20DVDrip%20X264.mp4	2	0.1
35	http://www.buzzedu.com/wp-content/themes/buzznew/images/inner-explore_banner.swf	14	0.7
36	http://www.buzzedu.com/wp-content/themes/buzznew/images/hc-home.swf	11	0.55
37	http://d13.zedo.com/OzoDB/b/2/1002817/V1/airtelneobroadband300x25012.swf?	12	0.6
38	http://developer.android.com/sdk/index.html	10	0.5
39	http://www.javalobby.org/forums/thread.jspa?threadID=16001&tstart=0	5	0.25
40	http://www.indiastudychannel.com/resources/116105-Best-site-for-Java-related-all-Tutorials.aspx	5	0.25
41	http://www.thecoolist.com/landscape-architecture-designs-10-modern-masterpieces/	10	0.5
42	http://www.infoworld.com/d/cloud-computing/what-cloud-computing-really-means-031	11	0.55
43	http://www.javadb.com/	19	0.95
44	http://auto.howstuffworks.com/	11	0.55
45	http://www.android.com/media/	9	0.45

Table 4. Sample set of unvisited URLs

No.	URL
n1	http://cquestionbank.blogspot.com/2011/08/c-objective-questions-and-answers-pdf.html
n2	http://www.freelancer.com/job-search/microcontroller-based-wireless-home-automation-latest/
n3	http://developer.android.com/guide/publishing/publishing.html
n4	http://mobile.tutsplus.com/tutorials/android/publish-to-android-market/

Table 5. Active user session

Session	Window	Active User Session (Page visits)				
1	1	1	-	-	-	-
1	2	1	2	-	-	-
1	3	1	2	n1	-	-
1	4	1	2	n1	n2	3
1	5	1	2	n1	n2	3
1	5	2	n1	n2	3	n3
1	5	n1	n2	3	4	n3
1	5	n2	3	4	n3	n4

Table 6. Matching clusters for new URLs

New URL	Matching Clusters
n1	1,3,8,13,14
n2	1,3,8,14
n3	3,4,10,11
n4	3,4,10,11,14

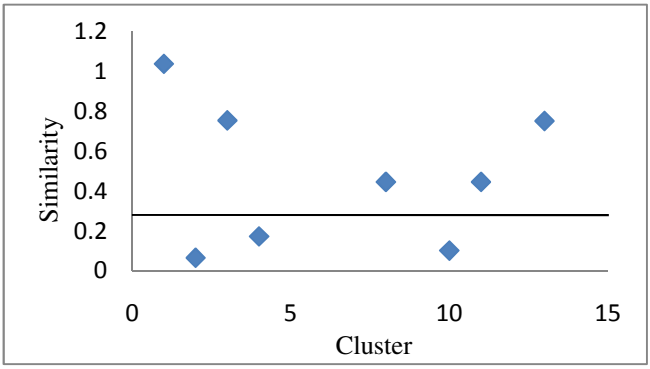


Fig. 2. Cluster similarity for URLs 1, 2, n1, n2 and 3

Table 7. Matching clusters

Session	Active Session	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
1	1	0.440	0.150												
1	1, 2	0.761	0.107												
1	1, 2, n1	0.999	0.087	0.577					0.577			0.577		0.577	0.577
1	1, 2, n1,n2	1.161	0.076	0.841					0.5			0.5		0.841	0.841
1	1, 2, n1,n2,3	1.038	0.067	0.755	0.174				0.447		0.102	0.44		0.752	
2	2, n1,n2,3,4	0.974		0.755	0.460				0.447	0.369	0.103	0.447		0.752	0.447
2	n1,n2,3,4,n3	0.752		1.022	0.757	0.447	0.447	0.447	0.447	0.37	0.47	0.752		1.019	0.752
2	n2,3,4,n3,n4	0.447		1.022	1.022	0.447	0.447	0.447		0.37	0.47	0.447	0.447	1.02	1.02

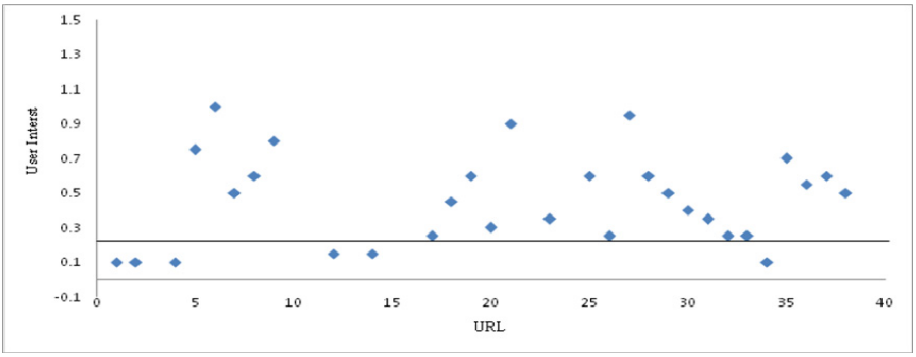


Fig. 3. Aggregate usage profile of cluster 1, 3, 8, 11 and 13

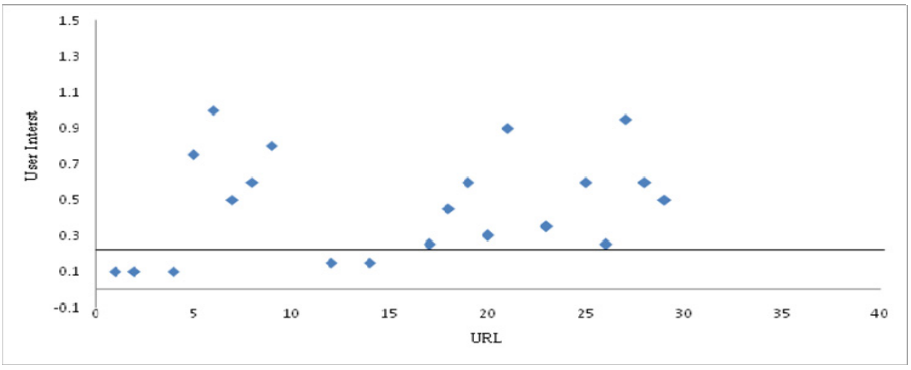


Fig. 4. Recommended pages

Table 8. Recommendation set

Session	Active Session	Recommended Pages
1	1	6,5,8,7,2
1	1, 2	5,9,11,7,10,2
1	1, 2, n1	6,5,15,8,16,7,13,12,14,2
1	1, 2, n1,n2	6,21,9,5,19,22,7,18,23,20,17
1	1, 2, n1,n2,3	6,27,21,9,5,19,28,8,7,29,18,23,26,17,20,25
2	2, n1,n2,3,4	27,21,9,5,8,25,28,10,24,29,38,20,17,26,39,40
2	n1,n2,3,4,n3	43,21,9,25,28,10,42,44,24,29,38,41,17,26,39,2
2	n2,3,4,n3,n4	9,5,8,22,25,28,10,42,24,29,41,45,23,26,39,40,2

References

[1] Sumathi, C.P., Padmaja Valli, R., Santhanam, T.: Automatic recommendation of web pages in web usage mining. International Journal on Computer Science and Engineering (IJCSE) 02(09), 3046–3052 (2010)

[2] Wang, T., Ren, Y.: Research on Personalized Recommendation Based on web Usage Mining Using Collaborative Filtering Technique, vol. 6(1) (January 2009)

[3] Thakur, M., Jain, Y.K., Silakari, G.: Query based Personalization in Semantic Web Mining. International Journal of Advanced Computer Science and Applications (IJACSA) 2(2) (February 2011)

[4] Forsati, R., Meybodi, M.R., Ghari Neiat, A.: Web page personalization based on weighted association rules. IEEE (2009)

[5] Eirinaki, M., Lampos, C., Paulakis, S., Vazirgiannis, M.: Web Personalization Integrating Content Semantics and Navigational Patterns (2004)

[6] Eirinaki, M., Lampos, C., Paulakis, S., Vazirgiannis, M.: Web Personalization Integrating Content Semantics and Navigational Patterns (2004)

[7] Mobasher, B.: WebPersonalizer: A Server Side Recommender System Based on Web Usage Mining

[8] Zhou, B., Hui, S.C., Fong, A.C.M.: Web Usage Mining for Semantic Web Personalization

[9] Eirinaki, M., Vazirgiannis, M.: Web mining for Web personalization. ACM Transaction on Internet Technology 3(1), 1–27 (2003)

[10] Srivastava, J., Cooley, R., Deshpande, M., Tan, P.: Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations 1(2), 2–23 (2001)

[11] Colley, R., Mobasher, B., Srivastava, J.: Web mining Information and pattern discovery on the World Wide Web. In: Proceedings of IEEE Interanational Conference Tools with AI, pp. 558–567 (1997)

[12] Fu, X., Budzik, J., Hammond, K.: Mining navigation history for recommendation. In: Proceedings of Fifth International Conference on Intelligent User Interfaces, pp. 106–112 (2000)

[13] Grey, M., Haddad, H.: Evaluation of Web usage mining approaches for user's next request prediction. In: Proceedings of the Fifth ACM International Workshop on Web Information and Data Management, pp. 74–81 (2003)
[14] Mobasher, B., Colley, R., Srivastava, J.: Automatic Personalization based on Web Usage Mining. Communication of ACM 43(8), 142–151 (2000)
[15] Mulevenna, M.D., Anand, S.S., Buchner, A.G.: Personalization on the net using Web mining. Communications of the ACM 43(8), 123–125 (2000)
[16] Sumathi, C.P., Padmaja Valli, R., Santhanam, T.: An Application of Session based clustering to Analyze Web Page of User Interest from Web Log Files. Journal of Computer Science, 785–793 (2010)
[17] Suresh, R.M., Padmajavalli, R.: Overview of data preprocessing in data and Web Usage Mining. In: Proceedings of the IEEE International Conference on Digital Information Management, Banglore, December 6-12, pp. 193–198 (2006)

A Parallel Fuzzy C Means Algorithm for Brain Tumor Segmentation on Multiple MRI Images

Aarthi Ravi, Ananya Suvarna, Andrea D'Souza,
G. Ram Mohana Reddy, and Megha

NITK Surathkal, Mangalore, India
aartravi@gmail.com

Abstract. The Fuzzy C Means (FCM) algorithm has been extensively used in medical image segmentation. But for large data sets the convergence of the FCM algorithm is time consuming and also requires considerable amount of memory. In some real time applications, like Content Based Medical Image Retrieval (CBIR) systems, there is a need to segment a large volume of brain MRI images offline. In this paper, we present an efficient method to cluster data points of all the images at once. The gray level histogram is used in the FCM algorithm to minimize the time for segmentation and the space required. A parallel approach is then applied to further reduce the computation time. The proposed method is found to be almost twice as fast as conventional FCM.

1 Introduction

In the field of medical diagnosis a variety of imaging techniques is presently available, such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI). MRI provides good contrast between the different soft tissues of the body, which makes it especially useful in imaging the brain. Image Segmentation is a process of partitioning an image into non-overlapped, consistent regions which are homogeneous with respect to some properties such as intensity, color and texture [1]. It is a vital step in analysis of medical images for computer aided diagnosis. The main objective of image segmentation in brain MRI images is to isolate a brain tumor from other regions of the brain.

In certain real time applications that support computer aided diagnosis like the CBIR systems, there is a need to process and analyze a large number of medical images. The processing of each image is time consuming owing to the large size of the image itself. Thus processing of large volumes of data must be done offline. In this paper we present a parallel histogram based fuzzy c means approach to efficiently cluster data points of all the MRI images together at once and segment the images to obtain the tumors.

2 Related Work

The computation of conventional FCM algorithm for the iterative operation is time consuming for large data sets and has a high amount of memory requirement for the

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 787–794.
springerlink.com

membership matrix. Modifications to overcome the drawbacks of FCM have been proposed by researchers.

Moh'd Belal Al-Zoubi et al. [2] have proposed a fast fuzzy clustering algorithm that is based on eliminating data points with a membership value lower than a threshold value. The choice of the threshold value is based on experimentations; hence the algorithm is not very efficient. Ming-Chuan Hung and Don-Lin Yang [3] have proposed a faster FCM algorithm that uses a two phase approach. Though this approach reduces computation time, additional memory is required for k-d tree and storing additional information like statistical information of the patterns in each block. S.R. Kannan et al. [4] have proposed a center knowledge method in order to reduce the running time of proposed algorithm. But the drawback here is the memory required for the distance table that is dependent on the size of the image.

The algorithm proposed by Ye Xiu Qing et al. [5] uses the gray level histogram in the FCM algorithm to minimize the time for segmentation and the space required for the membership matrix. The algorithms proposed by Weiling Cai et al. [6] and S. Zulaikha Beevi and M. Mohamed Sathik [1], speed up the conventional FCM and significantly reduce the execution time by clustering on grey level histogram rather than on pixels. The proposed methodologies are found to be efficient and robust to noise. The histogram based approach is also adopted by He Yangming and Dai Shuguang [8] and achieves great speed up. The method proposed by Arpit Srivastava et al. [9] uses a membership suppression mechanism which creates competition among clusters to speed up the clustering process. The drawback here is that the execution time depends on the size of the dataset. S. Rahimi et al. [7] and S. Murugavalli and V. Rajamani [8] have proposed parallel FCM based approaches for image segmentation. The parallel algorithms proposed divide all the image pixels equally among the processors so that each processor handles n/p data points (n is the number of pixels and p is the number of processors involved in the computation). Thus, the processing time reduces significantly.

In this paper we adopt the histogram based approach [5], thus reducing the data points to the number of gray levels in the image instead of the number of pixels. The histogram of all images is computed and the membership matrix is initialized based on all the histograms. Thus, the FCM has to be applied only once to cluster all the images. In this paper we also modify the parallel approach [7] by assigning each cluster to different processors. Each processor computes its cluster center and updates the membership matrix after each iterative operation in the FCM algorithm.

3 Proposed Methodology

3.1 Conventional FCM

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. It is based on minimization of the following objective function:

$$J(U,C)=\sum_{i=1}^{N}\sum_{j=1}^{C}(u_{ij})^{m}\parallel x_i-c_j\parallel^2 \tag{1}$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data and c_j is the d-dimension center of the cluster. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above. This iteration will stop when $\max_{ij}\left\{u_{ij}^{(k+1)}-u_{ij}^{(k)}\right\}<\varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps.

Steps:

1. Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$
2. At k-step calculate the centers vectors $C^{(k)}=[c_{ij}]$ with $U^{(k)}$

$$c_j=\frac{\sum_{i=1}^{N}u_{ij}{}^{m}\cdot x_i}{\sum_{i=1}^{N}u_{ij}{}^{m}} \tag{2}$$

3. Update $U^{(k)}$, $U^{(k+1)}$

$$u_{ij}=\frac{1}{\sum_{k=1}^{C}\left(\frac{\parallel x_i-c_j\parallel}{\parallel x_i-c_k\parallel}\right)^{\frac{2}{m-1}}} \tag{3}$$

4. If $\parallel U^{k+1)}-U^{(k)}\parallel<\varepsilon$, then STOP; otherwise return to Step 2.

The convergence of the conventional FCM algorithm is time consuming which makes it impractical for image segmentation.

3.2 Histogram Based FCM

A single gray level histogram comprising of multiple brain MRI images is computed. This histogram is used in the FCM algorithm, which enhances the speed of segmentation and at the same time reduces the space required for the membership matrix. The objective function is given by

$$J(U,C)=\sum_{l=1}^{L}\sum_{i=1}^{v}(u_{il})^{m}H(l)d^2(l,c_i) \tag{4}$$

where, H is the histogram of all images comprising of L gray levels. The computation of membership degrees of $H(l)$ pixels is minimized to that of only one pixel with l as grey level value.

The membership function u_{ij} and center c_i for histogram based FCM can be calculated as

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_i - c_j \|}{\| x_i - c_k \|} \right)^{\frac{2}{m-1}}} \tag{5}$$

$$c_j = \frac{\sum_{l=1}^{L} u_{il}{}^{m} .H(l)l}{\sum_{i=1}^{L} u_{il}{}^{m}} \tag{6}$$

where l is the gray level ranging from 0 to 255.

3.3 Proposed Parallel Histogram Based Approach to FCM

The proposed algorithm computes the histogram H of multiple MRI images that need to be segmented. The performance of Histogram Based FCM can be further enhanced by distributing computation and main memory usage. Thus, each cluster is assigned to a different processor. Each processor corresponding to a cluster computes its center and updates the corresponding row of the membership matrix in each iteration of the modified FCM algorithm.

In this paper, the brain is segmented into 4 clusters. Each processor p_j corresponds to a cluster c_j where j is between 1 and 4. The initiating processor P assigns each cluster c_j corresponding to a row r_j of the membership matrix U is assigned to each processor p_j. Each processor p_j computes the center of its cluster as follows:

$$c_j = \frac{\sum_{l=1}^{L} u_{il}{}^{m} .H(l)l}{\sum_{i=1}^{L} u_{il}{}^{m}} \tag{7}$$

Each processor p_j sends the computed centers back to the initiating processor P. The processor P then reassigns each row r_j of the membership matrix to processor p_j and sends the centers vector to each processor p_j. Each processor p_j updates row r_j of the membership matrix U.

$$u_{jl} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\| x_j - c_l \|}{\| x_j - c_k \|} \right)^{\frac{2}{m-1}}} \tag{8}$$

where j corresponds to the cluster or processor, l ranges from 0 to 255.

Proposed Methodology to Cluster Multiple Brain MRI Images

Step 1: Input Multiple Brain MRI Images
Step 2: Compute a single Histogram H of all the MRI images
Step 3: Initialize the membership matrix $U = [u_{jl}]$
Step 4: Initiating processor P assigns each cluster c_j to processor p_j
Step 5: Each processor p_j computes the center of its cluster
Step 6: Each processor sends the computed center c_j back to the initiating processor P
Step 7: Initiating processor P sends the center vector c to each processor p_j
Step 8: Each processor updates the row r_j of the membership matrix corresponding to its cluster c_j
Step 9: Each processor p_j sends the computed row r_j back to the initiating processor P
Step 10: If $\| U^{k+1)} - U^{(k)} \| < \varepsilon$ then, go to Step 4
Step 11: Output the Segmented Results

3.4 Extracting Tumor from Segmented Cluster

The cluster with the largest center value is chosen. All the points i.e. gray level values belonging to this cluster are stored in array C. Each brain MRI from the large set of MRI images is considered. The pixel values of points with gray value equal to gray values contained in array C to 255. Then perform opening and closing operations on the image using a disc as structural element. Check if blobs are present in the image. If the quality of the image is poor then the tumor may not be present in the cluster. Thus, the modified FCM algorithm must be applied again on the image if blobs are not present. If blobs are present then find the largest blob and set the pixel values of other blobs to zero. Fig. 2 shows results for extraction of tumors.

Steps of the Proposed Methodology to Cluster Multiple Brain MRI Images

Step 1: Input Multiple Brain MRI Images
Step 2: Set the pixel values of points with gray value equal to gray values con -tained in array C to 255
Step 3: Initialize the membership matrix $U = [u_{jl}]$
Step 4: Perform opening and closing operations on the image using a disc as structural element
Step 5: If blobs are present then, find largest blob
Else, Apply Modified FCM on the Input Brain MRI and go to Step 2
Step 6: Set the pixel values of other blobs to zero
Step 7: Output the Extracted Tumor

4 Results and Discussion

The proposed algorithm is found to be of reduced space and time complexity as compared to the conventional FCM.

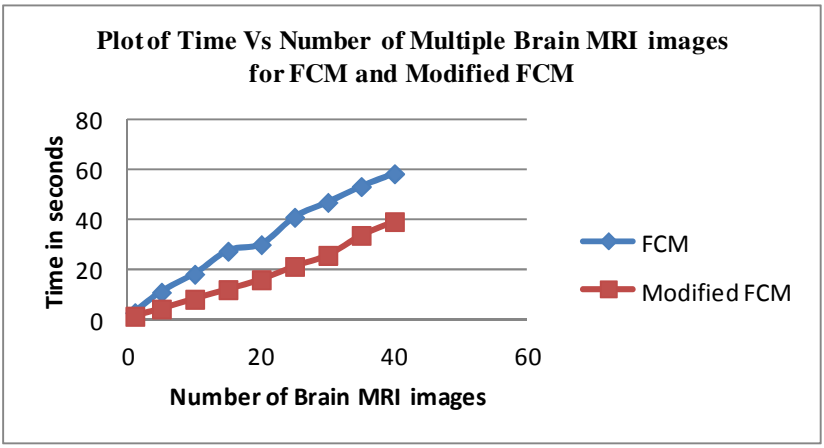


Fig. 1. Plot of Time Vs Number of Multiple Brain MRI images for FCM and Modified FCM

The proposed approach segments the brain MRI images into 4 clusters and uses 4 processors. The time required to cluster multiple MRI images is computed for the conventional FCM as well as the proposed parallel histogram based FCM and the results are compared. Fig. 1 shows the time taken to cluster the data set containing varied number of images.

The asymptotic efficiency of FCM and Modified FCM Algorithms are shown in Table 1.

Table 1. Space and Time Complexity of Clustering

Algorithm	Space Complexity (one image)	Space Complexity (n images)	Time Complexity (one image)	Time Complexity (n images)
FCM	O(dc)	O(ndc)	$O(dc^2i)$	$O(ndc^2i)$
Modified FCM	Cq	Cq	O(qci)	O(nqci) Ω(qci)

The asymptotic efficiency of the algorithm has following notations:

i number FCM over entire dataset
d number of data points
c number of clusters
q number of grey levels

The tumors from these clustered images are then extracted and separated from other parts of the brain using the method elucidated in the proposed methodology. The Modified FCM is applied to a sample of 3 brain MRI images. The output in Figure 2 shows the removal of brain portion from the cluster containing tumor.

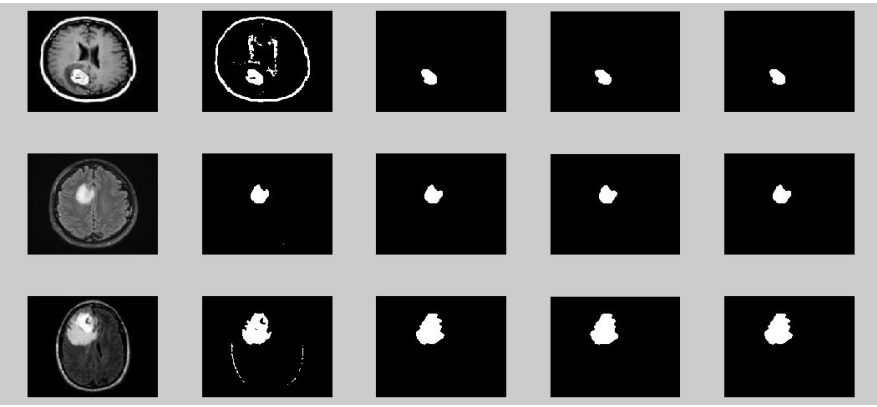


Fig. 2. Extraction of tumor from segmented cluster

5 Conclusion

The proposed parallel algorithm is found to be almost twice as fast as the conventional algorithm in spite of the overheads associated with parallelism. Large volumes of brain MRI images can be efficiently segmented at once using the proposed method. The proposed algorithm is independent of the size of the images to be segmented. Hence, it achieves a significant improvement over other parallel approaches which depend on the size of the image. The use of histogram based FCM in our algorithm reduces the space complexity significantly.

References

[1] Zulaika Beevi, S., Mohamed Sathik, M.: An Effective Approach for Segmentation of MRI images: Combining Spatial Information with Fuzzy C Means Clustering. Eur. J. on Scientific Res. 41(3), 437–451 (2010) ISSN 1450-216X

[2] Moh'd Belal, A.-Z., Hudaib, A., Shiboul, B.A.: A Fast Fuzzy Clustering Algorithm. In: Proceedings of the 6th WSEAS Int. Conf. on Artificial Intelligence, Knowl. Engineering (2007)

[3] Hung, M.-C., Yang, D.-L.: An Efficient Fuzzy C-Means Clustering Algorithm (2001); 0-7695-1 119-8/01 IEEE

[4] Kannan, S.R., Ramathilagam, S., Pandiarajan, R., Sathya, A.: Fuzzy Clustering Approach in Segmentation of T1-T2 brain MRI. International J. of Recent Trends in Engineering 2(1) (2009)

[5] Qing, Y.X., Hua, H.Z., Qiang, X.: Histogram Based Fuzzy C-Means Algorithm For Image Segmentation (1992), 0-8186-2920-7/92 IEEE; Cai, W., Chen, S., Zhang, D.-Q.: Fast and robust Fuzzy c-means clustering algorithms incorporating local information for Image segmentation. Pattern Recognition 40, 825–838 (2007)
[6] Rahimi, S., Zargham, M., Thakre, A., Chhillar, D.: A Parallel Fuzzy C-Means Algorithm for Image Segmentation. IEEE Annual Meeting of the Fuzzy Information 1, 234–237 (2004)
[7] Murugavalli, S., Rajamani, V.: A High Speed Parallel Fuzzy C-Mean Algorithm For Brain Tumor Segmentation. BIME Journal Volume 06(1) (2006)
[8] He, Y., Dai, S.: Application of Improved Fuzzy C- Means Clustering in Detecting Human Head. In: Sixth International Conference on Fuzzy Systems and Knowledge Discovery (2009)
[9] Srivastava, A., Asati, A., Bhattacharya, M.: A Fast and Noise-Adaptive Rough-Fuzzy Hybrid Algorithm for Medical Image Segmentation. In: IEEE International Conference on Bioinformatics and BioMedicine (2010)

Implementation of Web Search Result Clustering System

Hanumanthappa M. and B.R. Prakash

Bangalore University, Bangalore
hanu6572@hotmail.com,
brp.tmk@gmail.com

Abstract. Web search results clustering is an increasingly popular technique for providing useful grouping of web search results. This paper introduces a prototype web search results clustering engine that use the random sampling technique with medoids instead of centroids to improve clustering quality, Cluster labeling is achieved by combining intra-cluster and inter-cluster term extraction based on a variant of the information gain measure by using Modified Furthest Point First algorithm. M-FPF is compared against two other established web document clustering algorithms: Suffix Tree Clustering (STC) and Lingo, which are provided by the free open source Carrot2 Document Clustering Workbench. We measure cluster quality by considering precision , recall and relevance. Results from testing on different datasets show a considerable clustering quality.

1 Introduction

With the increase in information on the World Wide Web it has become difficult to find the desired information on search engines. The low precision of the web search engines coupled with the long ranked list presentation make it hard for users to find the information they are looking for. It takes lot of time to find the relevant information. Typical queries retrieve hundreds of documents, most of which have no relation with what the user is looking for. The reason for this is due to the user failing to formulate a suitable or specific enough query, and efforts have been made to use some form of natural language processing when processing a search query to try and understand the underlying concept the user is trying to get across. One solution to this problem is to enable more efficient navigation of search results by clustering similar documents together [1][2]. By clustering web search results generated by a conventional search engine, the search results can be organized in a manner to reduce user stress and make searching more efficient while leveraging the existing search capability and indexes of existing search engines [9].

2 Overview of M-FPF and Improving the FPF Algorithm

In this paper we improve the Furthest Point First algorithm from both the computational cost point of view and the output clustering quality. Since theoretically the FPF

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 795–800.
springerlink.com

algorithm as proposed by Gonzalez [12] is optimal (unless P = NP), only heuristics can be used to obtain better results and, in the worst case, it is not possible to go behind the theoretical bounds. We profiled FPF and analyzed the most computational expensive parts of the algorithm. We found that most of the distance computations are devoted to find the next furthest point. FPF clustering quality can be improved modifying part of the clustering schema. We describe an approach that use the random sampling technique to improve clustering output quality, we call this algorithm M-FPF[10][11]. Another crucial shortcoming of FPF is that it selects a set of centers not representative of the clusters. This phenomenon must be imputed to the fact that, when FPF creates a new center, it selects the furthest point from the previous selected centers and thus the new center can likely be close to a boundary of the subspace containing the data set. To overcame this we modify M-FPF to use medoids instead of centers.

3 Overview of Clustering and Labeling System

(1) **Querying one or more search engines:** The query entered by the user is redirected to the selected search engines. As a result of the search engine, a list of snippets describing Web pages relevant to the query. An important system design issue is deciding the type and number of snippet sources to be used as auxiliary search engines.

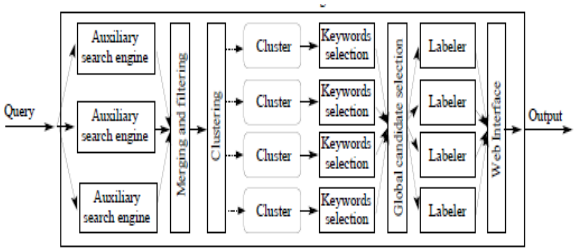


Fig. 1. Architecture of clustering and labeling system

With the rank of the snippet in the list returned by the search engine. Therefore the need of avoiding low-quality snippets suggests the use of many sources each supplying a low number of high-quality snippets.

(2) **Cleaning and filtering:** The input is then filtered by removing non-alphabetic symbols, digits, HTML tags, stop words, and the query terms. These latter are removed since they are likely to be present in every snippet, and thus are going to be useless for the purpose of discriminating different contexts. We then identify the language of each snippet, which allows us to choose the appropriate stop word list and stemming algorithm. Currently we use the ccTLD (Country Code Top Level Domain) of the url to decide on the prevalent language of a snippet.

(3) First-level clustering: We build a flat k-clustering representing the first level of the cluster hierarchy, using the M-FPF algorithm and the Generalized Jaccard Distance [5]. An important issue is deciding the number $k$of clusters to create. Currently, by default this number is fixed to 30, but it is clear that the number of clusters should depend on the query and on the number of snippets found. Therefore, besides providing a default value, we allow the user to increase or decrease the value of $k$to his/her liking. Clusters that contain one snippet only are probably outliers of some sort, and we thus merge them under a single cluster labeled "Other topics".

(4) Snippets re-ranking: In general users are greatly facilitated if the snippets of a cluster are listed in order of their estimated importance for the user. Our strategy is to identify an "inner core" of each cluster and "outliers". In order to achieve this aim we apply the FPF algorithm within each cluster as follows. Since FPF is incremental in the parameter k, we increment $k$up to a value for which it happens that the largest obtained cluster has less than half of the points of the input cluster.

(5) Candidate words selection: For each cluster we need to determine a set of candidate words for appearing in its label called as candidates. For each word that occurs in the cluster we sum the weights of all its occurrences in the cluster and pre-select the 10 words with the highest score in each cluster. We call this as local candidate selection, since it is done independently for each cluster. For each of the 10 selected terms we compute information gain IGm, [6]. The three terms in each cluster with the highest score are chosen as candidates. We call this as global candidate selection, because the computation of IGm for a term in a cluster is dependent also on the contents of the other clusters. Global selection has the purpose of obtaining different labels for different clusters. At the end of this procedure, if two clusters have the same signature we merge them.

(6) Second-level clustering: For second-level clustering we adopt a different approach, since metric-based clustering applied at the second level tends to detect a single large "inner core" cluster and several small "outlier" clusters. The second-level part of the hierarchy is generated based on the candidate words found for each cluster during the first-level candidate words selection. Calling $K$the set of three candidate words of a generic cluster, we consider all its subsets as possible signatures for second level clusters.

4 Experimental Evaluation

In this paper, precision is used to take into account both relevance and membership degrees. Since all three algorithms tested using overlapping clusters, modifying the weight of a result according to its membership degree is used to prevent variation of precision due to the same result being present in multiple clusters. M-FPF employs clusters which record membership degrees in the range [0, 1], while STC and Lingo appear to employ overlapping clusters, which does not define membership degree. In this case, membership degree is set to the inverse of the number of clusters of which a result is a member. M-FPF records relevance in the range [0, 1], however STC and

Lingo do not use relevance or sort results in any fashion, so they only use precision weighted by membership degrees in the tests that follow.

The other key problem is the data to be used for testing. For testing, sets of search results are downloaded and saved so they are identical between runs. Each dataset consists of 100 results, each with a title, snippet and URL. As the following results show, the algorithms' performance depends heavily on the dataset and its distribution of search results. This paper includes the results of M-FPF, STC and Lingo on four queries used in other papers [3], [8] (Jaguar, Apple, Java, Salsa), using the Google! search API.

4.1 Clustering Quality

Search Engine Results In all the tests that follow, the parameters of the three algorithms are left unchanged between runs. A fixed number of six clusters was chosen for all datasets, as there are at least ten topic labels and generation of too many clusters for a relatively small number of results can result in excessive fragmentation of categories. STC and Lingo were left to the defaults set by the Carrot2 software, M-FPF has the following default parameters set, unless otherwise noted: Nc = 10, the parameter is chosen to give on average a balance between precision and recall. The graphs show three bars for each of the algorithms: on the left is weighted precision, the middle is recall and on the right is a relevance score. A difference between precision and recall indicates a tendency of an algorithm to return only a small number of results that have a high probability of being correctly classified (high precision, low recall) or a large number of results in each cluster, with high overlap between clusters (low precision, high recall).

Figure 2 shows the performance of the three algorithms on two datasets: Jaguar and Apple. These two datasets are a fairly average case with three or four large clusters and two or three smaller clusters with low to moderate overlap between the clusters. M-FPF performs well in these cases, delivering a balance between precision and recall. All three algorithms show higher precision in the Jaguar dataset and higher recall in the Apple dataset, possibly indicating higher overlap in the later.

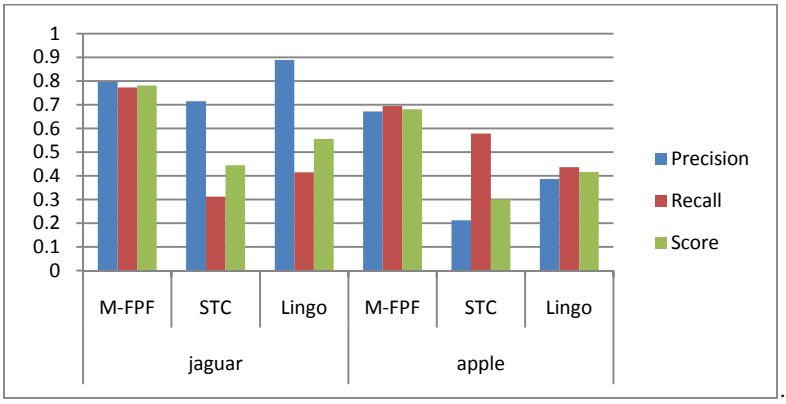


Fig. 2. Cluster quality measured using precision, recall and R1 score for the Jaguar and Apple datasets

M-FPF and Lingo was able to extract almost all of the major topics, while STC was unable to extract more than half. Lingo suffered from a low number of classified results (many results were binned in 'Other Topics' i.e. as outliers), which is expected from its design focus on cluster purity [4].

Figure 3 shows the performance of the three algorithms on two more challenging datasets: Java and Salsa. Both datasets are dominated by two large clusters and four or five smaller clusters, making it hard for the algorithms to effectively extract the topics. As a result, recall and overall score of all three algorithms drop significantly. The Salsa dataset also has very high overlap, and due to the default parameters for M-FPF, it performs similarly to Lingo while STC performs slightly better overall.

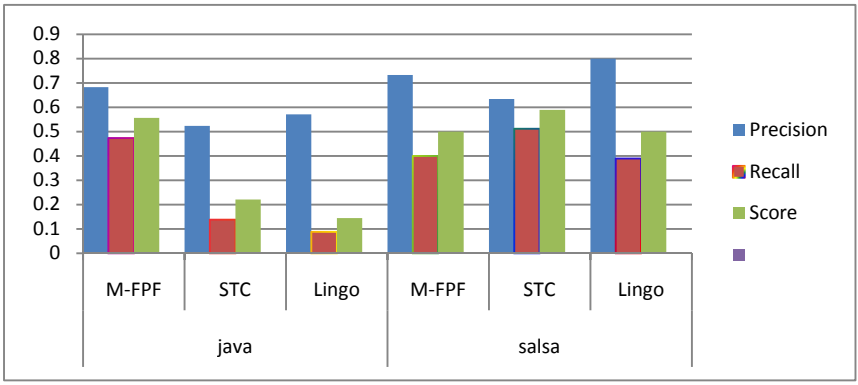


Fig. 3. Cluster quality measured using precision, recall and R1 score for the Java and Salsa datasets

5 Conclusion

This paper demonstrated M-FPF, a web search result clustering and the labeling tasks are performed on the fly by processing only the snippets provided by the auxiliary search engines, and use no external sources of knowledge. Clustering is performed by means of a modified version of the furthest-point-first algorithm.. Finally, M-FPF performs well compared to the established STC and Lingo algorithms, demonstrating both good quality clustering without using a more complex label driven approach to document clustering. Enhancing the performance of search engines and improving the usability of search results is an active area of research, and clustering web search results is only one way of doing this. however, improvements can be made its efficiency as well as the use of hierarchies to improve document organization.

References

1. Zamir, O., Etzioni, O.: Web document clustering: A feasibility demonstration. In: Proceedings of the 21st Annual International SIGIR Conference on Research and Development in Information Retrieval (1998)

2. Hanumanthappa, M., Prakash, B.R., Mamatha, M.: Improving the efficiency of document clustering and labeling using Modified FPF algorithm. In: Proceeding of International Conference on Problem Solving and Soft Computing (2011)
3. Geraci, F., Leoncini, M., Montangero, M., Pellegrini, M., Renda, M.E.: *FPF-SB*: A Scalable Algorithm for Microarray Gene Expression Data Clustering. In: Duffy, V.G. (ed.) HCII 2007 and DHM 2007. LNCS, vol. 4561, pp. 606–615. Springer, Heidelberg (2007)
4. Osinski, S., Weiss, D.: A concept-driven algorithm for clustering search results. IEEE Intelligent Systems 20(3), 48–54 (2005)
5. Charikar, M.S.: Similarity estimation techniques from rounding algorithms. In: Proceedings of the 34th Annual ACM Symposium on the Theory of Computing, STOC 2002, Montreal, CA, pp. 380–388 (2002)
6. Yang, Y., Pedersen, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of the14th International Conference on Machine Learning, ICML 1997, Nashville, US, pp. 412–420 (1997)
7. Ferragina, P., Gulli, A.: A personalized search engine based on Web-snippet hierarchical clustering. Special Interest Tracks and Poster Proceedings of the 14th International Conference on the World Wide Web, WWW 2005, Chiba, JP, pp. 801–810 (2005)
8. Crabtree, D., Gao, X., Andreae, P.: Standardized evaluation method for web clustering results. In: Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (2005)
9. Matsumoto, T., Hung, E.: Fuzzy Clustering and Relevance Ranking of Web Search Results with Differentiating Cluster Label Generation
10. Geraci, F., Pellegrini, M., Maggini, M., Sebastiani, F.: Cluster Generation and Cluster Labelling for Web Snippets: A Fast and Accurate Hierarchical Solution. In: Crestani, F., Ferragina, P., Sanderson, M. (eds.) SPIRE 2006. LNCS, vol. 4209, pp. 25–36. Springer, Heidelberg (2006)
11. Geraci, F., Pellegrini, M., Pisati, P., Sebastiani, F.: A scalable algorithm for high-quality clustering of Web snippets. In: Proceedings of the 21st ACM Symposium on Applied Computing, SAC 2006, Dijon, FR, pp. 1058–1062 (2007)
12. Gonzalez, T.F.: Clustering to minimize the maximum intercluster distance. Theoretical Computer Science 38(2/3), 293–306 (1985)

Data Mining in Online Social Games

Nazneen Ansari, Maahi Talreja, and Vaishali Desai

St. Francis Institute of Technology, Mumbai
sfitnaz12@yahoo.com,
{talreja55,vaishaliparikh1988}@gmail.com

Abstract. For thousands of years, people have been playing games of chance or wagering on the outcomes of various games and events. As a medium, the computer game is currently in a period of rapid development. From a design point of view, video games are becoming more complex and they are rapidly spreading to new platforms such as mobile phones, pocket computers, and websites. This paper aims to determine whether Association mining algorithm applied to an online social games database would provide the game designers with meaningful rules that would help improve the design of the game. A data set of online social gamer profile was created. The database contains various aspects of social games. The rules generated from association analysis would be of tremendous benefit to the gaming industry, as they can then use them to optimize game design features.

1 Introduction

Data Mining is about finding patterns and relationships within data that can possibly result in new knowledge. Data mining software applications includes various methodologies that have been developed by both commercial and research centers. These techniques have been used for industrial, commercial and scientific purposes[1]. Game designs is the process of designing the content, environment, storyline and characters and rules of a game. Game design using Association mining will aid game designers to know exactly what players want and expect from computer games. This knowledge in turn will enable them to take informed and strategic decisions while designing various aspects of the game. There is widespread consensus that games motivate players to spend time on task-mastering the skills a game imparts. Nevertheless, the literature reveals that a number of distinct design elements, such as narrative context, rules, goals, rewards, multi-sensory cues and interactivity, seem necessary to stimulate players' interest in the game [2]. The standard approaches of online Game design, consist of , User polls and surveys, Gaming forums and Market research. This paper suggests a new approach using Association mining to study online gamer activities, and identify frequent item sets, such as popular site, popular roles, popular maps etc. Such item sets can give us strong association rules. These rules can be translated by the game designer into knowledge, and use the same knowledge to enrich existing games, or create a new one that has a high chance of becoming a popular game.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 801–805.
springerlink.com

2 Association Mining

Association rule mining, one of the most important and well researched techniques of data mining, was first introduced by Agrawal, et al [3]. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. Association rules are widely used in various areas such as telecommunication networks, market and risk management, inventory control etc. Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub-problems. One is to find those itemsets whose occurrences exceed a predefined threshold in the database; those itemsets are called frequent or large itemsets. The second problem is to generate association rules from those large itemsets with the constraints of minimal confidence. In many cases, the algorithms generate an extremely large number of association rules, often in thousands or even millions. Further, the association rules are sometimes very large. It is nearly impossible for the end users to comprehend or validate such large number of complex association rules, thereby limiting the usefulness of the data mining results. Several strategies have been proposed to reduce the number of association rules, such as generating only "interesting" rules, generating only "nonredundant" rules, or generating only those rules satisfying certain other criteria such as coverage, leverage, lift or strength [4].This is a sample file. Please use this file to correctly typeset a submission to a conference published by Springer. The associated pdf file will help you to have an idea of what your paper should look like.

3 APRIORI

Apriori uses a complete, bottom-up search with a horizontal layout and enume ates all frequent item sets [5]. An iterative algorithm, Apriori counts item sets of a specific length in a given database pass. The main property of Apriori algorithm is that all nonempty subsets of a frequent item set must also be frequent.

3.1 Apriori Rules

Minimum support: 0.1 (10 instances)
Minimum metric <lift>: 1.5
Number of cycles performed: 18
Generated sets of large itemsets:
Size of set of large itemsets L(1): 26
Size of set of large itemsets L(2): 38
Size of set of large itemsets L(3): 12
Size of set of large itemsets L(4): 1

Best rules found:

1. Gender=m Devices=Desktop/Laptop 25 ==> why-play= fun and excitement Site=Facebook 10 conf:(0.4) < lift:(2.35)> lev:(0.06) [5] conv:(1.3)
2. why-play=fun and excitement Site=Facebook 17 ==> Gender=m Devices=Desktop/Laptop 10 conf:(0.59) < lift:(2.35)> lev:(0.06) [5] conv:(1.59)

3. why-play=fun and excitement 34 ==>Gender=m Devices=Desktop/Laptop Site=Facebook 10 conf:(0.29) < lift:(1.96)> lev:(0.05) [4] conv:(1.16)
4. Gender=m Devices=Desktop/Laptop Site=Facebook 15 ==> why-play=fun and excitement 10 conf:(0.67) < lift:(1.96)> lev:(0.05) [4] conv:(1.65)
5. Devices=Desktop/Laptop 40 ==> why-play=fun and excitement Site=Facebook 13 conf:(0.33) < lift:(1.91)> lev:(0.06) [6] conv:(1.19)
6. why-play=fun and excitement Site=Facebook 17 ==> Devices=Desktop/Laptop 13 conf:(0.76) < lift:(1.91)> lev:(0.06) [6] conv:(2.04)
7. Gender=m Devices=Desktop/Laptop why-play= fun and excitement 12 ==> Site=Facebook 10 conf:(0.83) < lift:(1.81)> lev:(0.04) [4] conv:(2.16)
8. Site=Facebook 46 ==> Gender=m Devices=Desktop/Laptop why-play=fun and excitement 10 conf:(0.22) < lift:(1.81)> lev:(0.04) [4] conv:(1.09)
9. Devices=Desktop/Laptop 40 ==> Gender=m why-play=fun and excitement Site=Facebook 10 conf:(0.25) < lift:(1.79)> lev:(0.04) [4] conv:(1.11)
10. Gender=m why-play=fun and excitement Site=Facebook 14 ==> Devices=Desktop/Laptop 10 conf:(0.71) < lift:(1.79)> lev:(0.04) [4] conv:(1.68)
From the rules, the designer gets a fair picture of the choices of the online social gamer. These association rules guides the designer in studing gamer activities and identifying frequent itemsets like popular device, site, frequency and length-oftime-played-social-game. For example, some of the best rule says:
Gender=m Devices=Desktop/Laptop 25 ==> why-play= fun and excitement Site=Facebook 10 conf:(0.4) < lift:(2.35)> lev:(0.06) [5] conv:(1.3)
Frequency=several times a day 36 ==> Leng-oftime- psg=1-2 years 12 conf:(0.33) < lift:(1.59)> lev:(0.04) [4] conv:(1.14)
Site=Facebook 46 ==> Devices=Desktop/Laptop why-play=fun and excitement 13 conf:(0.28) < lift:(1.57)> lev:(0.05) [4] conv:(1.11)

These best rules tells the game designer that online social gamers are male members, they play games for fun and excitement, on desktop/laptop platform frequency is several times a dayand site is facebook.

4 Association Mining for Designing Online Social Games

Weka(Waikato Environment for Knowledge Analysis), is a tool comprising of numerous machine learning algorithms that can be applied to Data Mining Problems. Weka was developed at the University of Waikato, New Zealand. It is an open source software issued under the GNU General Public License. Input to Weka is given as a data set. Weka permits the input data set to be in numerous file formats like CSV (comma separted values: *.csv), Binary Serialized Instances (*.bsi) etc. However, the most preferred and the most convenient input file format is the attribute relation file format (arff). So the first step in Weka always is taking an input file and making sure that it is in ARFF. Weka automatically convert .cv file inot .arff format. A data set of online gamers was created with the attributes like age, gender, devices, reason for playing social games, length-of time-playing-socialgames, length-of-game-play-session, frequency, site and currently-games played-once-a-week. A dataset (online-game.arff) was created in WEKA tool which consisted of 10 attributes and 100 records. We loaded the data set into WEKA, performed a series of operations using

WEKA's attribute and discretisation filters, and then performed association rule mining on the resulting data set. WEKA allows the resulting rules to be sorted according to different metrics such as confidence, leverage, and lift. In this example, we have selected lift as the crit ria. Furthermore, we have entered 1.5 as the minimum value for lift (or improvement). Lift is computed as the confidence of the rule divided by the support of the right-hand-side (RHS). In a simplified form, given a rule L => R, lift is the ratio of the probability that L and R occur together to the multiple of the two individual probabilities for L and R, i.e.,

lift = Pr(L,R) / Pr(L).Pr(R).

If this value is 1, then L and R are independent. The higher this value, the more likely that the existence of L and R together in a transaction is not just a random occurrence, but due to some relationship between them. Here we also change the default value of rules (10) to be 100; this indicates that the program will report no more than the top 100 rules (in this case sorted according to their lift values). The upper bound for minimum support is set to 1.0 (100%) and the lower bound to 0.1 (10%). Apriori in WEKA starts with the upper bound support and incrementally decreases support (by delta increments which by default is set to 0.05 or 5%). The algorithm halts when either the specified number of rules are generated, or the lower bound for minimum support is reached.

4.1 Results of Association Analysis of "games.arff" by WEKA Tool

Data Mining in Online Social Games
=== Run information ===
Scheme: weka.associations.Apriori -N 100 -T 1 -
C 1.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: social-game-stat2-
weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.Discretize-B5-M-
1.0-R7-weka.filters.unsupervised.attribute.Remove-
R1,6
Instances: 100
Attributes: 7
Gender
Devices
why-play
Leng-of-time-psg
Frequency
Site
currently-game-played-once-a-week
=== Associator model (full training set) ===

5 Conclusion

This paper shows how Association mining could be used to extract knowledge from online social gamer dataset to create strong rules that can guide the game design process. Using the strong rules in the design phase will enable the designers to preserve popular ingredients in new game titles and make successful games and tap into a wide pool of gamers, thereby generating commensurate revenue.

6 Future Scope

The above analysis has been done on video games and events, so accordingly the future scope for data mining in online game can be done on motion gaming along with latest game consoless.

References

1. Vamanan, R., Ramar, K.: Classification of Agricultural Land Soils: A Data Mining Approach. International Journal on Computer Science and Engineering 3(1), 379–384 (2011) ISSN : 0975-3397
2. Dodlinger, M.: Educational Video Game Design: A Survey of the Literature. Journal of Applied Educational Technology 4(1), 21–31 (2007)
3. Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216 (1993)
4. Kotsiantis, S., Kanellopoulos, D.: Association Rules Mining: A Recent Over view. GESTS International Transactions on Computer Science and Engineering 32(1), 71–82 (2006)
5. Anu Radha, T., Lavanya, P.: Recent Trends in Parallel and Distributed Apriori Algorithm. International Journal of Engineering Research and Applications 1(4), 1820–1822

Data Mapping in Intelligent Form Using Random Hierarchical Bit Format Enhancing the Security in Data Retrieval

Rahul Gupta[1], Nidhi Garg[2], and Preetham Kumar[1]

[1] Department of Information and Communication Technology
Manipal Institute of Technology,
Manipal University, Karnataka, India
rahul.gupta.engg@gmail.com,
preetham.kumar@manipal.edu
[2] Department of Computer Science
Jaipur National University, Jaipur, India
nidhigargjaipur@gmail.com

Abstract. The paper highlights a unique and secure way of retrieving data, where user can never identify the path followed by the system to reach the goal. The whole collection of data is represented in Boolean form which is stored in a matrix. This matrix can be used for quick retrieval of data as all the functions and properties of mathematical logic can be applied. This also helps a great deal in saving memory as the data is stored in Boolean form. Boolean answers to the auto generated questions asked by computer produces a path. The main advantage of this concept is, the data is highly secured as only the person who has the key matrix can understand the meaning of matrix. The data is normalized so the efficiency in the data retrieving is also increased. The computer puts an auto generated random set of questions each time when the data has to be retrieved, which results in formation of a hierarchical tree. Thus any malicious attempt to use the previously used key won't open the same lock and the attempt maker can be traced.

Keywords: Pattern Recognition, Bit-manipulation, Data Mining, Pattern Recognition, Matrix Generation.

1 Introduction

Retrieval of data by writing a query and then waiting for a table to appear on the window is not only time consuming but also requires skilled technicians to manage it. This is an algorithm where we don't have to write any query or the data names but we get the required solution. The main idea behind this concept is to supply the various data inputs to recognize the pattern and then use different functions to get a set of bounded solutions. A huge collection of database is mapped in the form of a characteristics matrix. Its elements are considered as objects. These objects later help the computer to trace a path to reach the final output. The user will be asked certain questions and it is answered in Boolean form either yes or no. These are logically

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 807–812.
springerlink.com

implemented as false (0) or true (1) by the computer. The non favorable choices are logically marked with a 0 and are removed by the computer. In this manner the fittest of the options survive.

Auto generated questions are being asked by the computer and a pattern is being recognized [1]. With its help, a matrix is generated containing only the favorable outputs. Results are used to lead the user to next question which in turns forms a path. There can be various paths being running at the same time. Hence a set of outputs is gained. It is based on eliminating the unfavorable options, thus the final set of answers can be considered to be robust. In this method, the user is completely unaware of the next question. This helps in maintaining data as abstract. Only the master matrix knows about the data path and the meaning of the Boolean tree formed. Any other person cannot retrieve the path without the master key.

2 Mapping of Data

Our brain stores information in a relational manner. When we come across any object be another person, new movie, place, clothes etc. brain explores their characteristic features and remember it .When we discuss, we use the information or data we learnt, to define that object. Our mind associates those things/incidences on the basis of those features and this form our perceptions as well. By this article we focus on how the concept of elimination of choices and survival of the fittest can be applied to a set of solution space which we have got from the neural networks result [2]. The result will be accurate and closer to the answer than the conventional neural network and requires no technical acquaintance as the user has to just answer the questions and a road map is developed on its own to get more and more information moving across it.

3 Question Theory and Path Generation

The question theory defines how the computer system manipulates in computational form what a normal human being observes from his surroundings. Every entity has an associated name, its place, gender, characteristics,unique identity or feature,relative things etc. The object can be a place, person, occupation, definition, mathematical value or even an alphabet . Mathematical functions(integration) are used for generating their permutations which can help in generating a matrix easily. In mathematical terms, every entity is unique but each has certain properties or qualities commonly associated with it. This can be used to link it with other entities like same size, color, weight, value etc. for example a number can be described as even, odd, perfect square, multiple of certain numbers, factors, factorial, prime, etc. These common properties are used to implement our concept of elimination. By using the set of questions, it keeps on eliminating the unfavourable options directing us closer to the answer at every step.[1]

The flow is always linear and there is no need of looping back

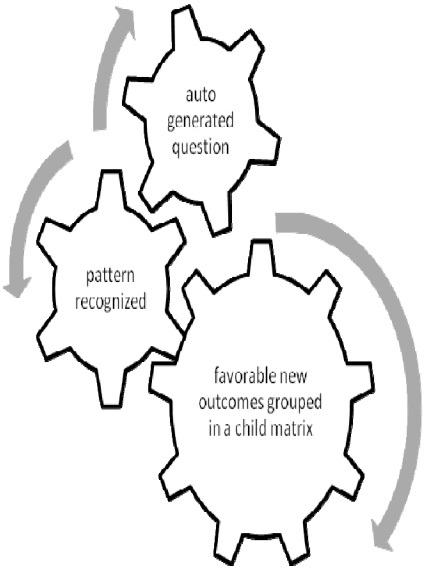


Fig. 1.

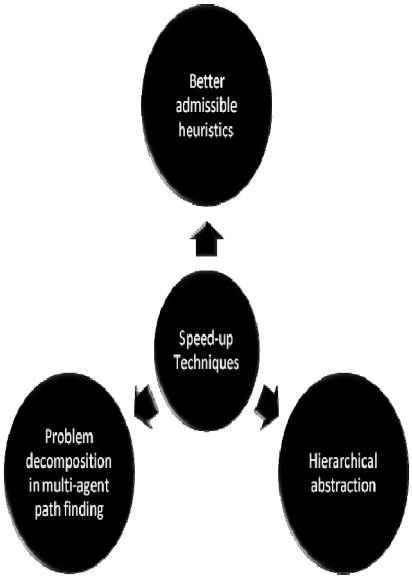


Fig. 2.

There is a continuous tracing of path. As is clear from the diagram, the auto-generated question set is under manipulation on the basis of recognition of a pattern in the answers given by the user. The answers help in building a path with the help of basic information and with every question, more significant details. Data has been saved in an object-oriented manner, using all its features and applications. We can say that here we use reverse retrieval i.e. data object is being looked after on the basis of its properties. Later, sub matrix grids are established. These grids store the pointers to the master matrix. These pointers are used to point to a data element in the master matrix. They are also termed as "keys". The data elements for which user answers positively, are automatically either marked or put into a new matrix named child matrix. Once a path is traversed to reach the goal, will not necessarily be the same when the user puts in the same query [3]. This technique is not for general purposes. Hence its time and memory constraints need not be a matter of concern.

4 Representation in Matrix Form

The whole concept is divided into two set of matrices. The first matrix called the master matrix holds the questions and is never changed for a particular problem. The second matrix is called the key matrix which contains all the solution with respect to a particular question. Each row in the matrix is associated with certain properties and all the questions that are there in the row are directly related to that. Each element in the key matrix is a answer to the master matrix which contains all the questions. The key matrix comprises of the answers given by user. Since the user is bound to answer only in Boolean format, the matrix contains either 0 or 1. Any attempt made to search

the address of a data object will go in vain. This is because though the address would be same, address owner will be different.

The example given below portrays how this concept works in real life. There is a master matrix which holds all the information but is of no use as it is just a formal structure; the key matrix holds the answer of the master matrix. The most interesting feature is, if anybody knows the key matrix but does not hold the master matrix nothing can be done. Combination of both the matrix is required to get to the solution. And during data exchange the data flows through key matrix and thus it is impossible to trace without making use of master matrix. This makes this algorithm unique and highly secured. The master matrix holds the questions whose answers are to be given in just yes (1) or no (0). Based on the answers provided in 1st master matrix the subsequent master matrices are generated [2].

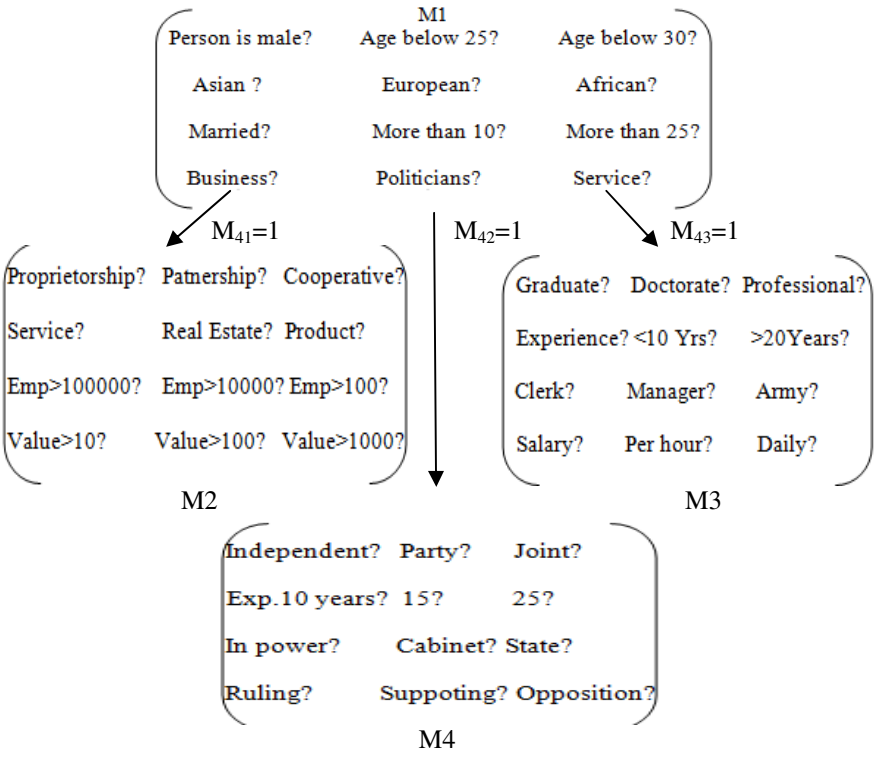


Fig. 3. Matrices to depict flow of control from one to another

If $M1_{41}$= 1 which means that the person has business as its profession. He/she is redirected to another matrix which holds the questions related to business explicitly. In this way each matrix has some common property with each row defining the property of a particular attribute. This each element in matrix behaves independently and explaining more about a particular value. Now the matrix M2 is generated when user has chosen business as its profession. So it contains the basic information

regarding business. Now this matrix is also linked to other matrices based on the user choices. Like if the persons owns a company which deals in real estate then $M_{22}=1$ after filling this matrix user is directed to matrix containing questions related to real estate. So in this way a hierarchy is followed through which we come across more and more information at each level without compromising on security.

If earlier $M1_{42}=1$ then it means a person's profession is politician. The master matrix that will be generated will be having the questions related to politics. The user only get to see the following matrix if he has selected politicians as its profession in the matrix M1. If the user has chosen $M1_{43}=1$, it indicates that the user has chosen service as his profession so now all the questions related to services will be asked. A new matrix comes into picture which holds the questions related to services. In this way many more profession can be added and can be directed to the secondary matrix as done before. There is no constraint that the number of elements in each row should be same thus increasing the flexibility. The above matrix M4 is used to gain the information of person's work. So in this way the entire information of a person can be generated by using the concept of 0's and 1's. Thus the hierarchical matrix holds the complete information about a person. So the matrix is also linked by the pointers to direct the set of master matrix in this way security is maintained.

4.1 Structure

Any entity is considered as an object. Various tuples are created initially. They contain information about that particular object. The questions being asked are based on these properties. New matrices with same fields and different tuples are automatically generated each time a question is asked. Each new matrix is anticipated again and a new and simpler database comes into play. One object is related with different entities in different tables or matrices. When an object is given acceptance by the user, the related tuples in all associated matrices are used to generate a new matrix.

4.2 Data Retrieval

A computer provides the user with certain questions in a sequence. That question is answered by the user in either yes (i.e. 1) or no (i.e. 0). When a 0 is received, the matrix which had that option as the primary object is rejected.

We can also say that when a 1 is encountered, a new matrix is formed with all the favorable outputs. One option points towards various different outputs, so more than one matrix can be formed. Using the same question, many matrices are formed. The challenge of attaining the same set of solutions while provided with different set of questions is over powered by the multiple matrices continuously being formed at logical level. These matrices are continuously modified or transformed into a next level. As the level increases, size of the database under consideration decreases and a path is created .This path is not unique and the probability of retrieving it again is very low.

5 Applications

This theory has applications in:-

- Medicine – If there is a new and typical case for the doctor he can tell the state of patient in the Boolean form and can get the most close and likely event of the past that has occurred and the cure that was given.
- Security – Whenever there is a terrorist activity description of the event can be given to recognize the pattern.
- Traffic control - Recognizing and saving records for any accidents.
- Sports- The performance state of the teams is evaluated as each player's information, ground details, opposition team's complete facts and figures is given as input and using this self learning algorithm explained before the most likely result can be concluded.
- Business- The current scenario of the market is given as input and the prediction in the market can be made easily especially to tell that which stock will rise and which will come down.
- Aptitude tests- By asking a set of questions it can tell about a person's interests, behavior and mental level where on the basis of first questions answered the proceeding questions will appear. This will be highly useful to take real life decisions too.
- Law- The time and cost can be reduced by analyzing the law case with the help of our question theory.

References

[1] Gupta, R., Garg, N.: Optimum unitization of self learning algorithm in artificial neural network. In: Proceedings 4th International Conference on Computer and Automation Engineering, ASME 2012, pp. 255–260 (2012) ISBN 978-0-7918-5994-0

[2] Mirzaaghazadeh, A., Motameni, H.: Using Neural Network in Pattern Recognition. In: Proceeding of Iran Computer Conference (2002)

[3] Teshnehlab, M., Watanabe, K. (eds.): Intelligent control based on flexible neural networks. Kluwer Academic Publishers, Dordrecht (1999) ISBN 0-7923 -5683-7; Automatica 38(3), 564–569 (March 2002)

[4] Yasunobu, S., Miyamoto, S.: Automatic train operation system predictive fuzzy control. In: Sugeno, M. (ed.) Indusrrial Applicarions of Fuzzy Control. North-Holland, Amsterdam (1985)

[5] Foster, I., Kesselman, C., Nick, J., Tuecke, S.: The Physiology of the Grid: an Open Grid Services Architecture for Distributed Systems Integration. Technical report, Global Grid Forum (2002)

Effective Unit Testing Framework for Automation of Windows Applications

A.N. Seshu Kumar and S. Vasavi

V.R. Siddhartha Engineering College, Vijayawada
{seshu1203,vasavi.movva}@gmail.com

Abstract. The major concern of software industry is software quality and reliability. Unit Testing is a practical approach to improve the quality and reliability of a software. Unit testing is usually performed by programmers and is the base for all other tests such as integration testing and system testing. Unit Testing can be done manually (and/or) automatically. This paper presents "White.NUnit" framework that automates the unit testing of windows applications. The automated unit tests are written by the developers after the completion of functionality coding. We found that the number of defects got reduced when automated unit tests are written iteratively similar to test driven development. This framework proved that significant portions of windows application can be automatically tested without manual intervention. This reduces the Manpower involved in testing each and every unit of the application and increases the quality of the software product.

1 Introduction

The process of a software development includes white box testing which is a technique used by software developers to verify whether their code works as expected. Unit Testing is a kind of white box testing where individual units of software are tested. The intension of unit testing is to check whether each and every unit (module) of a software works as per the developer's expectation. Unit Testing can be done manually (and/or) automatically. According to the research conducted [6] 65% of bugs can be caught by unit testing alone. Unit Testing finds the defects in each and every unit of the application at initial level of testing. It increases the confidence levels of developer by ensuring that their code is working correctly. According to Industry research the cost of fixing a defect in Quality Assurance (QA) is 100 times more than fixing it during development. So, Unit Testing ensures that code works correctly and improves the quality, reliability and cost of the application software development. We initially performed unit testing manually where it was a time consuming task and hence we intend to choose automation of unit testing. To automate unit testing, it is necessary to write test scripts called unit tests. Once the unit tests are available, it is easy to automate the unit testing. The White.NUnit is an open source automation framework for automating the windows applications. This paper focuses on methodology and framework for automation of unit testing.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 813–822.
springerlink.com

2 Background and Related Work

The automated unit tests for any application can be written in two ways:

- Before code implementation
- After code implementation

If unit tests are written before any code implementation then it is known as Test Driven Development. If Unit tests are written after the code implementation then it is known as Test After Development.

2.1 Test Driven Development

In Test Driven Development, the developer writes automated unit tests for the new functionality they are about to implement. It is a software engineering process that follows small development cycle. In industry, while coding a software application the development cycle that they follow is shown below in figure 1.

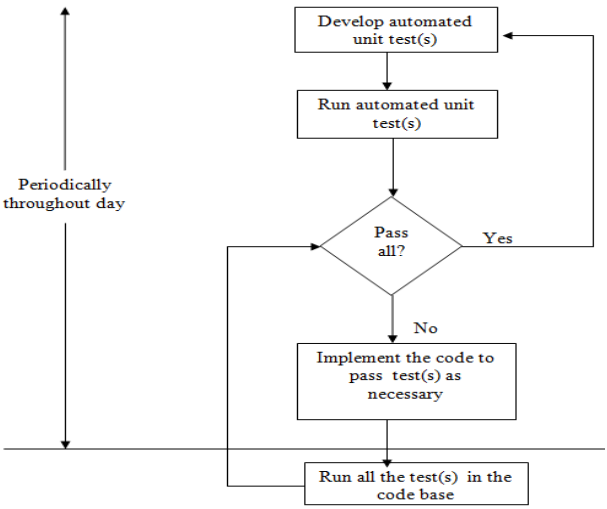


Fig. 1. Test Driven Development Cycle

Let us see the detailed description of development cycle of test-driven development.

2.1.1 Develop Automated Unit Tests

Without writing any code to implement feature, test cases must be written by the developer in the initial stage by collecting the specifications and requirements in the form of user stories (or) use cases that covers all the requirements and conditions. This makes developer to focus on requirements and coding in this manner makes the code consistent.

2.1.2 Run Automated Unit Tests

Run the automated Unit tests to ensure that they fail because there is no implemented code yet.

2.1.3 Implement the Code to Pass Test(s)

Developer needs to write the code for those cases that are failed in the previous test. The code that developer writes should not add any other unpredicted functionality.

2.1.4 Run All the Tests in Code Base

Once development is done, run all the automated tests in the code base. If all the tests are pass then develop automated unit tests for other features of the application and repeat the same process otherwise implement the code necessary to make the test pass and run the automated unit test(s).

Finally, once the development of code for the application and unit testing are done, the developer may restructure the code for better readability (or) improving the performance. The advantage of above written unit tests is that whatever changes the developer may make to the code now, it won't affect the existing functionality of the build [11], as the test cases written earlier defines the requirement and specifications of the application.

2.2 Test After Development

In Test After Development, the developer writes automated unit tests after the completion of code for the application. It is also a software engineering process that follows small development cycle. The development cycle for test after development is shown in figure2.

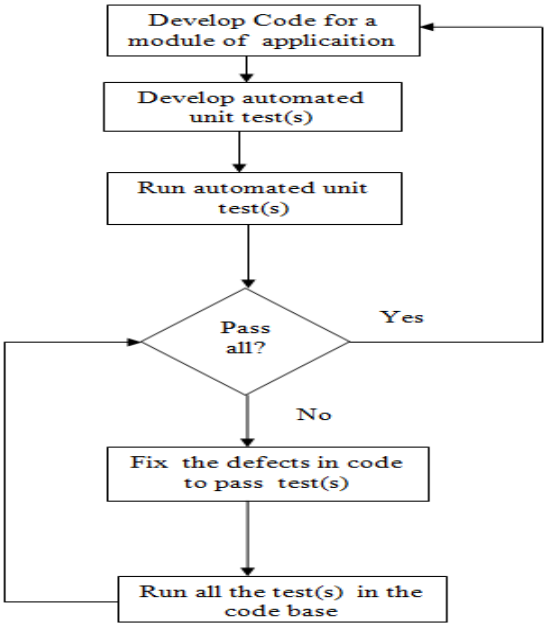


Fig. 2. Test After Development Cycle

Let us see a brief explanation of the test after development cycle:

- Initially develop the code for few features of an application.
- Develop the automated unit tests that cover all the features that are developed.
- Run automated unit tests and if all the tests are passed then continue the development of the remaining features of application.
- Otherwise, fix the defects that are found during unit testing, repeat the testing of automated unit tests.
- Repeat the above process till the features of the entire application are done.

These are the two approaches for unit testing the application. We have used test after development approach in our development process.

To automate the unit testing we have an open source framework Nunit which is a free unit testing framework which is not seamlessly integrated with an IDE. It loads test assembly in different application domain and keeps a watch on targeted assembly for any changed events. But the problem with this is it just validates the internal functionality of any method written in an application and it doesn't have any interaction with the user interface. So, we have another open source automation framework with the name "white.NUnit" for automation of user interface applications that needs the support of Nunit.

3 Methodology

Initially we explain the methodology for development of application and then for the automation of unit testing. The methodology that we followed for development of our application is agile methodology. The agile methodology is based on iterative and incremental development throughout the life-cycle of a project .It reduces the risk by developing application in short span of time. Development accomplished in one unit of time (generally up to two weeks) what we followed is called an iteration. All the required functionality may not be covered in one iteration for releasing the project. But it will be covered in multiple iterations. The idea is to have a defect free application available at the end of each iteration.

Once development is completed in the iteration we perform automated unit testing. The test after development approach is followed in our development process. The methodology for automation of unit testing is shown in figure 3.

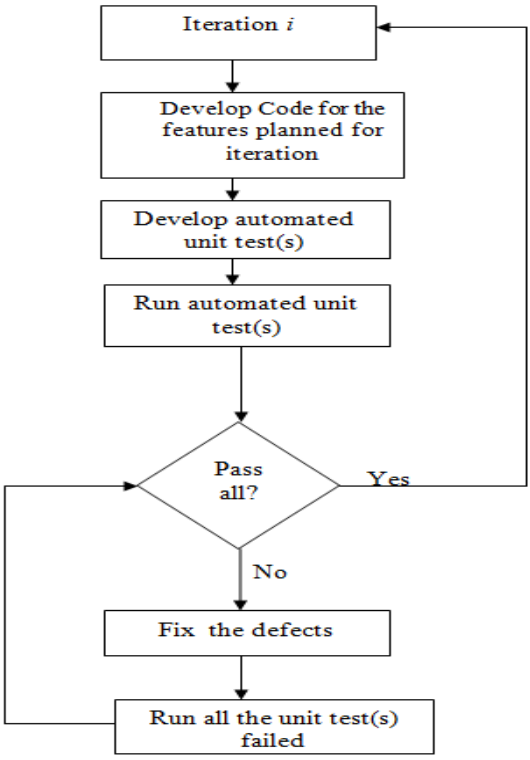


Fig. 3. Automation Methodology

In the above figure iteration *i* indicate the number of iteration in development of application.

4 Requirements for Framework

The minimum requirements of unit testing framework for automation of widows application are:

Operating System: Windows XP SP 2+ or Windows 7
Platform: Microsoft Visual Studio 2005, 2008 or 2010.
Software: NUnit which is the Open Source software (OSS) for which we'll create unit tests.
PC with minimum configuration.

The Nunit software can be downloaded easily from the internet for free except Microsoft Visual Studio for which you need to get a License.

5 Unit Testing Automation Framework

White.NUnit is a open source unit testing automation framework for windows applications in .Net. Since the template for the unit tests (test cases) are not generated automatically in white, it means that developer who is not aware of unit tests, he has to get some training on using this framework. The unit tests are written in a separate project. The framework supports access to all the controls in an application while writing unit tests. Once we are done with the unit tests we can automate the unit testing of application using this framework. This framework doesn't have a separate setup file to run unit tests it provides dll's that are to be added in the project of unit tests and run the unit tests by loading the project dll into user interface provided by NUnit. The set of steps involved in using this framework are:

- .Installation of Nunit Software.
- Setting Up an Environment.
- Write Unit Tests.
- Run Unit tests.

5.1 Installation of Nunit Software

The binaries and documentation of NUnit can be downloaded from their website. It comes as an installer package or can also be downloaded in a standalone folder which is archived inside a file. Once downloading is completed, install the setup file which is quite simple process.

5.2 Setting Up an Environment

To start with White.NUnit, we have to create a test project to contain the White.NUnit tests manually. Once we create a project, it requires certain amount of nontrivial work to finally start writing a test case. Before we start writing a test case, we should include NUnit's and White's linked libraries to the project. To include them, we need to add references to the Core library of NUnit i.e. Nunit.Core, the whole NUnit framework which is in NUnit.framework dll and the core library of White i.e., White.Core and all the dll's of white framework.

5.3 Write Unit Tests

Since the environment is all set for the framework, we need to write unit tests. To write unit tests, first any developer has to search for the document of the source software and then start writing unit tests. Since we are writing the unit tests for automation of windows application we have to access the each and every control in the application and validate them. So, the developer of unit tests must be familiar with the code of the application to access the controls in unit tests.

In the case of White.NUnit, once a developer writes class, he has to create a Test project, or if the Test project is already there, he has to manually write separate class to write the Unit test as there are no integrated features built in inside an IDE.A sample unit test is shown in following figure 4.

```
[Test]
public void Case001English()
{
    //Retrieving Login window from application
    Assert.IsNotNull(SampleApp);
    List<Window> windows = this.SampleApp.GetWindows();
    Window LoginWin = this.SampleApp.GetWindow("Form1");
    Assert.IsNotNull(LoginWin);
    ListBox languages = LoginWin.Get<ListBox>("listBox1");
    Assert.IsNotNull(languages);
    languages.Items[4].Click();
}
```

Fig. 4. A Sample Unit Test

5.4 Run Unit Tests

Now, to execute the tests written for White.NUnit, once again, developer should be familiar with the whole process. First he needs to compile the test code to generate a binary file and then load that binary into White.NUnit and then run the tests for the desired functions. The following is the GUI of white.NUnit framework to execute unit tests.

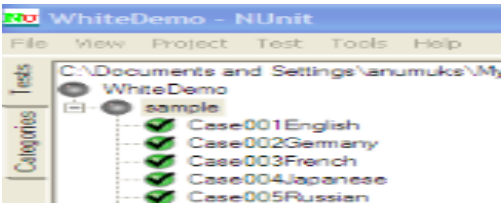


Fig. 5. White.NUnit GUI to execute Unit Tests

This framework allows the user to view the user interface that it automates.The following is the figure that shows the user interface being loaded during execution and perform unit testing.

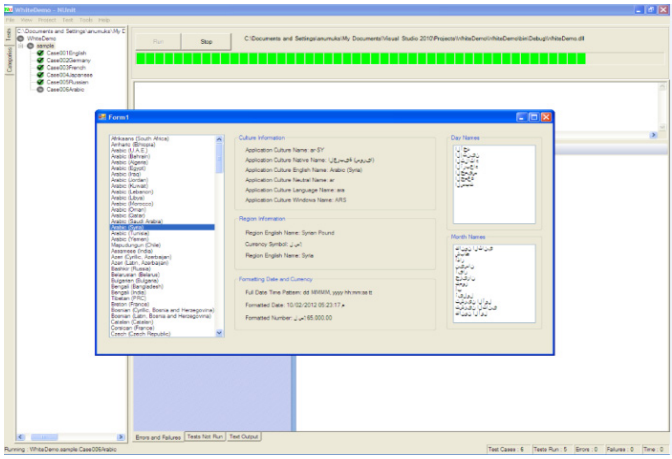


Fig. 6. Sample Screen of White.NUnit Framework

6 Result Analysis

We made analysis and generated a report considering some usability and performance tests of White.Nunit.

Table 1. White.NUnit Metrics Report

Metric	White.NUnit
Installation	Easy
Setting up environment	Easy
Avg Time taken to write First Unit Test	6 mins
Avg Time to rin First Unit Test	<1 min
Document/Internet Forums	Very good

We used the framework for a real time application that has number of features. Initially manual unit testing was done which was time consuming and then we used this framework, developed unit tests and started automating the unit testing. We found lot of time difference compared to manual testing.

Application has 10 features and for each we developed unit tests separately approximately 200.We execute all the cases in 4 hrs of time. When we did manually it takes 8 hrs of time per person in the team. We are totally 10 members in our team.

Total time taken by the team is 80 hrs.

Time saved by automation of unit testing is

$$T_{s=} T_m - T_a$$

T_s -- Time Saved by automation
T_m -- Time taken for manual unit testing
T_a –Time taken for automation unit testing

For our application the time saved due to automation is

$$T_s = 80 - 4$$
$$= 76 \text{ hrs}$$

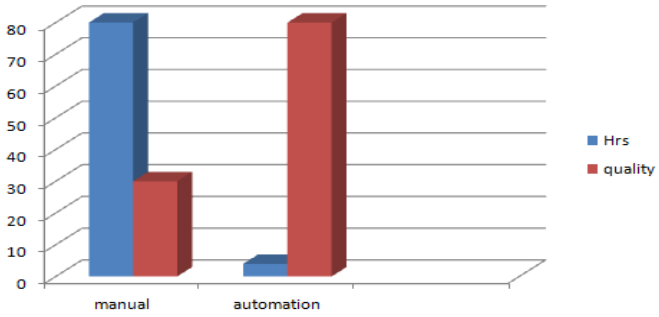


Fig. 7. Results in Graph

The above figure shows graphical representation of time saved through automation in terms of time and quality.

7 Conclusion

White.NUnit framework is more useful to automate the windows applications developed in .Net. It saves developer's time by identifying the defects in application at early stages. In order to produce application software with high reliability, maintainability and keeping testing costs low, the automation of unit testing should be preferred.

References

1. Geras, A., Smith, M., Miller, J.: A Prototype Empirical Evaluation of Test Driven Development. In: International Symposium on Software Metrics (METRICS), Chicago, IL, pp. 405–416 (2004)
2. Restivo, A., Aguiar, A.: Towards detecting and solving aspect conflicts and interferences using unit tests. In: Proceedings of the 5th Workshop on Software Engineering Properties of Languages and Aspect Technologies, Vancouver, British Columbia, Canada, p.7-es (March 2007)
3. George, B.: Analysis and quantification of Test Driven Development Approach MS Thesis. Computer Science Raleigh. North Carolina State University, NC (2002)
4. George, B., Williams, L.: An Initial Investigation of Test-Driven Development in Industry. In: ACM Symposium on Applied Computing, Melbourne, FL, pp. 1135–1139 (2003)

5. Smith, B., Williams, L.: A Survey on Code Coverage as a Stopping Criterion for Unit Testing. North Carolina State University Technical Report TR-2008-22 (2008)
6. Beizer, B.: Software Testing Techniques, 2nd edn. (1990)
7. Larman, C., Basili, V.: A History of Iterative and Incremental Development. IEEE Computer 36(6), 47–56 (2003)
8. Ho, C.-W., Johnson, M.J., Williams, L., Maximilien, E.M.: On Agile performance Requirements Specification and Testing. In: Agile 2006, Minneapolis, MN, pp. 47–52 (2006)
9. Gelperin, D., Hetzel, W.: Software Quality Engineering. In: Fourth International Conference on Software Testing, Washington, DC (June 1987)
10. Arora, H.: Test Driven Development with integrated Microsoft Unit Testing Framework Environment. Mid-term Paper, NCSU Software Engineering 2010, Raleigh, North Carolina, USA, October 7, pp. 1–6 (2010)
11. Williams, L., Nagappan, N., Maximilien, E.M.: Realizing quality improvement through test driven development: results and experiences of four industrial teams. Empirical Software Engineering 13(3), 289–302 (2008)
12. Runeson, P.: A Survey of Unit Testing Practices. IEEE Software, 22–29 (July/August 2006)
13. Tool Evaluation of Microsoft's Unit Testing Framework. Himanshu Arora, Department of Computer Science, North Carolina State University
14. Basili, V.R., Shull, F., Lanubile, F.: Building Knowledge Through Families of Experiments. IEEE Transactions on Software Engineering 25(4), 456–473 (1999)
15. Williams, L., Kudrjavets, G., Nagappan, N.: On the Effectiveness of Unit Test Automation at Microsoft. In: Proceedings of the ISSRE, pp. 81–89 (November 2009)

A New Optimization Method Based on Adaptive Social Behavior: ASBO

Manoj Kumar Singh

Manuro Tech Research, Bangalore, India
mksingh@manuroresearch.com

Abstract. The interactions and influence taking place in the society could be a source of rich inspiration for the development of novel computational methods. In this paper a new optimization method called "Adaptive social behavior optimization (ASBO)" derived from abstract inherent characteristics of competition, influence and self-confidence which are involved behind making a successful social life especially with human society is presented. The characteristics of dynamic leadership and dynamic logical neighbors along with experienced self capability are taken as fundamental social factors to define the growth of individual and in result of whole society. For each entity of a society, characteristics and affect of these three factors are not being constant for whole life span, rather than function of time and present status. To define this dynamic characteristic under a social life, in ASBO, help of self-adaptive mutation strategy is opted. To establish the applicability of proposed method various benchmark optimization problems are taken to obtain the global solutions. Performance comparison between ASBO and various variation of PSO, which is another well established optimization method based on swarm social behavior, is also presented. Proposed method is simple, more generalized and free from parameters setting in working and very efficient from performance perspectives to achieve the global solution.

Keywords: Social Structure, Global optimization, Competition, Influence, Self adaptive mutation.

1 Introduction

Social behavior is a novel natural mechanism for survival and computing model based on this can be utilized to solve difficult problems efficiently and reliably. In recent year it has proven possible to identify, abstract and exploit the computational principles underlying in social structure and deploy them for scientific and industrial purposes. Beauties of such computing model are simplicity in principle and do not require the auxiliary knowledge of the problem.

The ability of an individual to mutually interact is a fundamental social behavior that is prevalent in all human and insect societies. Social interactions enable individuals to adapt and improve faster than biological evolution based on genetic inheritance alone. This is the driving concept behind the optimization algorithm introduced in this paper that makes use of the competition and influence available within a formal

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 823–831.
springerlink.com

society. Particle swarm optimization (PSO) [1][2] and Ant colony optimization(ACO) [3][4] are two very successful and established computing model already justifying the importance of above statements. These two computing model having the bias reference of social life activities either with respect to species like bird , fish or like ant.

2 Social Structure and Influencing Behavior

A group of living entities bonded through homogeneous behavior and/or physical characteristics .The formation of bonding increase the chance of survival manifold and promoting for innovation explicitly as well implicit manner. In most of the living species including human, animals or swarm individual element is not having enough quality to make him survive with high fitness for a longer period of time. In result cooperation with other element force them to form a society, where individual contribution utilize to complete or drive the progress, examples are numerous like human society, ant colony, fish schooling etc. In general the advantages to be a pert of society are: (i) increase the forging capability, (ii) increase the reproductive efficiency,(iii) increase predator avoidance(iv) increase possibility of new innovation .Social behavior is always a fascinating subject for evolutionary biologist and social philosopher. Several attempts have been made with various different species societies to understand how as a group they are so successful with respect to certain objectives. Their individual capability as a group success has given stimulation to develop the similar computing model to solve the optimization problem. Various models already exist and they are successful with respect to certain aspect. Particle swarm optimization, Ant colony optimizations are among of them.

Each society must have a leader with respect to objective as a deriving force. The time period for a leadership is dynamic; hence emergence of new leader helps the society to have better option of development with the new innovation. An individual capability of the member makes the society diverse and help to find better innovation and a leader. Neighborhood concept play very important role in society to make the individual competitive and provides the support to increase the fitness of individual. Broadly depends upon requirements two types of neighborhoods can be defined (i) logical neighborhood (ii) geographical neighborhood. Logical neighborhood can be defined as neighbor's w.r.t objective status. While geographical neighborhood related with surrounding position location of individual. It's not necessary that logical neighbors should be same as geographical neighbors. From the perspective of innovation, logical neighbors play more important role than geographical neighbors.

General format of a social structure taken in this work is assumed to have number of subpopulation under the same social structure and environment. As in case of human society there are number of subgroups having their individual recognition. Each subgroup is having the development mostly independently at the initial stage and after certain periods of time depends upon requirement; individuals form subgroups are selected to achieve the objectives. This process provides a better chance of diversity in solution perspective while encourage competition and cooperation. Development of individual entity depends upon number of influencing factors available under circumstances like leadership, neighbors, social reaction, self sensitivity etc.

3 Proposed ASBO Methods as Leader, Neighbors and Self (LNS) Model

With this simplified macro model of influencing environment in the human society, a mathematical model is developed to achieve the objective of global optimization. Assuming, a population containing numbers of individual each represents the solution of problem in hand. Each individual representing solution is appeared in direct form (not in coded format).Each individual is having a fitness value, derived by objective function. Individual having the maximum value of fitness treated as leader at present time. A group of individuals having next nearest higher value of fitness compare to an individual are treated as neighbors for that particular individual. The change in existing status because of influences is innovated by each and every member of population using eq. (1) and the next location of status given by eq. (2).

$$\Delta X(i+1) = Cg*R1*(Gbi - Xi) + Cs*R2*(Sbi - Xi) + Cn*R3*(Nci - Xi); \quad (1)$$

$$X(i+1) = Xi + \Delta X(i+1) \quad (2)$$

Where ΔX (i+1) represents the new change in i'th dimension of an individual element and Cg, Cs and Cn are adaptive progress constants ≥ 0;
Ri, $\forall$ i=1, 2, 3, are uniformly distributed random number in range of [0 1];
Gb: global best individual at present population;
Sb: self best for an individual;
Nc: center position of a group formed by an individual and its neighbors,

For a D-dimensional problem, Gb , Sb & Nc represent vectors of D-dimension.

Gb = [Gb1, Gb2, Gb3, Gb4, …… GbD];
Sb = [Sb1, Sb2, Sb3, Sb4, …… …SbD];
Nc = [Nc1, Nc2, Nc3, Nc4, …… NcD];

3.1 Working Process of ASBO

A population with good enough members is defined initially, in which each member represents solution. Initially for all members this is random values. Each member contains one more three dimensional vector, representing the information about LNS constant. Initializations of these vectors are also random. A fitness function is defined with respect to problem in hand and a fitness value obtained for each and every member of the population. Member having maximum value of fitness declared as global best (leader) at the present time. For each member neighbors factor is calculated by taking mean of neighbors values (in this paper, three members having more nearest fitness value taken as neighbors).Self best initialize for each member at the beginning is same as initial solution. Self-adaptive mutation strategies are applied to get the new set of progress constant. Gaussian mutation has applied for random perturbation in mutation. With the set of 2N possible progress constant, a set of N member selected who are having maximum fitness and accordingly N progress constant set are selected

for searching of the new position. Because each and every member having very different position and fitness values hence its necessary they should have according values of progress constant parameters rather than a unique fixed value for everyone and in all situations. Using eq. (1) and eq. (2), new status of position is achieved to create the next population.

3.2 Self Adaptive Parameters Setting

Mutation strategy for progress constant is based on concept given by Schwefel [14] as shown below.

(a) A population of N trails solution initialized. Each solution taken as a pair of real valued vector (pi, σi), for all i Є {1, 2, 3}, with three dimension corresponding to the number of progress variables. The initial components of each pi, for all i Є {1, 2, 3} were selected in accordance with a uniform distribution ranging over a presumed solution space. The values of σi, for all i Є {1, 2, 3}, the so called strategy parameters were initially set to some value.

(b) One offspring (pi', σi') generated from each parent (pi, σi) by eq. (3) and eq. (4)

$$p'i(j) = pi(j) + \sigma i(j).\, N(0,1) \tag{3}$$

$$\sigma'i(j) = \sigma i(j) \exp(\tau' N(0,1) + \tau Nj(0,1)), \forall j \in \{1, 2, 3\}. \tag{4}$$

Where pi (j), p'i (j), σi (j), σ'i (j) denote the jth component of the vectors xi, x'i, σi, σ'i respectively. N(0,1) is a random number from Gaussian distribution. Nj(0,1) is a random number, sampled a new value for each counter j using Gaussian distribution. The scaling factors τ and τ' are robust exogenous parameters which have been set to $(\sqrt{(2\sqrt{n})})^{-1}$ and $(\sqrt{(2n)})^{-1}$.

There are two phases under the whole process to get the global solution.

(i) M number of different population having same population size (PZ) initially is taken and ASBO method is applied independently up to fixed, say P'th, number of iterations. At the end, values of fitness and all progress constants are stored for each and every member from each final population. This phase will help to maintain the diversity and in result better exploration to localize the region of solution.

(ii) From all final population, depends upon the fitness, members who are having best PZ number of fitness values are selected to form new population and their existed progress constants are also taken to form the second stage single population. Over this newly generated population ASBO is applied to get the final solution. This phase will help to get the optimal solution in faster manner.

Table 1. Benchmark problems opted for performance comparison

Function	Range	$f\min$
$F_1 = \sum_{i=1}^{N-1}\left[100\left(x_{i+1}-x_i\right)^2+\left(x_i-1\right)^2\right]$	$[-30,30]^N$	0
$F_2 = \sum_{i=1}^{N}\left[\sum_{j=1}^{i} x_j\right]^2$	$[-100,100]^N$	0
$F_3 = \exp\left(-0.2\sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2}\right)-\exp\left(\frac{1}{N}\sum_{i=1}^{N}\cos(2\pi x_i)\right)+20+e$	$[-30,30]^N$	0
$F_4 = \sum_{i=1}^{N}\left[x_i^2-10\cos(2\pi x_i)+10\right]$	$[-5.12,5.12]^N$	0
$F_5 = \frac{1}{4000}\sum_{i=1}^{N}x_i^2-\prod_{i=1}^{N}\cos\left(\frac{x_i}{\sqrt{i}}\right)+1$	$[-600,600]^N$	0

4 Experimental Setup

I have taken test functions as those used by [5] in order to make comparison with my results. Experiments were carried out over 5 test functions given in Table 1 of 30 dimensions were run for self terminating situation, which is defined as consistency in result for 500 iteration with precision of 1e-200.All experiments were run 10 times independently and the average of the best fitness values were recorded. Population sizes were maintained to 500.Different variations of PSO also taken for comparison purpose and results are shown in Table 2 and in Table 3.Plot of function values with generation are shown in Fig.1 to Fig.4.Each figure upper part represent the initial development phase for 10 different population independently and lower part represents the function value by extracting the best 500 among all from first phase .From performance as shown in Table 2 it is clear that for function F1,F2,F5, ABSO outperform the various variation of cooperative coevolved PSO methods with high margin and for F3,F4 performance is very respectful and competitive.

Table 2. Performance comparison of ASBO with various variation of CCPSO [5]

Fun	CCPSO -s6	CCPSO -s6,rg	CCPSO -s6,aw	CCPSO -s6,rg,aw	CCPSO -H6	**ASBO**
F1	2.28E+07 (5.83E+07)	7.58E+06 (3.77E+07)	1.76E+01 (3.99E+00)	2.41E+01 (1.06E+01)	_	**4.92E-18 (1.54E-17)**
Rank	[5]	[4]	[2]	[3]		**[1]**
F2	2.70E-20 (7.03E-20)	9.95E+03 (2.58E+04)	4.33E-17 (3.03E-16)	1.29E+00 (5.33E+00)	1.40E-29 (1.1E-29)	**1.15E-202 (00E+00)**
Rank	[3]	[6]	[4]	[5]	[2]	**[1]**
F3	4.00E-01 (2.83E+00)	7.05E-15 (6.26E-15)	2.13E-14 (4.95E-15)	**4.44E-16 (3.49E-31)**	2.73E-12 (2.0E-12)	1.44E-13 (1.11E-13)
Rank	[6]	[2]	[3]	**[1]**	[5]	[4]
F4	2.22E+01 (6.64E+00)	8.16E+00 (5.74E+01)	**0.00E+00 (0.00E+00)**	**0.00E+00 (0.00E+00)**	7.78E-01 (1.87E-01)	2.38E+00 (2.05E+00)
Rank	[5]	[4]	**[1]**	**[1]**	[2]	[3]
F5	2.6E-02 (2.32E-02)	1.99E+01 (9.92E+01)	1.39E-02 (2.62E-02)	2.59E-03 (1.83E-02)	5.24E-02 (1.19E-02)	**2.47E-15 (1.94E-15)**
Rank	[4]	[6]	[3]	[2]	[5]	**[1]**
Avg.Rank	4.6	4.4	2.6	2.2	3.5	**2**
Final Rank	6	5	3	2	4	**1**

Performances for four different functions F1, F3, F4 and F5 have shown in Table 3 with different variation of particle swarm optimization in various literatures. Result in terms of rank of the performance proposed method much ahead compares to other methods. Convergence characteristics for functions F1, F2, F3 andF4 in both phases of proposed method have shown in Fig1 to Fig4.Each figure is having two parts, upper parts have shown the convergence for all the ten different population for 100 fixed iterations and lower part represents the convergence of final population selected best from first phase populations.

Table 3. Performance comparison of proposed ASBO with various variations of PSO

function	F1 Mean/SD/ Rank	F3 Mean/SD/ Rank	F4 Mean/SD/ Rank	F5 Mean/SD/ Rank	Avg. Rank	Final Rank
GPSO	40.70	**1.31E-14**	26.03	2.12E-02		
[6]	32.19	**2.08E-15**	07.27	2.18E-02	[6]	[5]
	[8]	**[1]**	[5]	[10]		
LPSO	28.08	8.20E-08	35.07	1.53E-03		
[7]	21.79	6.73E-08	06.89	4.32E-03	[6]	[5]
	[7]	[6]	[7]	[4]		
SPSO-20 [8]	3.13	1.28	56.00	8.47E-03		
	3.48	1.00	15.75	9.79E-03	[7.25]	[8]
	[2]	[10]	[9]	[8]		
SPSO-40 [8]	13.50	3.73E-02	41.03	7.48E-03		
	14.63	0.19	11.09	1.25E-02	[7]	[7]
	[4]	[9]	[8]	[7]		
FIPS	25.12	2.33E-07	65.10	9.01E-12		
[9]	0.51	7.19E-08	13.39	1.84E-11	[6.25]	[6]
	[6]	[7]	[10]	[2]		
HPSO-TVAC	23.91	7.29E-14	9.43	9.75E-03		
	26.51	3.00E-14	3.48	8.33E-03	[5.25]	[3]
[10]	[5]	[3]	[4]	[9]		
DMS-PSO	41.58	1.84E-14	27.15	6.21E-03		
	30.25	4.35E-15	06.02	8.14E-03	[5.75]	[4]
[11]	[9]	[2]	[6]	[6]		
CLPS	11.36	3.66E-07	**9.09E-05**	9.02E-09		
[12]	09.85	7.57E-08	**1.25E-04**	8.57E-09	[3.75]	[2]
	[3]	[8]	**[1]**	[3]		
OPSO	49.61	6.23E-09	6.97	2.29E-03		
[13]	36.54	1.87E-09	3.07	5.48E-03	[5.75]	[4]
	[10]	[5]	[3]	[5]		
ASBO	**4.9E-18**	1.44E-13	2.38E+00	**2.47E-15**		
	1.5E-17	1.11E-13	2.05E+00	**1.94E-15**	**[2]**	**[1]**
	[1]	[4]	[2]	**[1]**		

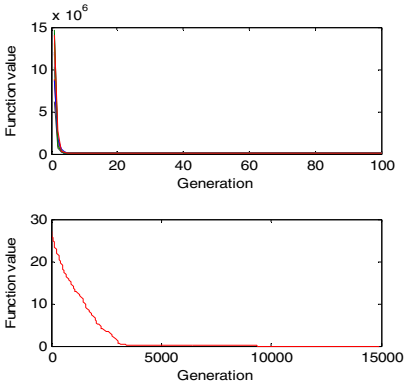


Fig. 1. Generalized Rosenbrock Function (F1)

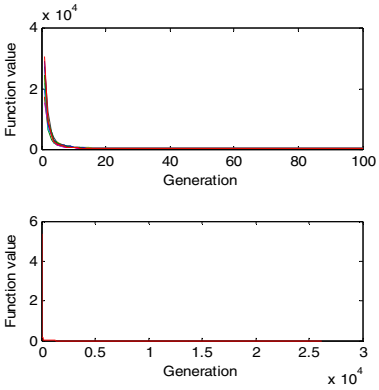


Fig. 2. Quadratic finction (F2)

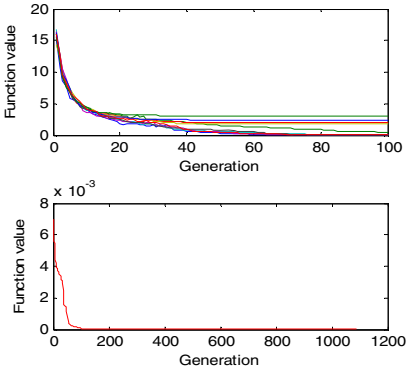


Fig. 3. Ackley Function (F3)

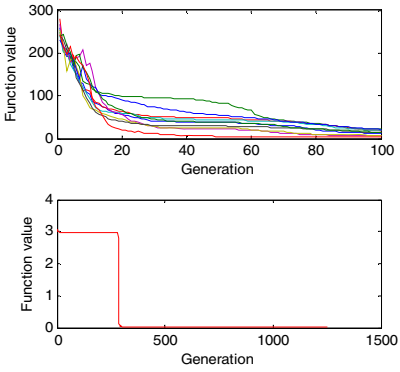


Fig. 4. Generalized Rastringin Function (F4)

5 Conclusions

A new method of optimization inspired by social life of species (especially with human) is presented. Concept of inspiration, confidence and competition are taken as social operator to improve the individual fitness. Dynamic state of individual is obtained by mutation strategy. Importance of logical neighbors in social life has great importance because they stimulate the entity up to great extent. Comparison with existing solution concept by cooperative coevolving PSO and other various variations are presented. Proposed model is simple, adaptive and very efficient. It is hope that optimization by ASBO will open a new path for research and various more macro parameters of social life will transform as a set of operator to improve quality of solution further.

Acknowledgments. I would like to say thanks to Reeta Singh and all associated members of Manuro tech research for criticism, suggestions and valuable discussion over the presented work.

References

[1] Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks 1995, vol. 4, pp. 1942–1948.

[2] Clerc, M., Kennedy, J.: The Particle Swarm—Explosion Stability, and Convergence in a Multidimensional Complex Space. IEEE Trans. Evol. Comput. 6(1), 58–73 (2002)

[3] Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: optimization by a colony of cooperating agents. IEEE Trans. Systems, Man, Cybernet.-Part B 26(1), 29–41 (1996)

[4] Dorigo, M., Stutzle, T.: Ant Colony Optimization. MIT Press, Cambridge (2004)

[5] Li, X., Yao, X.: Tackling high dimensional nonseparable optimization problems by cooperatively coevolving particle swarms. In: IEEE Congress on Evolutionary Computation, CEC 2009, pp. 1546–1553 (2009)

[6] Shi, Y.H., Eberhart, R.C.: A modified particle swarm optimizer. In: Proc. IEEE World Congr. Comput. Intell., pp. 69–73 (1998)

[7] Kennedy, J., Mendes, R.: Population structure and particle swarm performance. In: Proc. IEEE Congr. Evol. Comput., pp. 1671–1676 (2002)

[8] Particle Swarm Central, `http://www.particleswarm.Info`

[9] Mendes, R., Kennedy, J., Naves, J.: The fully informed particle swarm: Simpler, maybe better. IEEE Trans. Evol. Comput. 8(3), 204–210 (2004)

[10] Ratnaweera, A., Halgamuge, S., Watson, H.: Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients. IEEE Trans. Evol. Comput. 8(3), 240–255 (2004)

[11] Liang, J.J., Suganthan, P.N.: Dynamic multi-swarm particle swarm optimizer. In: Proc. Swarm Intell. Symp., pp. 124–129 (June 2005)

[12] Liang, J.J., Qin, A.K., Suganthan, P.N., Baskar, S.: Comprehensive learning particle swarm optimizer for global optimization of multimodal functions. IEEE Trans. Evol. Computation 10(3), 281–295 (2006)

[13] Ho, S.-Y., Lin, H.-S., Liauh, W.-H., Ho, S.-J.: OPSO.: Orthogonal particle swarm optimization and its application to task assignment problems. IEEE Trans. Syst., Man, and Cybern. A 38(2), 288–298 (2008)

[14] Back, T., Schwefel, H.P.: An overview of evolutionary algorithm for parameter optimization. Evolutionary Computation 1(1), 1–23 (1993)

[15] Singh, M.K.: Localization in wireless sensor network using merged population based particle swarm optimization: LMPPSO. In: 2011 Annual IEEE India Conference (INDICON), December 16-18 (2011)

Design and Implementation of Interval Type-2 Single Input Fuzzy Logic Controller for Magnetic Levitation System

Anupam Kumar, Manoj Kumar Panda, and Vijay Kumar

Indian Institute Technology Roorkee, Roorkee-247667, India
anuanu1616@gmail.com
{pandadec,vijecfec}@iitr.ernet.in

Abstract. This paper deals with the design and implementation of an interval type-2 single input fuzzy logic controller (IT2SIFLC) for a Magnetic Levitation System (MLS). The type-2 fuzzy controller is designed in such a way that it can be implemented with signed distance method. With help of signed distance, interval type-2 fuzzy input variable in our simple fuzzy logic controller called type-2 single input FLC. This research work is mainly focused on suspending the steel ball without any mechanical support in desired position with help of an efficient controller which uses less number of rules and less processor's time complexity. For this task we have proposed an interval type-2 single input fuzzy logic controller (IT2SIFLC) based on theory of type-2 fuzzy logic systems (T2FLS) and single input theory of fuzzy logic control. Which has the advantage of the total number of rules are abruptly reduced compared to IT2FLC. Fuzzy logic based on interval value sets is capable of modelling the uncertainty and precision in a better way. However, in real time application uncertainty associated with the available information is always occurs. The proposed controller performance is compared with the conventional fuzzy logic controller i.e. type-1 fuzzy logic controller (T1FLC), IT2FLC controller designed with the help of interval type-2 fuzzy inference system toolbox in the MATLAB-Simulink environment. Simulation results show that the proposed controller is fast with high degree of uncertainty. Simulation results analysed for all the controllers and validated in the real time model of the MLS. The proposed IT2SIFLC is surpassing the performance obtained with other controllers and is cleared from the computed time response parameters.

Keywords: Magnetic Levitation System, IT2FLC, Fuzzy logic control, IT2Single Input FLC, real-time plant.

1 Introduction

Magnetic Levitation System (MLS) is very unstable, nonlinear complex and usable system that can be applied in many application area such as in high speed transport, magnetic bearing system, vibration isolation, levitation of wind power generation and fusion Energy Materials processing in magnetic levitation furnaces**.** Magnetic levitation system is highly non-linear and unstable. Non-linearity is present due to

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 833–840.
springerlink.com

electromechanical dynamics. This is challenging and interesting task for control engineers and researcher to control MLS. PID control is classical technique for controlling of nonlinear system. It is simple, easy to implement and applicable but process of design is linear. In reference of fuzzy [1] it is shown that PID can be made nonlinear by using fuzzy logic. PID control is classical technique for controlling of nonlinear system. It is simple, easy to implement and applicable but process of design is linear. In reference of fuzzy [1] it is shown that PID can be made nonlinear by using fuzzy logic.

Type-1 FLC is unable to handle the linguistic and numerical uncertainties which are associated with dynamic unstructured environment. But type-2 fuzzy sets have the capability to determine the exact membership function for a specified fuzzy set [8]. In the design of type-1 fuzzy systems, uncertainty is limited with the linguistic uncertainty contained within the definition of variables. It is assumed that there is no uncertainty in the definition of membership functions, although parameters of the membership functions are determined by the opinions of experts who have different experiences and knowledge [6]. Unlike a type-1 fuzzy set where the membership grade is a crisp number in [0, 1], a type-2 fuzzy set is characterized by a fuzzy membership function. Each element of a type-2 fuzzy set is a fuzzy set in [0, 1]. A type-1 fuzzy set is characterized by a two-dimensional, membership function whereas a type-2 fuzzy set is characterized by a three-dimensional membership function [8, 12].

The rest of the paper is organized as follows: Section 2 briefly describes mathematical model of MLS. Section 3 reviews about the type-2 fuzzy sets and interval type-2 fuzzy logic systems (IT2FLS). Section 4 describes the design and implementation of proposed Interval type-2 single input Fuzzy logic controller (IT2SIFLC) and other existing controllers for MLS their results in section 5, followed by conclusion presented in section 6.

2 Mathematical Model of MLS

The Magnetic levitation system is steel ball levitation system, in which a steel ball will levitate in air without any mechanical support. As shown in fig.1 MLS consists of an electromagnet coil, steel ball, sensor, and LED source. MLS mathematical model [4] is based on the ball kinematics and electrodynamics equations. Applying Newton's second law of motion in vertical direction.

$$F(i,x) + mg = m\left(\frac{d^2x}{dt^2}\right) \tag{1}$$

Where F(i,x), x, m, i and g are air gap, steel ball mass, current and acceleration of gravity respectively.

$$F(i,x) = K(i/x)^2 \tag{2}$$

Where K is constant. For linearized at equilibrium point (i_0, x_0) in equation (2)

$$F(i,x) = F(i_0,x_0) + F_i(i_0,x_0)(i-i_0) + F_x(i_0,x_0)(x-x_0) \tag{3}$$

In which $F(i_0,x_0)$ is equilibrium force when the air gap is x_0 and current i_0.

$$F(i_0,x_0) = mg \tag{4}$$

$$K_i = F_i(i_0, x_0) = \left.\frac{\delta F(i,x)}{\delta i}\right|_{i=i_0, x=x_0} = \frac{2Ki_0}{x_0^2} \tag{5}$$

$$K_x = F_x(i_0, x_0) = \left.\frac{\delta F(i,x)}{\delta x}\right|_{i=i_0, x=x_0} = -\frac{2Ki_0^2}{x_0^3} \tag{6}$$

where K_i and K_x the stiffness coefficient. From (4),(5),(6),(3)and(1) then we get

$$F(i,x) = K_i i + K_x x = F(i_0, x_0) \tag{7}$$

$$m\frac{d^2x}{dt^2} = K_i(i - i_0) + K_x(x - x_0) \tag{8}$$

The voltage equation of the electromagnetic coil is given by

$$U(t) = Ri(t) + L(di/dt) \tag{9}$$

Now taking Laplace transform and putting $mg = -K(i_0 / x_0)^2$, then system

$$m\frac{d^2x}{dt^2} = K_i(i - i_0) + K_x(x - x_0) = \frac{2Ki_0}{x_0^2} i - \frac{2Ki_0^2}{x_0^3} x \tag{10}$$

$$x(s)s^2 = \frac{2Ki_0}{mx_0^2} i(s) - \frac{2Ki_0^2}{mx_0^3} x(s) \tag{11}$$

$$\frac{x(s)}{i(s)} = \frac{-1}{As^2 - B} \tag{12}$$

Where $A = i_0/2g$, $B = i_0/x_0$, Define the input and output variable as U_{in}, U_{out}.

$$G(s) = \frac{U_{out}(s)}{U_{in}(s)} = \frac{K_x x(s)}{K_a i(s)} = \frac{-\left(\frac{K_s}{K_a}\right)}{As^2 - B} \tag{13}$$

Then the system state variables are $x_1 = U_{out}, x_2 = \dot{U}_{out}$ and state equations are as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{2g}{x_0} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ -\frac{2gK_s}{i_0 K_a} \end{bmatrix} u_{in} \tag{14}$$

Real system model parameter are as m=0.022kg, Ka=5.8929, Ks=458.7204, i_0 = =0.6105A, $x_0 = 0.03m$ and r(radius of ball)=0.0125m, these parameter put in eq.14

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 653.4 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 2499.1 \end{bmatrix} u_{in}$$

$$y = x_1$$

After calculating the pole of system, it is found that system has open loop poles on the left plane. Therefore the GML system is unstable system.

3 Interval Type-2 Fuzzy Logic Systems

The concept of a fuzzy set of type-2 was introduced by Zadeh as an extension of concept of an ordinary fuzzy set [5]. A FLS that uses at least one type-2 fuzzy set is called a type-2 FLS.The background material in this section is taken from [11].An interval type-2 fuzzy set (IT2FS) $\tilde{A}$ is characterized as [7],[28]:

$$\tilde{A} = \int_{x\in X} \int_{u\in Jx\subseteq[0,1]} 1/(x,u) = \int_{x\in X} [\int_{u\in Jx\subseteq[0,1]} 1/u]/x \tag{16}$$

Where x, the primary variable, has domain X; $u \in U$, the secondary variable, has domain J_x at each $x \in X$; J_x is called the primary membership of x and is defined in (5); and, the secondary grades of $\tilde{A}$ all equal 1. Note that (16) means $\tilde{A} : X \rightarrow \{[a,b]: 0 \le a \le b \le 1\}$. Uncertainty about $\tilde{A}$ is conveyed by the union of all the primary memberships, which is called the footprint of uncertainty (*FOU*) of $\tilde{A}$ (Fig.3), i.e.

$$FOU(\tilde{A}) = \bigcup_{\forall x\in X} J_x = \{(x,u) : u \in J_x \subseteq [0,1]\} \tag{17}$$

The upper membership function (UMF) and lower membership function (LMF) of $\tilde{A}$ are two type-1 MFs that bound the FOU. The UMF is associated with the upper bound of FOU associated with the lower bound of FOU ($\tilde{A}$) and is denoted by ($\tilde{A}$) and is denoted by $\overline{\mu}_{\tilde{A}}(x), \forall x \in X$ and the LMF is $\underline{\mu}_{\tilde{A}}(x), \forall x \in X$ i.e.

$$\overline{\mu}_{\tilde{A}}(x) \equiv \overline{FOU(\tilde{A})}, \forall x \in X \tag{18}$$

$$\underline{\mu}_{\tilde{A}}(x) \equiv \underline{FOU(\tilde{A})}, \forall x \in X \tag{19}$$

Note that J_x is an interval set, i.e.

$$J_x = \{(x,u) : u \in [\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x)] \tag{20}$$

So that FOU(Ã) in (17) can also be expressed as

$$FOU(\tilde{A}) = \bigcup_{\forall x\in X} [\underline{\mu}_{\tilde{A}}(x), \overline{\mu}_{\tilde{A}}(x)] \tag{21}$$

For continuous universes of discourse X and U, an embedded IT2FS $\tilde{A}_e$ is

$$\tilde{A}_e = \int_{x\in X} [1/u]/x, \quad u \in J_x \tag{22}$$

Note that (22) means: $\tilde{A}_e : X \rightarrow \{u: 0 \le u \le 1\}$. The set $\tilde{A}_e$ is embedded in $\tilde{A}$ such that at each x it only has one secondary variable (i.e., one primary membership whose secondary grade equals 1) Examples of $\tilde{A}_e$ are $1/\overline{\mu}_{\tilde{A}}(x)$ and $1/\underline{\mu}_{\tilde{A}}(x)$, $\forall x \in X$.

4 Controller Design and Implementation

The purpose of the controller is to keep the steel ball suspended in air, at the nominal equilibrium position by controlling the current in the magnet.

4.1 The Design Methodology for T1FLC and IT2FLC

First the T1FLC controller is designed and implemented for the control of MLS. In case of, IT2FLC is designed with the help of interval type-2 fuzzy inference system toolbox in the MATLAB software [9-10]. The controller is designed based on the block diagram shown in Fig.2.The two inputs considered for the design of controller are error (E) and change in error (CE). The controlled output (CV) is fed to the plant. In this proposed controller design an interval type-2 Gaussian membership function with uncertain mean is chosen where the standard deviation value is fixed. The membership function plot is shown below in the Fig.6. The membership function is expressed.

$$\mu(x) = \exp\left\{\frac{-1}{2}\left[\frac{(x-m)}{\sigma}\right]^2\right\} \quad m \in [m_1, m_2] \tag{23}$$

4.2 The Design Methodology for IT2SIFLC

Single input fuzzy logic controller uses a single input variable. Generally, input variable in IT2FLC are error (E) and change of error (CE) taken. In this paper, we presented the new variable signed distance method, which has the advantage of the total number of rules are abruptly reduced compared to T2FLC. Signed distance method [2-3] is applicable, when the rule base of FLC is in skew-symmetric form. The rule base for two inputs (error and rate of change of error) FLC is given in fig 7. For any combination of (E, CE), the output membership function will lie in any one of the diagonal line (L_{NL}, L_{NM}, L_{NS}, L_Z, L_{PS}, L_{PM}, L_{PL}). Main diagonal line (L_Z) can be representation by

$$\text{Diagonal line:} \quad \dot{e} + \lambda e = 0 \tag{24}$$

Where variable λ is the slope, the distance from any point $(e, \dot{e})$ to the main diagonal line can be written as

$$\text{Distance:} \quad d = \frac{\dot{e} + \lambda e}{\sqrt{1+\lambda^2}} \tag{25}$$

Depending on the distance d, the new rule base can be constructed from fig.7 given in fig.8. Rule table is one dimensional and contains only seven rules. The membership function for input (d) and control output (CV) is shown in fig.6.

5 Simulation Results and Discussion

The proposed controller IT2SIFLC is applied for the control MLS and its performance is plotted in the fig.3 and is compared with the other controller's performance and is summarized in table 1. Fig.4 shows the error response of position. To test the robustness of the controller, the MLS is simulated by varying the parameter of plant without changing the structure of the controllers. Output response, in Fig.5 for the T1FLC is not satisfactory.

The performance comparison table 4 clearly indicates that the settling time of proposed IT2SIFLC is very less compared to other controllers and the overshoot is reduced with minimum rise time. There is no steady state error in the output by using proposed controller.

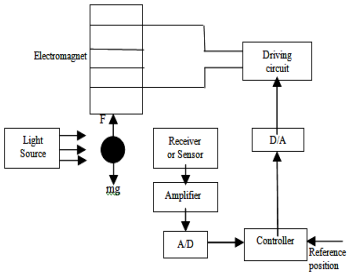


Fig. 1. Block diagram of MLS

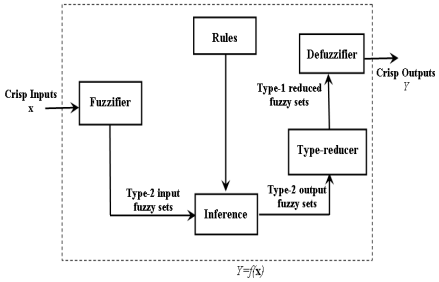


Fig. 2. Interval Type-2 FLC

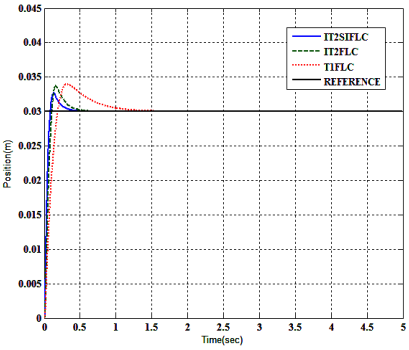


Fig. 3. Output response for controllers

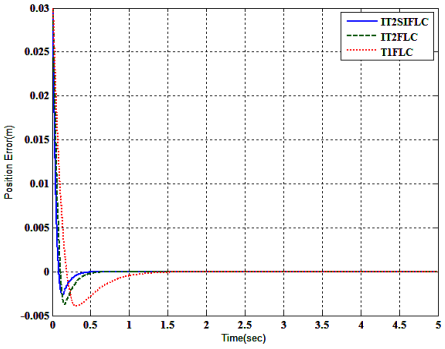


Fig. 4. Error Output response for controllers

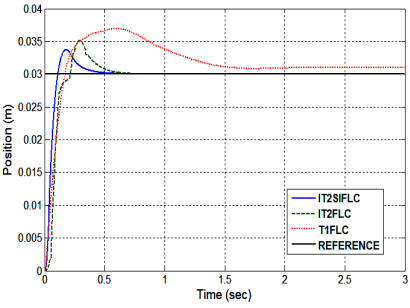


Fig. 5. Output response with parameter variation

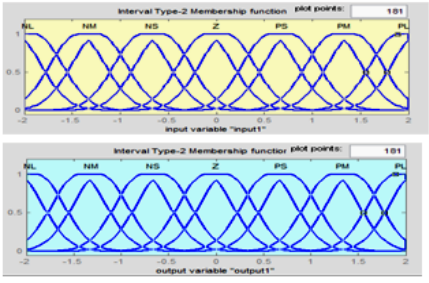
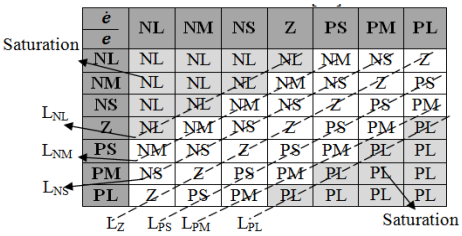


Fig. 6. Membership functions for input d and Control Output

Table 1. Performance comparison of controllers

S.N	Overshoot (%)	Rise time(s)	Settling time(s)
IT2SIFLC	6.67	0.1	0.3
IT2FLC	12.00	0.12	0.48
T1FLC	13.30	0.17	1.10



ė / e	NL	NM	NS	Z	PS	PM	PL
NL	NL	NL	NL	NL	NM	NS	Z
NM	NL	NL	NL	NM	NS	Z	PS
NS	NL	NL	NM	NS	Z	PS	PM
Z	NL	NM	NS	Z	PS	PM	PL
PS	NM	NS	Z	PS	PM	PL	PL
PM	NS	Z	PS	PM	PL	PL	PL
PL	Z	PS	PM	PL	PL	PL	PL

Fig. 7. IT2FLC rule base

d	L_{NL}	L_{NM}	L_{NS}	L_Z	L_{PS}	L_{PM}	L_{PL}
u	NL	NM	NS	Z	PS	PM	PL

Fig. 8. IT2SIFLC Rule base

6 Conclusion

This paper presents an interval type-2 single input fuzzy logic controller for a magnetic levitation system. The effectiveness of the proposed controller has been investigated through simulation studies. The simulation shows better performance of the proposed controller compared to an IT2FLC and IT1FLC controller. The proposed controller does not require heavy computations and much processor complexity, therefore implementation is feasible. The performance comparison is summarized in table 1 and clearly indicates that the proposed controller is fast and giving improved performance compared to other existing controllers. The proposed controller is more robust compare to conventional T1FLC with parameter variation and disturbances. The simulation results are validated with the experimental real time model of MLS, developed by googol tech.

Acknowledgments. This research work is supported by the MHRD, Govt. of India and Indian Institute of Technology, Roorkee.

References

1. Hu, B., Mann, K.I., Gosine, R.G.: New Methodology for Analytical and Optimal Design of Fuzzy PID Controllers. IEEE Trans. on Fuzzy System 7(5), 521–538 (1999)
2. Choi, B.-J., Kwak, S.-W., Kim, B.K.: Design of a Single-Input Fuzzy Logic Controller and Its Properties. Fuzzy Sets and Systems 106(3), 299–308 (1999)
3. Choi, B.J., Kwak, S.W., Kim, B.K.: Design and Stability Analysis of Single-Input Fuzzy Logic Controller. IEEE Transaction on Systems, Man and Cybernetics-Part B: Cybernetics 30(2), 303–309 (2000)
4. Googol Technology Ltd., GML series Magnetic Levitation System User Manual and Experimental Book (2007)

5. Hagras, H.: Type-2 FLCs: a new generation of fuzzy controllers. IEEE Comput. Intell. Mag. 2(1), 30–43 (2007)
6. Mendel, J.M., Hagras, H., John, R.I.: Standard background material about interval type-2 fuzzy logic systems that can be used by all authors, http://www.ieee-cis.org/_files/standards
7. Mendel, J.M., John, R.I.B.: Type-2 fuzzy sets made simple. IEEE Trans. Fuzzy Syst. 10(2), 117–127 (2002)
8. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning, Parts 1, 2, and 3. Information Sciences 8, 9, 199–249, 301-357, 43-80 (1975)
9. Castilo, O., Melin, P., Castro, J.R.: Computational intelligence software for interval type-2 fuzzy logic. In: Proc. Workshop on Building Computational Intelligence and Machine Learning Virtual Organizations, pp. 9–13 (2008)
10. Castilo, O.: IT-2 fuzzy logic toolbox for use with MATLAB. Software developed by research group of O. Castillo at Tijuana IT Mexico
11. Dereli, T., Baykasoglu, A., Altun, K., Durmusoglu, A., Turksen, I.B.: Industrial applications of type-2 fuzzy sets and systems: A concise review. Computers in Industry 62, 125–137 (2011)
12. Maity, S., Sil, J.: Color image segmentation using type-2 fuzzy sets. Int. Journal of Comup. and Elect. Eng. 1(3), 376–383 (2009)

Mining Negative Association Rules from Multiple Data Sources on the Basis of Local Pattern Analysis

T. Ramkumar[1], S. Selvamuthukumaran[1], S. Hariharan[2], and V. Harikrishnan[1]

[1] A.V.C. College of Engineering, Tamilnadu, India
ramooad@yahoo.com
[2] TRP Engineering College, Tirichirappalli, Tamilnadu, India

Abstract. While positive association rules identify the co-occurrences of frequent item-sets, negative association rules find out the negation relationships of frequent items-sets by forming occurrence of an item-set characterized by the absence of others. The notion of negative association rule is very useful in customer-driven domains such as market basket analysis for identifying products that conflict with each other. When data are scattered in multiple data sources that are located in different regions, negative relationships among frequent item-sets are very important for arriving decisions both at strategic and branch levels. This paper presents an approach for mining negative association rules from multiple data sources and synthesizing global negative association rules voted by most of the data sources. The proposal has been justified experimentally by using a bench marked database found in UCI machine learning repository.

1 Introduction

With the advent of information and communication technologies, enterprises business transactions are scattered over multiple branches located in different regions. An interstate business organization has a need over analyzing their branch databases which are distributed in many sites. Traditional way of mining such distributed databases has been done using centralized mining approach. The centralized mining integrates all the branch databases into a huge repository and mining will be taken place. Such approach is called as 'Mono-Mining' which is data warehouse oriented. This approach suffers significant limitations such as information and privacy factors of branch databases, huge amount of data movement from branch to central head and high investment in hardware and software. To overcome such drawbacks, researchers focus on local pattern analysis strategy as an alternative solution for distributed data mining. On the basis of local pattern analysis strategy, patterns from individual databases are mined and forwarded to the centralized process. Due to the pattern forwarding nature, issues related to integration of heterogeneity in databases, privacy are sorted out. Besides market basket analysis, Association rule mining algorithms are employed in various domains such as web usage mining, intrusion detection and bioinformatics. An association rule is in the form of Antecedent → Consequent [2]. Two measures namely support and confidence is used to asses the interestingness of association rule. A strong rule is one which satisfies the minimum support and minimum confidence threshold values chosen by the domain expert. Like positive association rules,

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 841–846.
springerlink.com

negative association rules are also very important in data analysis and decision-making process since it identifies the occurrence of one frequent item by the absence of others. A strong negative association rule identifies the negative relationship between frequent item-sets. if the positive association rule is represented in the form of A→B, then the negative association rule would be represented either in one of the following form (i) A→¬B (ii)) ¬A → B (iii) ¬A → ¬B, where 'A' and 'B' are the frequent item-sets. Since databases are scattered over different regions with distinct features, identifying negative relationship among forwarded patterns and synthesizing global negative association rules at central head will be very useful both at branch level and central level. In this paper, an approach for synthesizing negative association rule which are voted by most of the branch databases is presented.

Brin et. al [4] introduced the negative relationships between variables by using chi-square based model. To determine the nature of relationship, a correlation metric has been proposed by them. Savasere et. al[8] combined positive frequent item-sets with domain knowledge in the form of taxonomy to mine negative associations. Wu et. al[9] derived an algorithm for generating both positive and negative association rules. They have contributed a new measure called mininterest for a better pruning of the frequent item-sets generated. Antonie and Zaine [3] proposed an algorithm that extends the support-confidence framework with a sliding correlation coefficient threshold. Dong et.al[5] proposed MLMS (Multiple Level Minimum Supports) model which uses multiple level minimum supports to discover infrequent itemsets and frequent itemsets simultaneously. Also they have refined their work by proposing a new measure VARCC[6] (Valid Association Rule Based on Correlation Coefficient and Confidence) which combines correlation coefficient and minimum confidence to generate positive and negative association rules correctly from the frequent and infrequent item-sets discovered by the MLMS model. Zhang et al. [11] introduced the local pattern analysis strategy, which efficiently overcomes limitations caused by centralized mining. Wu and Zhang [10] proposed a weighting model for synthesizing high-frequency rules from multiple databases.. Adhikari and Rao [1] made an attempt for synthesizing heavy association rules in multi-database environment and reported whether a heavy association rule is high-frequent or exceptional in nature. Ramkumar and Srinivasan [7] proposed a weighting model for synthesizing high-frequent association rules from multiple data sources based on the transactions-population of data sources.

2 Mining Negative Rules from Multiple Data Sources

Let A and B are the frequent item-sets, then the positive association rule is represented as A→B and the corresponding negative association rules are represented either in any of the forms (A→¬B) or (¬B→A) or (¬A→¬B), where "¬A", "¬B" represent the negation of item-sets A and B respectively.

$$\text{supp}(\neg A) = 1 - \text{supp}(A) \tag{1}$$

$$\text{supp}(A \rightarrow \neg B) = \text{supp}(A) - \text{supp}(AUB) \tag{2}$$

$$\text{supp}(\neg A \rightarrow B) = \text{supp}(B) - \text{supp}(AUB) \tag{3}$$

$$\text{supp}(\neg A \rightarrow \neg B) = 1 - \text{supp}(A) - \text{supp}(B) + \text{supp}(AUB) \tag{4}$$

$$conf(A \rightarrow \neg B) = (supp(A) - supp(AUB)) / supp(A) \quad (5)$$

$$conf(\neg A \rightarrow B) = (supp(B) - supp(AUB)) / (1 - supp(A)) \quad (6)$$

$$conf(\neg A \rightarrow \neg B) = (1 - supp(A) - supp(B) + supp(AUB)) / (1 - supp(A)) \quad (7)$$

Synthesizing negative association rules mined from individual data sources consists of the following steps; (i) Generate set of negative association rules from the branch data sources which satisfy local support and local confidence thresholds (ii) Eliminate un-interesting negative association rules by using rule selection threshold; $\min.\gamma_{effective}$ (iii) Synthesize negative association rules on the basis of site weights and local support and local confidence values. (iv) Identify global negative association rules whose global support and global confidence values are above than the user specified threshold value. Let $S_1, S_2,\ldots,S_m$ be the set of branch databases taking part in the process of mining negative association rules. $w'_{s1}, w'_{s2},\ldots,w'_{sm}$ are the weights corresponding to the transactions-populations of the data sources. In general, w'_{sj} is the un-normalized site weight of site j, based on the transactions-population. The Normalized weight of a site is the ratio between the transactions-population in the respective data-source and the total of the transactions-populations of all the participating sites.

$$\text{Normalized weight of Site } j = w_{sj} = \frac{w'_{sj}}{\sum_{j=1}^{m} w'_{sj}} \quad (8)$$

$$\gamma_{effective}(Ri) = \sum_{j=1}^{m} \delta(i, j) * w_{sj} \quad (9)$$

$$Supp_G(Ri) = \sum_{j=1}^{m} w_{sj} * Supp_j(Ri) \quad (10)$$

$$Conf_G(Ri) = \frac{Supp_G(Ri)}{Supp\_ante_G(Ri)} \quad (11)$$

$\delta(i, j) = 1$ if Ri is present in site j otherwise $\delta(i, j) = 0$ which is the percentage of votes received from different data sources for a given rule on the basis of the transactions-populations of the corresponding data sources. The local support and local confidence of a negative rule Ri at site j is represented as $Supp_j(Ri)$ and $Conf_j(Ri)$ respectively. Support for the antecedent of negative rule Ri at site j is represented as $Supp\_ante_j(Ri)$. $Supp_G$, $Conf_G$ are the synthesized support and synthesized confidence respectively. Let us apply the proposed approach for the example shown in Table 1.

Table 1. Rules voted by the sites

Rule	S1 (20000)	S2 (10000)	S3 (5000)	S4 (5000)
R1=A→¬B	0.45,0.60	0.27,0.40	0.35,0.50	0.30,0.60
R2=¬C→D	0.25,0.40	0.35,0.60	0.25,0.40	-
R3=¬E→¬F	0.40,0.50	-	-	-
R4= ¬G→H	-	-	-	0.35,0.45

$W'_S = 20000 + 10000 + 5000 + 5000 = 40000;$
$w_{s1} = 0.50;\ w_{s2} = 0.25;\ w_{s3} = 0.125;\ w_{s4} = 0.125\ ;$
$W_{R1} = 1 * 0.50 + 1 * 0.25 + 1 * 0.125 + 1 * 0.125 = 1.0;$
$W_{R2} = 1 * 0.50 + 1 * 0.25 + 1 * 0.12 = 0.875;$
$W_{R3} = 1 * 0.50 = 0.50;$ WR4=1*0.125=0.125

By Choosing $\min.\gamma_{\text{effective}}$ as 0.50, rule R_4 is eliminated. Though R3 and R4 are voted by a single site, R3 is participating in the synthesizing process due to its respective data source, S1's transactions-population. The synthesized global support and confidence values are calculated as

$Supp_G(R1) = (0.50 * 0.45) + (0.25 * 0.27) + (0.125 * 0.35) + (0.125 * 0.30) = 0.3737$
$Conf_G(R1) = (0.50 * (0.45/0.60)) + (0.25 * (0.27/0.40)) + (0.125 * (0.35/0.50)) +$
$(0.125 * (0.30/0.60)) = 0.6937$
$Supp_G(R2) = (0.50 * 0.25) + (0.25 * 0.35) + (0.125 * 0.25) = 0.2187$
$Conf_G(R2) = (0.50 * (0.25/0.40)) + (0.25 * (0.35/0.60)) + (0.125 * (0.25/0.40)) = 0.5364$
$Supp_G(\neg E \rightarrow \neg F) = (0.50 * 0.40) = 0.20$
$Conf_G(\neg E \rightarrow \neg F) = 0.50 * (0.40/0.50) = 0.40$

3 Experimental Study

To evaluate our proposed approach, we have made an experimental investigations on mushroom data set found in UCI machine learning repository. The data set contains 8124 transactions with 120 distinct items and the average length of transaction is 24. The entire database has been divided into four data sets namely, S1, S2, S3 and S4 with respective transactions-populations of 3000, 2000, 2000 and 1124. The minimum support threshold has been chosen as 0.30 for generating negative frequent itemsets/association rules from individual sites and for global synthesizing process. The effective vote-rate has been chosen as 0.66 to find candidature for synthesizing process. By the proposed approach, the normalized site weight of S1, S2, S3 and S4 are 0.3692, 0.2461, 0.2461 and 0.1383 respectively. We found 24 negative association rules that are emerged as candidature for synthesizing and supported by all the four sites. (effective vote rate =1).Over 37 negative association rules supported by three sites are also emerged as candidature for global rule synthesizing process by satisfying the effective vote rate of 0.66. We found 79 negative association rules, supported by two sites only and fail to emerge as candidature for rule synthesizing. While Synthesizing, 25 negative association rules among the candidature fail to attain the global threshold value of 0.30. The rest of 36 rules are turned as global negative association rules and their corresponding synthesized support and confidence values are shown in Table 2.

Table 2. Global Negative association rules with their synthesized support and confidence

Rule-Id	Rule	$Supp_G$	$Conf_G$
2	¬ 10→36	0.4436	0.7385
3	¬ 63→90	0.3683	0.9386
4	¬ 59→90	0.3388	0.9335
5	¬ 53→90	0.3545	0.8191
6	¬ 53→86	0.4082	0.9431
7	¬ 63→86	0.3915	0.9975
8	¬ 63→85	0.3924	1
9	¬ 59→86	0.3619	0.9973
10	¬ 59→85	0.3629	1
11	¬ 53→85	0.4328	1
16	¬ 10→86	0.5771	0.9607
17	¬ 10→85	0.6007	1
18	36→¬ 63	0.3235	0.3858
20	¬ 53→36	0.3776	0.8726
21	¬ 52→85	0.5672	0.9584
22	¬ 63→34	0.3902	0.9944
23	¬ 59→34	0.3607	0.994
24	¬ 53→34	0.4069	0.9403
27	¬ 3→34	0.4365	0.7573
30	¬ 6→34	0.4919	0.778
31	¬ 56→86	0.3998	0.74
33	¬ 10→34	0.5037	0.8385
35	¬ 56→85	0.4019	0.7439
38	¬ 6→90	0.4674	0.7393
42	¬ 6→86	0.4919	0.778
44	¬ 10→90	0.4839	0.8055
45	¬ 28→85	0.3689	0.5998
46	¬ 6→85	0.4938	0.7811
48	¬ 3→90	0.409	0.7097
49	¬ 3→86	0.4366	0.7575
50	¬ 3→85	0.4379	0.7599
53	¬ 13→85	0.4964	0.6685
54	¬ 52→86	0.4906	0.865
58	¬ 56→34	0.3997	0.7398
59	¬ 52→34	0.4906	0.865
60	¬ 93→85	0.3126	0.5594

4 Conclusion

In multi-database mining context, discovering significant patterns from multiple databases has been accounted as an important issue always. While positive association rules identify the co-occurrence of frequent items, negative association rules are used to find out items that conflict with each other. In this paper, negative association rules are mined from multiple databases using local pattern analysis strategy and synthesized on the basis of data source weights. Example and experimental investigations presented in the paper claim the validity of the proposed approach.

References

1. Adhikari, A., Rao, P.R.: Synthesizing heavy association rules from different real data sources. Pattern Recognition Letters 29, 59–71 (2008)
2. Agrawal, R., Srikant, R.: Fast Algorithms for Mining Association Rules. In: Proceedings of the Twentieth International Conference on Very Large Databases, Chile, pp. 478–499 (1994)
3. Antonie, M.-L., Zaïane, O.R.: Mining Positive and Negative Association Rules: An Approach for Confined Rules. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) PKDD 2004. LNCS (LNAI), vol. 3202, pp. 27–38. Springer, Heidelberg (2004)
4. Brin, S., Motwani, R., Silverstein, C.: Beyond market basket: Generalizing association rules to correlations. In: Proceedings of SIGMOD, pp. 265–276 (1997)
5. Dong, X., Niu, Z., Shi, X., Zhang, X.-d., Zhu, D.: Mining Both Positive and Negative Association Rules from Frequent and Infrequent Itemsets. In: Alhajj, R., Gao, H., Li, X., Li, J., Zaïane, O.R. (eds.) ADMA 2007. LNCS (LNAI), vol. 4632, pp. 122–133. Springer, Heidelberg (2007)
6. Dong, X., Niu, Z., Zhu, D., Zheng, Z., Jia, Q.: Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model. In: Tang, C., Ling, C.X., Zhou, X., Cercone, N.J., Li, X. (eds.) ADMA 2008. LNCS (LNAI), vol. 5139, pp. 444–451. Springer, Heidelberg (2008)
7. Ramkumar, T., Srinivasan, R.: Modified algorithms for synthesizing high-frequency rules from different data sources. Knowledge and Information System 17, 313–334 (2008)
8. Savasere, A., Omiecinski, E., Navathe, S.: Mining for strong negative associations in a large database of customer transactions. In: Proceedings of ICDE, pp. 494–502 (1998)
9. Wu, X., Zhang, C., Zhang, S.: Efficient Mining of both Positive and Negative Association Rules. ACM Transactions on Information Systems 22, 381–405 (2004)
10. Wu, X., Zhang, S.: Synthesizing high-frequency rules from different data sources. IEEE Transactions on Knowledge and Data Engineering 15, 353–367 (2003)
11. Zhang, S., Wu, X., Zhang, C.: Multi-Database Mining. IEEE Computational Intelligence Bulletin 2, 5–13 (2003)

Recognition of Hand Punched Kannada Braille Characters Using Knowledge Based Multi Decision Concept: Basic Symbols

Srinath S. and C.N. Ravi Kumar

Image Processing and Computer Vision Lab, JSS Research Foundation
Department of Computer Science and Engineering,
S.J. College of Engineering
Mysore City, Karnataka State, India - 570 006
{srinath_mysore,kumarcnr}@yahoo.com

Proverb says "A blind person who sees is better than seeing person who is blind".

Abstract. Technology has made the visually impaired life more comfortable. Access to textual information has become simple to the visually impaired community. Braille is widely used as communication tool for visually impaired people. The reading and writing communication for the visually impaired is available through the Braille language all over the world. Braille notations are available for most of the languages all over the world. Kannada is a south Indian language, which follows syllable writing method. People who work in association with visually impaired and can't understand Braille notation, require conversion of Braille document into a normal language representation. This paper presents an efficient algorithm for converting a Kannada Braille symbols into its equivalent normal version.

Hand punched Kannada Braille document is used for the experimentation. The document image is segmented to extract Braille characters one at a time and compared with the knowledge base created to identify the equivalent normal version of it. Six different knowledge bases are created to reduce the search time complexity.

Keywords: Braille, Optical character recognition, Nearest -Neighbour classifier, Optical Braille Character Recognition, Kannada OBR.

1 Introduction

Globally an estimated 45-50 million people all over the world are blind and nearly 140-150 million have low vision and this number increasing every year. Louis Braille developed the system of embossed writing nearly 200 years back. It is the accepted communication media for among the blind community for reading and writing. Except some modifications the Braille notations remains same. Today blind and partially sighted people communicate widely through Braille.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 847–857.
springerlink.com

A System of embossed dots is used in the Braille language. Every character of a language is represented using 6 dots arranged in 3 rows and 2 columns and the dots are numbered from 1 to 6. Having 6 dots in this pattern, there are possible 64 dot combinations as shown in Figure -1[18].

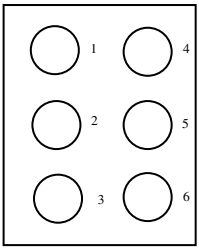


Fig. 1. The six dots of a Braille cell

Braille notations can be adapted to any natural languages. The different combinations of the dots are used to represent different characters of the language. So a language having character set up to 64 can be easily moulded with the Braille notations. Every language has its own Braille writing, however the sighted people by looking at the Braille document cannot understand in what language it is written.

Many attempts have been made to optically recognize embossed Braille using various methods [1-17].

The dimensions of Braille dots have been set according to the tactile resolution of the fingertips of a person. The horizontal and vertical distance between dots in a character, the distance between cells representing a word and the inter-line distance are also specified by the library of congress. Dot base diameter is approximately 1.5 mm. The distance between the centres of two dots within a character cell is approximately 2.3mm horizontally and 2.5mm vertically. The distance between dots in adjacent cells is approximated to 3.75mm horizontally and 5mm vertically as shown in Figure -2.

The standard Braille sheet is of size 11 inches wide and 11.5 inches in height. A Braille sheet contains 25 lines horizontally and 40-42 Braille cells in each line [18].

Braille writing cannot be processed with standard optical character recognition (OCR). This is due to the fragmented nature of the characters (character is a collection of embossed dots), and the fact that characters on both side can be seen simultaneously. A different approach has to be considered. Hence the need for development of Optical Braille character recognition system.

A Braille optical recognition system is interesting and useful due to several reasons:

It is an excellent communication tool for sighted people (who do not know Braille) with blind writing.

It is a cut-price alternative to Braille to Braille copy machine instead of the current complex devices which use a combination of heat and vacuum to form Braille impressions.

Braille writing is read using the finger so it is necessary to touch the document, for this reason the book after many readings will deteriorate.

It is useful to store a lot of documents written by blind authors in Braille that were never converted to digital information.

It also helps to preserve and multiply large volumes of manually crafted Braille books, as it will be difficult to retype due to the special rules that apply in Braille.

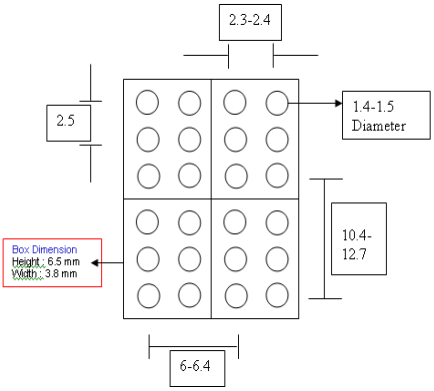


Fig. 2. Spacing between Braille characters and spacing between dots within a Braille character (Dimensions in Millimetres)

1.1 Kannada Braille

Kannada is one of the major Dravidian languages of Southern India and one of the earliest languages evidenced epigraphically in India. Kannada, is the official language of the state of Karnataka situated in the southern part of India. Nearly 50 million people all over the world speak kannada and is is whose native speakers are called Kannadigas (ಕನ್ನಡಿಗರು *Kannadigaru*), number roughly 50 million making it the 27th most spoken language in the world. It is one of the scheduled languages of India and the official and administrative language of the state of Karnataka [19].

The language uses forty-nine phonic letters, divided into three groups: *swaragalu* (vowels – thirteen letters); *vyanjanagalu* (consonants – thirty-four letters); and *yogavaahakagalu* (neither vowel nor consonant – two letters: the *anusvara* ಂ and the *visarga* ಃ), similar to the vowels and consonants of English. The character set is almost identical to that of other Indian languages. The Kannada character set is as shown in Figure -3

ಅ ಆ ಇ ಈ ಉ ಊ ಋ ೠ ಌ
ಎ ಏ ಐ ಒ ಓ ಔ ಅಂ ಅಃ
ಕ ಖ ಗ ಘ ಙ
ಚ ಛ ಜ ಝ ಞ
ಟ ಠ ಡ ಢ ಣ
ತ ಥ ದ ಧ ನ
ಪ ಫ ಬ ಭ ಮ
ಯ ರ ಲ ವ ಶ ಷ ಸ ಹ ಳ

Fig. 3. Kannada Character Set

Kannada Braille is used widely at schools for blind in the Karnataka state of India. This paper aims at developing a Kannada Optical Braille Recognition (Kannada OBR) system, which means converting a Kannada Braille document into its equivalent normal Kannada version as soft copy. Braille notations for the Kannada symbols are as shown in Figure – 4.

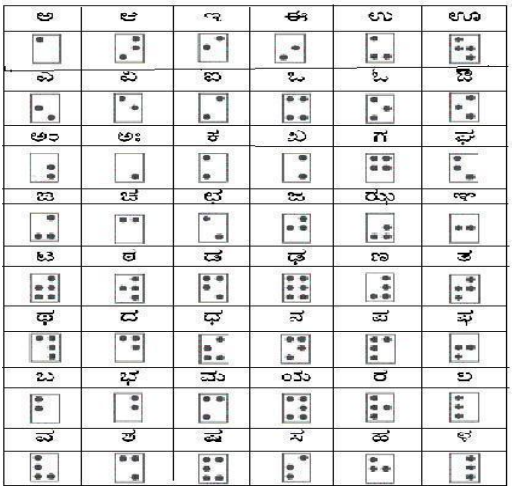


Fig. 4. Normal Kannada characters and its respective Braille version characters

Several researchers have made efforts to recognize Braille documents. Different approaches have been introduced to recognize and analyse Braille documents of English, Arabic and other languages, but so far no productive effort has been executed to develop an OBR system for the Kannada language to the best of our knowledge.

Early work was carried out by J. Dubus [1] lacto Braille to recognize Braille characters.

Mennens and their team in their work [2, 3], proposed algorithm to recognize both single and double sided Braille documents.

Both English and Chinese Braille conversion was proposed in the paper authored by C.Ng and V.Lau [4].

Murry and Dais designed a handheld device which handles the scanning as well as the translation of Braille document [5-6].

A hybrid system using a neural network to solve the recognition problem was designed by Morgavi and Morando [7].

An algorithm was proposed based on the half character recognition at a time in the work of Lisa Wong and his team [8].

Image processing techniques were used by Nestro Falcon and his team for Braille writing recognition [9].

An effort to recognize Arabic Braille recognition was found in the paper presented by Abdul Malik and his team in 2007 [10] and in the paper presented by Shanjun ZHANG and his team [15].

Ahenfei Tai and his team presented an adaptive Braille document parameter estimation method to automatically determine the skewness, indentations and spacing in both vertical and horizontal directions [11].

Amany Al-Saleh and team separated Braille image as recto and verso dots, which enabled them to analyse and recognise Braille document punched on both sides of a sheet in one scan [12].

With the popularity of Mobile devices new techniques were introduced to detect the Braille characters using mobile devices [13, 14].

Looking into the previous works on Braille character recognition, literature survey does not show any development for the recognition of Kannada Braille characters. Adapting Braille to the English language is easy compared to South Indian languages including Kannada. The reason is number of basic characters in Kannada is more than 500, where as in English we have only 26. The 64 combinations available by the 6 dots of a Braille cell can be used easily for representing all characters of English.

This paper aims at developing algorithms to recognize the hand punched basic symbols of the Kannada Braille language

Input to the algorithm is Braille document Image, punched in Braille Kannada on only one side of a Braille sheet and is digitized using a HP DeskJet F4200 with optical resolution of 2400x2400 DPI. Experiments have been done using digital cameras, by taking picture of the Braille document from different distances. But we realized that scanner results in more efficient, suitable, easy and economical output. If a camera is used, along with the resolution, the distance from which the image is captured of the document also matters.

The successive section elaborates on other aspects like: The Section 2 discusses about the different steps followed in the design of the algorithm. The design of knowledge base used in the design of two levels nearest neighbourhood method is discussed in Section 3. Experimental results can be seen in Section 4. Section 5 gives the conclusion remarks.

2 Proposed Algorithm

Acquire the Digital Image of the Braille document using a scanner. The Braille document used in the experiment is hand punched by blind persons on only one side of a thick white sheet. Braille slate and stylus is used in embossing the Braille characters. Basic symbols of Kannada characters and numbers are punched.

As we are not interested in the colour of the sheet or background, the digitized image is converted into binary image by applying threshold and then processed.

The digitized binary image will have variation in colour intensity for the dots of the Braille character due to the embossed nature of it. When the scanner light falls on it the reflection angle will give us the information about the presence of dot or absence, for a character.

In the process of scanning along with the actual dots some other area might also carry the predicate of the character hole. This is computed based on the number of connected components. In the experiment conducted it is observed that if it is less than 30 connected component having the same predicate as that of the real Braille cell is considered as noise and is eliminated.

Braille documents do not include different font size and style. Each Braille character is of same size, the character spacing and line spacing is also of predefined measurement. This feature is utilized effectively, to segment the given document. The Braille document image is segmented line by line and then character by character.

As the Braille character is a collection of dots which are not connected. The dots of one character may be miss-interpreted as dots of neighbouring characters. Care has been taken to compute the distance between inter-character and inner-character.

When you look at the Braille document it looks like a collection of dots. The challenge lies in segmenting the Braille dots into character lines and each line to characters.

When a black and white Braille image is taken, we have some information regarding the presence of dots. The embossing nature of the dots reflects the scanner light in a different angle compared to other areas and black spots are created. Each pixel row of the Braille document is verified and row numbers containing top edge and bottom edge of the Braille dots are noted from top to bottom of the Braille document. The row numbers containing the top edges of the Braille dots are used to understand whether Braille character dots are spread in 3 lines or 2 lines. Segmenting the Braille document is as shown in the Figure – 5. Based upon this information, the dots are grouped.

In the next stage, the left and right edge of each dot is identified. Again the distance between each dot is used to group the dots vertically into a character box. Once this is achieved, one character is extracted at a time and processed.

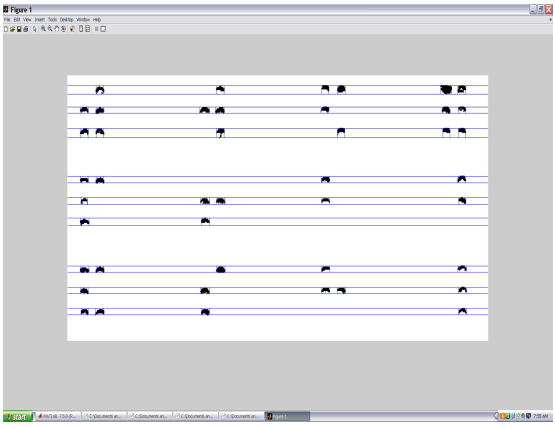


Fig. 5. Segmented Braille documents

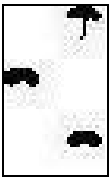


Fig. 6. Extracted Single Kannada Braille character

The character extracted from the document is as shown in figure 6 and is subdivided into 6 regions. The division is done so that the six dot positions of the Braille character are individualized.

A histogram of each subregion is computed. Then based on the threshold value six different variables are assigned with value one or zero.

In the first level we locate the positions of 1's and 0's in the extracted Braille character and based on the positional values the first level neighbour is identified in terms of the appropriate knowledge base. Now the searching is reduced to a single knowledge base with less number of elements.

Once the knowledge base is known, in the second level, the nearest neighbour is identified within it.

Character is compared each record of the knowledge base and the successful equivalent normal Kannada character is written into the text file. Bharaha Kannada text editor is used for writing the normal Kannada character. Process is repeated with the subsequent characters on one line and for all the lines of a document.

Different types of knowledge bases are created to reduce the time complexity. Detailed discussion about the knowledge base created is discussed in the next section.

Instead of using single level database, the two levels, reduces the computing time. The time complexity is reduced based on the character, the number of dots present in it and its position. When considered 49 basic characters of Kannada, an estimated time reduction in the proposed method is as given in table 1.

Table 1. Estimated time reduction

Sl. No	Dots position in the Braille character	Estimated % of reduction in the time complexity
1.	Only first and Second rows filled	73
2.	Only second and third row filled	85
3.	First and third row filled	73
4.	Only first column filled	90
5.	Only second column filled	88
6.	All the three rows are used to form a character	60

The proposed algorithm has the following steps.

Step 1: Digitize the document

Step 2: Convert into Binary Image

Step 3: Eliminate the noise

Step 4: Perform line Segmentation

Step 5: Segment the Characters.

Step 6: Recognize the Braille character and equivalent normal Kannada character is recorded into another document.

Step 7: Repeat the process with all the characters in a line and for all the lines of Braille document.

Step 8: Stop

3 Knowledge Base

Knowledge base is created which contains the data set designed with the knowledge. Different types of knowledge base are created to make the time complexity efficient.

The knowledge sets are created using MS-Excel. The different knowledge set is as shown below.

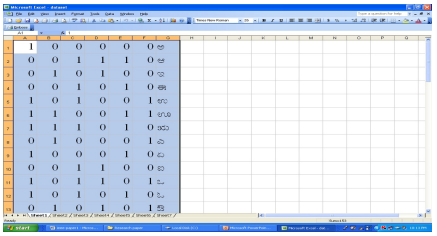


Fig. 7. Knowledge set for all 49 characters of Kannada Braille

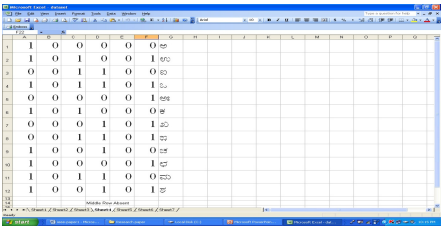


Fig. 10. First and third row filled

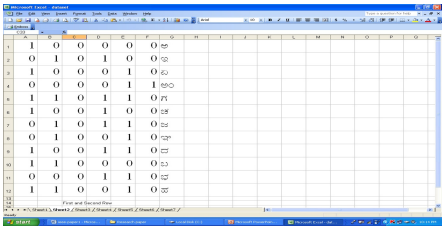


Fig. 8. Only first and Second rows filled

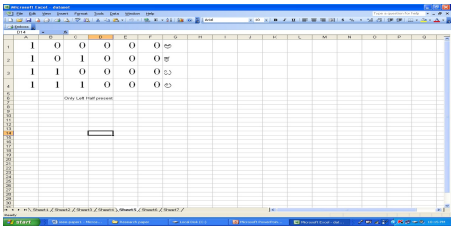


Fig. 11. Only first column filled

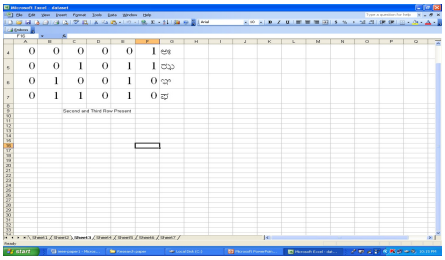


Fig. 9. Only second and third row filled

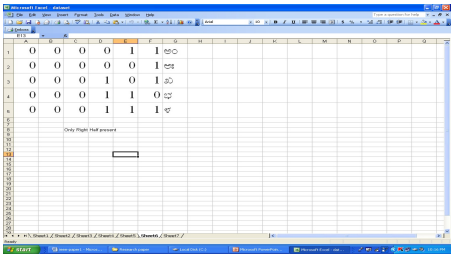


Fig. 12. Only second column filled

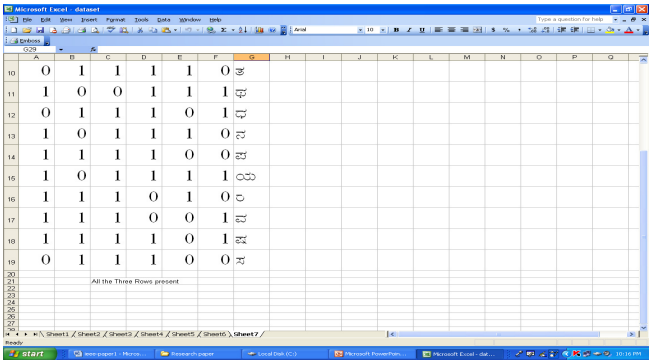


Fig. 13. All the three rows are used to form a character

4 Results

For the experimentation purpose, 10 different visually impaired students studying in the class 10 of the local blind school were requested to emboss all the basic characters of Kannada on a thick white sheet. Only one side of the sheet is used and embossing is done using a Braille slate and stylus.

The experiment we conducted has successfully converted all the basic characters of the Kannada Braille into normal Kannada. As the experiment is conducted in an ideal situation having Kannada Braille characters embossed on only one side of white thick sheet, more than 99% accuracy is achieved.

Accuracy cited above is based on the experiments conducted considering all the 10 Braille documents. The extracted characters of the Braille document is then compared with the knowledge base created.

The actual Braille document image and its related output is as shown below. Figure – 14 Shows the scanned Braille document image containing the basic symbols of Kannada language. Figure – 15 Shows the output of the algorithm after processing each Braille character, one at a time.

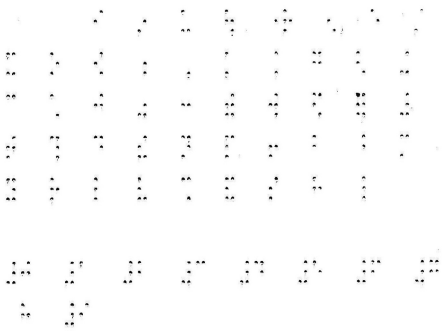


Fig. 14. Braille Document Image

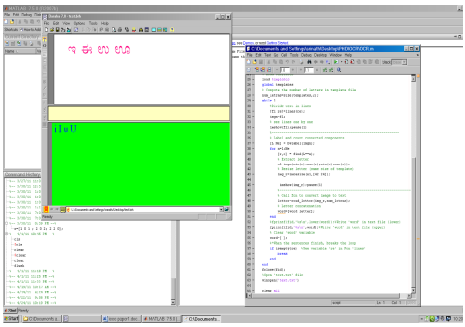


Fig. 15. Normal Kannada Equivalent written into Bharaha document file

5 Conclusion

Multi decision concept using Histogramic approach is used in recognizing the Hand punched Kannada Braille characters. We have created 6 different knowledge bases to reduce the time complexity. All the basic characters of Kannada Braille were recognized. The experiment has facilitated the converting the hard copy Braille document into a soft copy version of it. The experimentation produced good and satisfactory results in the ideal conditions.

To the best of our knowledge, it is the first and important mile stone in achieving recognition and mapping of Hand Punched Kannada Braille Characters into normal readable version.

Acknowledgement. My sincere thanks to JSS Research Foundation for providing all facilities in conducting experiments related to this paper.

References

[1] Dubus, J.P., Benjelloun, M., Devlaminck, V., Wauquier, F., Altmayer, P.: Image Processing techniques to perform an autonomous System to translate relief Braille back into ink called LectoBraille. In: IEEE 10th International Conference in Medicine and Biology Society, New Orleans, pp. 1584–1585 (November 1988)

[2] Mennens, J.: Optical recognition of Braille writing. IEEE, 428–431 (1993)

[3] Mennens, J., Va Tichelen, L., Francois, G., Engelen, J.J.: Optical Recognition of Braille writing using Standard Equipment. IEEE Trnsactions on Rehabilitation Engineering 2(4) (December 1994)

[4] Ng, C., Ng, V., Lau, Y.: Regular feature extraction for recognition of Braille. In: Proceedings of Third International Conference on Computational Intelligence and Multimedia Applications, ICCIMA 1999, pp. 302–306 (1999)

[5] Murray, I., Dias, T.: A portable device for optically recognizing braille - part i: hardware development. In: The Seventh Australian and New Zealand Intelligent Information Systems Conference 2001, pp. 129–134 (2001)

[6] Murray, I., Dias, T.: A portable device for optically recognizing braille - part ii: software development. In: The Seventh Australian and New Zealand Intelligent Information Systems Conference 2001, pp. 141–146 (2001)

[7] Morgavi, G., Morando, M.: A neural network hybrid model for an optical Braille recognitor. In: International Conference on Signal, Speech and Image Processing, ICOSSIP 2002 (2002) CD ROM

[8] Wong, L., Abdulla, W., Hussmann, S.: A software Algorithm prototype for Optical Recognition of Embossed Braille. In: Proceeding of 17th International Conference on Pattern Recognition, ICPR 2004 (2004)

[9] Falcón, N., Travieso, C.M., Alonso, J.B., Ferrer, M.A.: Image Processing Techniques for Braille Writing Recognition. In: Moreno Díaz, R., Pichler, F., Quesada Arencibia, A. (eds.) EUROCAST 2005. LNCS, vol. 3643, pp. 379–385. Springer, Heidelberg (2005)

[10] Al-Salman, A.M., Alohai, Y., ALKanhal, M., AIRajith, A.: An Arabic Optical Braille Recognition System. In: ICTA 2007, Hammamet, Tunisia, pp. 12–14 (2007)

[11] Tai, Z., Cheng, S., Verma, P.: Braille Document Parameters Estimation for Optical Character Recognition. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Porikli, F., Peters, J., Klosowski, J., Arns, L., Chun, Y.K., Rhyne, T.-M., Monroe, L. (eds.) ISVC 2008, Part II. LNCS, vol. 5359, pp. 905–914. Springer, Heidelberg (2008)

[12] Al-Saleh, A., El-Zaart, A., AlSalman, A.: Dot Detection of Optical Braille Images for Braille Cells Recognition. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 821–826. Springer, Heidelberg (2008)

[13] Rantala, J., Raisamo, R., Lylykangas, J., Surakka, V., Raisamo, J., Pakkanen, K.S.T., Hippula, A.: Methods for Presenting Braille Characters on a Mobile Device with a TouchScreen and Tactile Feedback. IEEE Transactions on Haptics 2(1) (January-March 2009)

[14] Zhang, S., Yoshino, K.: A Braille Recognition System by the Mobile Phone with Embedded Camera (2010)

[15] Al-Shamma, S.D., Fathi, S.: Arabic Braille Recognition and Transcription into Text and Voice. In: 5th Cairo International Biomedical Engineering Conference, Cairo, Egypt, December 16-18 (2010)

[16] Srinath, S., Ravi Kumar, C.N.: An Insight into Optical Braille Character Recognition since its conceptualization. International Journal of Computer Applications 33(6) (November 2011)

[17] Srinath, S., Ravi Kumar, C.N.: Recognition of Kannada Braille Characters using Histogram Approach. In: National Conference on Advancement in Computer Applications, FISAT, Angamaly Ernakulam Dist, Kerala, India (December 2011)

[18] http://www.dimensionsguide.com/dimensions-of-a-standard-braille-sheet/

[19] http://en.wikipedia.org/wiki/Kannada

Vascular Tree Segmentation in Fundus Images Using Curvelet Transform

Rupu Kumari[1], Charul Bhatnagar[2], and Anand Singh Jalal[2]

[1] Banasthali University, Jaipur
roop4frnd@gmail.com
[2] GLA University, Mathura
charul@gla.ac.in, anandsinghjalal@gmail.com

Abstract. In retinal images, vessel segmentation methods are an important component of circulatory blood vessel analysis systems. This paper introduces an effective approach to segment the vessels in the fundus images. The fundus images are first enhanced using curvelet transform, then segmentation is performed using morphological operations with a modified structuring element and length filtering. The proposed method has been tested on 40 images of the DRIVE database. The results demonstrate that the proposed algorithm segments blood vessels in the retinal images effectively with an accuracy of 94.33%.

1 Introduction

The eye can be affected by a number of systemic diseases. Problems in the eye may be the first presentation of a systemic disease or patients with known systemic problems may need to have their eyes specifically checked for complications. Some common systemic problems are Diabetic Retinopathy, Hypertensive Retinopathy etc.

Reliable vessel extraction is a primary step for retinal image analysis and processing because vessels are the predominant and most stable structures appearing in the fundus images. Like rest of the body, there are two types of blood vessels in retina - arteries and veins. A blockage in either retinal veins or arteries can affect sight. A retinal vascular tree is distinctive enough to each individual and can be used for biometric identification. Retinal vessel tree segmentation has some clinical objective, like evaluation of retinopathy prematurity, implementation of screening program for diabetic retinopathy, arteriolar narrowing, vessels diameter measurement to diagnose hypertensive retinopathy, cardiovascular diseases and retinal vessel tortuosity to characterize hypertensive retinopathy [1]. However, there are many challenges which need to be handled for robust segmentation of the blood vessels [2]. Vessels may have low contrast; some vessels may have similar intensity as that of the background; the vessel width may vary from 1 pixel to 12 pixels.

There is a need to develop an algorithm to segment blood vessels in a fundus image in short time and with high accuracy. In this paper, a two step methodology is proposed for the segmentation of the vascular tree in fundus images. In the first step, a curvelet transform based approach is used to enhance the retinal image. Then mathematical morphology, global thresholding and length filtering is applied to segment the vascular tree in the fundus image.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 859–864.
springerlink.com

Candes et al. [3] introduced a multiscale transform which is known as curvelet transform. The two important features of curvelet as geometrical transform are scaling and the directionality. These features make curvelet transform more suitable for sparse representation of objects with edges. It is also useful for handling image singularity better than other multiscale transform. In [4] [5] the authors introduced the second generation of curvelet transform (also known as Discrete Curvelet Transform (DCT)) based on a frequency partition technique. The DCT is very efficient in representing curve like edges. Therefore, it is a good choice to represent retinal vessel tree.

2 Proposed Methodology

In the proposed approach, first the enhancement of retinal vessels is performed by the curvelet transform. It is then followed by a segmentation process.

Fig. 1. Steps in Retinal Vessel Tree Segmentation

2.1 Retinal Image Contrast Enhancement Using Curvelet Transform

Since the curvelet transform represents edges better than any other multiscale transform, it is therefore well suited for multi-scale edge enhancement. Curvelet co-efficient can be modified in order to enhance edges in an image and to get better contrast of the image. For a fundus image, the following non-linear function, as defined in [6], is used to modify the value of the curvelet co-efficient.

$$E_c(x,\sigma) = 10 \quad \text{if} \quad x \le c\sigma$$

$$E_c(x,\sigma) = 5*\left(\frac{x-c\sigma}{c\omega}\left(\frac{m}{c\sigma}\right)^p + \frac{2c\sigma - x}{c\sigma}\right) \text{ if } x < 2c\sigma.$$

$$E_c(x,\sigma) = 5*\left(\frac{m}{x}\right)^p \text{ if } \quad 2c\sigma \le x \angle m$$

$$E_c(x,\sigma) = 10*\left(\frac{m}{x}\right)^s \text{ if } \quad x > m.$$

In the above equation σ is the noise standard deviation [7], s introduces the dynamic range of compression, p determines the degree of non-linearity and c is the normalized parameter, if its value is larger than 3, it guarantees that noise will not get amplified. m can be derived from noise standard deviation as $m = k_m\sigma$ where k_m is independent of curvelet co-efficient, here $k_m = 10$.

Curvelet enhancement of Retinal image contrast consists of the following steps:

1. Calculate the curvelet co-efficient of the input image $I_{(x,y)}$. A set of bands B_i is obtained, where each band B_i contains C_i co-efficient correspond to a given resolution level.
2. Estimate the noise standard deviation σ of input image $I_{(x,y)}$.
3. Calculate σ_i for each band i of curvelet transform and for each band i multiply each curvelet coefficient $B_{i,k}$ by $E_c\left(\left|B_{i,k}\right|,\sigma_i\right)$
4. Reconstruct the enhanced image from the modified curvelet coefficients.

2.2 Vascular Tree Segmentation in Fundus Images

Steps for vascular tree segmentation:

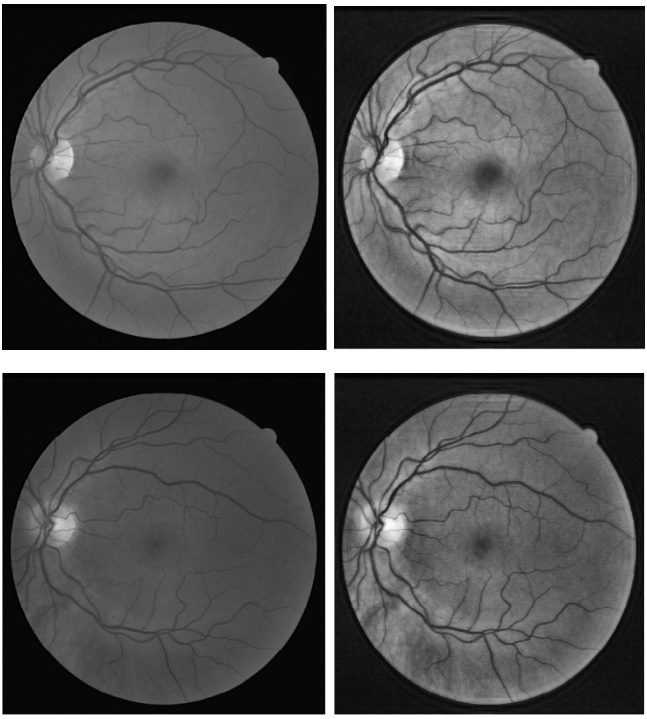
1. Enhance the fundus image using curvelet transform.
2. Get the fundus mask, by applying global thesholding [8] on green band of fundus image followed by morphological closing.
3. Get region of interest (ROI), by multiplying fundus mask with curvelet enhance image.
4. Apply morphological opening with a structuring element to enhance image to suppress the retinal vessel from retinal images while preserving the shape and size of other objects in the image [9].
5. Subtract morphologically opened image by curvelet enhanced image.
6. Apply Otsu's thesholding [8] to get the binary image.
7. Apply length filtering on binary image.

3 Results and Discussion

The proposed algorithm has been tested on 40 retinal images of the DRIVE database [10]. Out of 40 colour fundus images of the DRIVE data set, 33 do not have any signs of disorder while 7 show signs of mild early disorders. In order to evaluate the performance of the proposed approach, structure similarity [11] and peak SNR [12] measures are used. The structural similarity index is computed between the segmented vascular tree using the proposed method and manually segmented vascular tree (provided with DRIVE Database). Structural similarity can be measured as follows:

$$SSIM = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2Y_{x,y} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(Y_x^2 + Y_y^2 + C_2\right)}$$

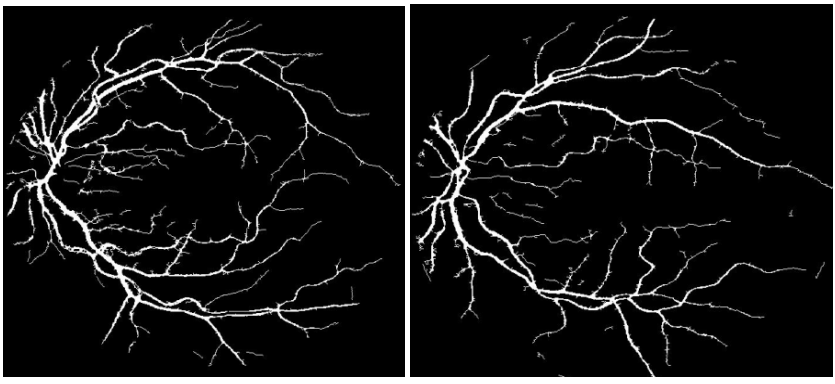
where μ_x, μ_y are average intensities, Y_x^2, Y_y^2 are variances, $Y_{x,y}$ is the covariance of I_x (Segmented image using the proposed method), I_y (Ground truth image) and $C_1 = (0.01 * L)^2$, $C_2 = (0.03 * L)^2$, L is the dynamic range of pixel values.



(a) (b)

Fig. 2. a) Green band of 01_test.tif and 05_test.tif images of DRIVE Database, b) Enhanced image using curvelet transform

Fig. 2 shows the results of green band of original image and its contrast enhanced image using curvelet transform. Fig. 3 shows the result of vascular tree segmentation by the proposed method.



(a) (b)

Fig. 3. Vascular tree of a) 01_test.tif, b) 05_test.tif images of DRIVE Database

To evaluate the image contrast enhancement, one of the objective measures is Peak SNR (Signal Noise Ratio) [12]. PSNR evaluates the intensity changes of images between original and enhanced images. PSNR can be calculated as:

$$PSNR = 20.\log_{10} \frac{MAX_I}{\sqrt{MSE}}$$

MAX_I is the maximum possible pixel value of the image and MSE is mean squared error calculated as:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [F_o(i,j) - F_e(i,j)]^2$$

where F_o and F_e are original and enhanced image respectively.

Table 1. Quantitative assessment

Measure	Average (40 images of DRIVE)	Standard deviation
SSIM	0.94	1.56
PSNR	26.43	2.08

Table 1 illustrates that average structure similarity between ground truth images of vascular tree and the segmentation results using the proposed method segmentation is .94 and enhanced image PSNR is 26.43.

4 Conclusions

This paper introduced a new algorithm for automatic vascular tree segmentation in fundus images. Retinal vessels have some specific characteristic; which makes it's segmentation difficult. In the proposed approach, we have exploited a curvelet transform to enhance the contrast of an image. FDCT has high ability to represent edges in the images, so in proposed algorithm curvelet transform coefficients are used to enhanced retinal vessel in the image and make it a good candidate for segmentation. Ostu's thesholding method is used for the vessel segmentation and length filtering to remove the false edges, while preserving the thin vessel accurately. The experimental results illustrate that the proposed approach gives satisfactory results.

References

1. Youssif, A., Ghalwash, A., Ghoneim, A.: Optic Disc Detection from Normalized Digital Fundus Images by Means of a Vessels' Direction Matched Filter. IEEE Trans. on Medical Imaging 27(1) (2008)
2. Michal, S., Charles, V.S.: Retinal Vessel Extraction using Multiscale Matched Filters, Confidence and Edge Measures. IEEE Trans. on Medical Imaging 25(12), 1531–1546 (2006)

3. Candès, E., Demanet, L.: Curvelets and Fourier Integral Operators. C. R. Math. Acad. Sci. 336(5), 395–398 (2003)
4. Candes, E., Demanet, L., Donoho, D., Ying, L.: Fast Discrete Curvelet Transforms. Society for Industrial & Applied Mathematics 5(3), 861–899 (2006)
5. Jianwei, M., Plonka, G.: The Curvelet Transform. IEEE Signal Processing Magazine 27, 118–133 (2010)
6. Starck, J., Murtagh, F., Candes, E., Donoho, D.: Gray and Color Image contrast Enhancement by the Curvelet Transform. IEEE Trans. Image Processing 12(6) (2003)
7. Zhen, Z., Jin-Sha, Y., Qiang, G., Ying-Hui, K.: Wavelet Image De-noising Method Based on Noise Standard Deviation Estimation. In: Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, Beijing (2007)
8. Otsu, N.: A Threshold Selection method from Gray level Histograms. IEEE Trans. Syst., Man, Cybern. S2A-9(1), 62–66 (1979)
9. Gonzalez, R., Woods, R.: Digital Image Processing, 3rd edn., pp. 627–679. Prentice-Hall, NJ (2008)
10. Niemeijer, M., Staal, J., Ginneken, B., Loog, M., Abràmoff, M.D.: Comparative Study of Retinal Vessel Segmentation Methods On a New Publicly Available Database. In: Proc. SPIE—Med. Image., vol. 5370, pp. 648–656 (2004)
11. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image Quality Assessment: From Error Visibility to Structural Similarity. IEEE Trans. on Image Processing 13(4), 600–612 (2004)
12. Turaga, D., Yingwei, C., Caviedes, J.: No Reference PSNR Estimation for Compressed Pictures. In: Proceedings International Conference on Image Processing, vol. 6 (2002)

A Density-Based Clustering Paradigm to Detect Faults in Wireless Sensor Network

Sourav Kumar Bhoi, Sanjaya Kumar Panda, and Pabitra Mohan Khilar

Department of Computer Science and Engineering
National Institute of Technology, Rourkela, India
{souravbhoi,sanjayauce}@gmail.com,
pmkhilar@nitrkl.ac.in

Abstract. Event detection using wireless sensor network is an emerging area of research nowadays in distributed environment. In geographical regions, it is a great area of research to set the sensors for event (volcanic eruptions) detection by taking local decisions. But due to failure of nodes in these regions, it is difficult to detect the event. In this paper, we have proposed an approach of detecting the fault by using Density-Based Clustering method. Our main idea is to form a density-based cluster in which the nodes within the cluster have same behavior (faulty or active). The cluster is formed by using ε-Neighborhood, in which the Density-Reachability and Density-Connectivity concepts are used to get the Density-Based Cluster. By this method, the faults are detected as the nodes which are not in the cluster. Our observation shows better results in modeling a Fault-Detection Paradigm to detect the faults in the network.

1 Introduction

Sensor networks are the most essential and integrated components of data communication. This consists of base stations which help in transmitting and receiving the data [7]. The nodes are mainly responsible for collecting the sensed data. These networks are mainly used in military purpose and detecting the volcanic eruptions.

Event detection is the detection of the events by using the wireless sensor networks [2, 3]. Nowadays volcanic eruptions are growing in a large manner. So, there should be detection of faults in case of node failure. Here, we detect the fault by using the sensor reading as '1' and '0'. '1' represents the detection of the event and '0' represents no detection. So, these are local decisions by which we detect the faults in the network.

So, to detect the faults in the sensor network and to make it a fault-tolerant network we have proposed an approach of detecting the fault in the node by using *Density-Based Clustering* method [8]. Our main idea is to form a density-based cluster in which the nodes have the same behavior (faulty or active). The cluster is formed by using the ε-neighborhood, in which the *Density-Reachability and Density- Connectivity* is done to get the *Density-Based Cluster*. The paper is organized as follows: section 2 presents the related work done in this field. Section 3 presents the preliminaries, where this discusses about the assumptions taken in modeling the fault-detection paradigm. In section 4, we have described about the proposed approach with

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 865–871.
springerlink.com

a description and pseudocode. Section 5 presents the illustration and observations taken. Section 6 presents the conclusion part of our proposed algorithm and at last references are presented.

2 Related Works

In the recent years many fault-detection techniques have been developed. E. Ould-Ahmed-Vall et al. [14] presented a geometric based approach to detect local detection error in wireless sensor networks in which they have used the concept of convexity. Krishnamachari et al. [9] proposed an algorithm for event region detection. Chen et al. [5] corrected the errors in distributed-Bayesian algorithm. Luo et al. [11] enhance the model in [9] by discussing about the two sources. An assumption has been taken in which, all nodes have the same probability of failure [9], [11]. E. Ould-Ahmed-Vall et al. [15] proposed an approach of considering different probability of failure levels.

Multifunction/ Multimode devices [19] are used, in which a single terminal offers multiple interfaces and deployment of overlay network occurs for survivability of IEEE 802.11. Sahoo et al. [17], [16] proposed a model for survivability against AP failure in IEEE 802.11. They have taken two main phases: *Design phase* and *Fault Response phase* to detect and check the faults in the network. Chen et al. [4] describe a technique to enhance the connection reliability in WLANs. But, due to the presence of redundant APs *co-channel interference* problems.

Li et al. analyzed the network performance by region-disjoint and node-disjoint constraints [10]. Sen et al. proposed a technique of capturing the faults by using the concept of region based connectivity [18]. Newmayer et al. identified the failed parts of a network and model the event (disaster) as a circular cut or a line by considering the graph models and also analyzed the failures in the nodes [12, 13]. Feyessa et al. proposed a technique of randomly generating the networks which supports geometric constraints for survivability [6]. Applegate et al. developed algorithms to compute the optimal restorations paths and analyzed the performance of route restoration in ISP network [1].

In our paper, we have discussed about the density-based clustering method to detect the faults in the wireless sensor networks for event detection.

3 Preliminaries

3.1 Assumptions

Assumptions taken in our proposed fault-detection method are as follows:

1) A location aware sensor network is considered in which a node knows its location and the location of other nodes [8].
2) An arbitrary wireless sensor network is considered initially.
3) If all the neighbors of node ‘p’ detect an event and node p does not detect the event, then there may be failure or fault in node ‘p’ according to density-based approach, it is not in the density-based cluster.

4) If all the neighbors of node 'p' are unable to detect the event and node 'p' detects the event than there may be a fault in ode'p' according to density-based approach, it is not in the density-based cluster.
5) If a node in a cluster has a local decision, then the nodes inside the density-based cluster have also the same local decision (all have same decision).
6) If a node inside the cluster is unable to detect an event, and other points detect the event then there is a fault in the node.
7) A radius 'ε' is defined according to the circle (coverage).
8) The nodes which are core objects (having minimum nodes) are considered to be trusted nodes.
9) The nodes forming the clusters are authenticated to each other, if there is an attack to a node inside the cluster then we can say the node is faulty in the cluster.

3.2 Network Model

We have considered a network model as an arbitrary wireless sensor network in which many arbitrary nodes are spread over an area. It is shown in figure 1.

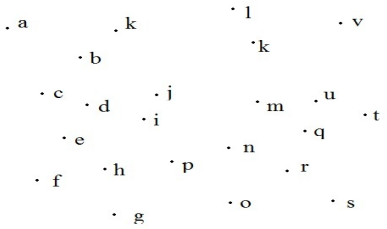


Fig. 1. Arbitrary Wireless Sensor Network

3.3 Fault-Detection Model

In our approach, the node fault is detected by using density-based clustering. In this method, we find the nodes which are in a cluster and having same behavior. By taking a node as a point we find a density based region in which points are density-reachable. We connect the nodes by using the density connectivity technique. If the circle of a node has minimum points (assumed according to the coverage) then it is called as *Core Object* (b, c, d according to Figure 2) [8]. After this, we find the point 'p' which is density-reachable from point 'q', then points to get a density-based cluster. This points form a cluster of same behavior. Series of points are extracted which are density reachable and then we connect these Figure 2 shows the density-based fault-detection model.

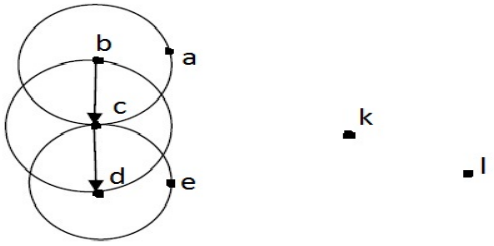


Fig. 2. Fault-Detection Model

4 Proposed Approach

4.1 Fault-Detection Model

In our approach, we have proposed a density-based clustering method in which we find a density-based cluster by joining the nodes by using a technique of density connectivity. In this method, we come to know about the faulty nodes which are not in the cluster. First, find a point and according to the ε-neighborhood of the point we form a circle. 'ε' shows the radius of the circle with respect to the node. If the ε-neighborhood of a node or point contains minimum number of points then this node is called core object [8]. In our algorithm, a set of nodes are given (arbitrary network) 'N', a node 'n' is density-reachable from a point 'm', if 'n' is within the radius (ε-neighborhood) of 'm' and 'm' is a core object [8].

A node n_1 is density-reachable from node n_2 , if there is series of nodes n_1, n_2, …,n_n, where n_1= m and n_n=n such that n_{i+1} is directly density-reachable from n_i (with respect to ε and minimum points). A node is density-connected if there is an node 'o' belongs to 'N' and o is density-reachable to both n and m (with respect to minimum points and ε) [8].

So, like this we find the points or nodes which are of same behavior (according to the assumptions taken) and then come to know which node is faulty and which node is not faulty.

4.2 Proposed Algorithm

The pseudocode for designing a fault-detection model is as follows:

Pseudocode for Fault-Detection model:

1. Given a set of points or nodes 'N'.
2. 'ε' is represented as the radius of the circle which is defined according to the size of the area (coverage).
3. Minimum points in a circle are chosen to control the cluster size.

4. Form the density-reachable points by checking the two conditions:

   (i) the points are close enough to each other:

   $$distance\ (n1,n2) < \varepsilon$$

   (ii) there are enough points in its neighborhood

5. Form density-connected points, to find a density-based cluster.
6. If (the points are in a density-based cluster)
   {
   The points or nodes within the cluster have same behavior (faulty or active).
   }
   else
   {
   Fault is detected, with different behavior (faulty or active) than the points inside the cluster.
   }
7. Fault is detected.

5 Illustration and Observation

5.1 Illustration

We have considered an illustration in which we have an arbitrary wireless network. By using the density-based clustering technique, the fault-detection model is created to detect the faults. Here, we have chosen '3' minimum points in a circle. Figure 3 shows an arbitrary network containing nodes.

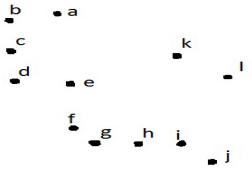


Fig. 3. Arbitrary Wireless Sensor Network

5.2 Observation

In this observation, we have considered four different cases, according to the assumptions. Final fault-detection model is shown in figure 4.

1) Case 1: If all the neighbors of node 'b' detect an event and node 'k' does not detect the event, then there may be a failure or fault in node 'q' according to density-based approach, it is not in the density-based cluster and may have different local decision.

2) *Case 2:* If all the neighbors of node 'g' cannot detect an event and node 'l' detect the event, then there may be a failure or fault in node 'l' according to density-based approach, it is not in the density-based cluster and may have different local decision.
3) *Case 3:* If the node 'j' has a decision, then the nodes (i, h, g, f e.g.) inside the density-based cluster have also the same decision (all have same decision).
4) *Case 4:* If the node inside the cluster shows a faulty local decision and other nodes shows right decision, then there is a fault in that node according to the density-based approach (when there is a fault in the cluster).

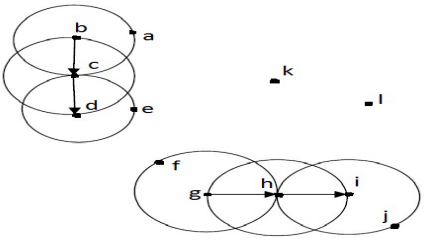


Fig. 4. Final Fault-Detection Model

6 Conclusion

Our algorithm shows better results in detecting the faults and detecting the event. It has a computational complexity of $O(n^2)$, where n is the number of nodes. Event detection is a wide application of wireless sensor network, where fault in the nodes can degrade the performance of the network. So, detection of faults is a necessity task nowadays. Our main idea here is to form a density-based cluster in which the nodes have the same behavior (faulty or active). By this method, the fault in the network is detected as the nodes which are not in the set of density-connected objects. Our observation shows better results in modeling a fault-detection paradigm to detect faults in the network.

In future, implementing this approach in geographical regions is a great area of research. We can add the concepts of accuracy in detecting faults and can add strong fault tolerating techniques to make it a robust fault-tolerant system.

References

1. Applegate, D., Breslau, L., Cohen, E.: Coping with network failures: routing strate-gies for optimal demand oblivious restoration. In: Proc. ACM SIGMETRICS, pp. 270–281 (2004)
2. Brooks, R., Griffin, C., Friedlander, D.: Self-organized distributed sensor network entity tracking. International Journal of High Performance Computing Applications 16(3) (2002)
3. Chamberland, J., Veeravalli, V.: Distributed detection in sensor networks. IEEE on Signal Processing 51(2) (2003)

4. Chen, D., Kintala, D., Garg, S., Trivedi, K.S.: Dependability enhancement for IEEE 802.11 wirelessLAN with redundancy techniques. In: Proceedings of the International Conference on Dependable Systems and Networks, pp. 521–528 (June 2003)
5. Chen, Q., Lam, K., Fan, P.: Comments on distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. IEEE Transactions on Computers 54(9) (September 2005)
6. Feyessa, T., Bikdash, M.: Geographically-sensitive network centrality and survivability assessment. In: 2011 IEEE 43rd Southeastern Symposium on System Theory (SSST), pp. 18–23 (March 2011)
7. Forouzan, B.A.: Data Communication and Networking, 4th edn. Tata Mcgraw-Hill (2006)
8. Han, J., Kamber, M.: Data mining, 2nd edn. Elsevier (2006)
9. Krishnamachari, B., Iyengar, S.: Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. IEEE Transactions on Computers 53(3) (March 2004)
10. Li, R., Wang, X., Jiang, X.: Network survivability against region failure. In: 2011 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), pp. 1–6 (September 2011)
11. Luo, X., Dong, M., Huang, Y.: On distributed fault-tolerant detection in wireless sensor networks. IEEE Transactions on Neural Networks (2005) (to appear)
12. Neumayer, S., Modiano, E.: Network reliability with geographically correlated failures. In: IEEE INFOCOM 2010 (March 2010)
13. Neumayer, S., Zussman, G., Cohen, R., Modiano, E.: Assessing the vulnerability of the fiber infrastructure to disasters. In: IEEE INFOCOM 2009, pp. 1566–1574 (April 2009)
14. Ould-Ahmed-Vall, E., Riley, G.F., Heck, B.S.: A geometric-based approach to fault-tolerance in distributed detection using wireless sensor networks. School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta
15. Ould-Ahmed-Vall, E., Riley, G.F., Heck, B.S.: A distributed fault tolerant algorithm for event detection using heterogeneous wireless sensor networks. In: Proceedings of the 45th IEEE Conference on Decision and Control, CDC 2006 (2006) (under review)
16. Sahoo, M., Khilar, P., Majhi, B.: A redundant neighbourhood approach to tolerate access point failure in IEEE 802.11 WLAN. In: Fourth International Conference on Industrial and Information Systems, ICIIS 2009, pp. 28–31 (December 2009)
17. Sahoo, M.N., Khilar, P.M.: Survivability of IEEE 802.11 wireless LAN against AP failure. International Journal of Computer Applications in Engineering, Technology and Sciences (IJCA- ETS), 424–428 (April 2009)
18. Sen, A., Murthy, S., Banerjee, S.: Region-based connectivity - a new paradigm for design of fault-tolerant networks. In: IEEE International Conference on High Performance Switching and Routing (HPSR), Paris, France, pp. 1–7 (June 2009)
19. Snow, A.P., Varshney, U., Malloy, A.D.: Reliability and survivability of wireless and mobile networks. IEEE Computer, 49–55 (July 2000)

Ant Colony Optimization for Data Cache Technique in MANET

R. Baskaran[1], P. Victer Paul[2], and P. Dhavachelvan[2]

[1] Department of Computer Science and Engineering,
Anna University, Chennai, India
baskaran.ramachandran@gmail.com
[2] Department of Computer Science,
Pondicherry University, Puducherry, India
{victerpaul,dhavachelvan}@gmail.com

Abstract. Mobile ad hoc networks (MANETs) are collection of distributed nodes which communicate using multi-hop wireless links. In order to manage the frequent topology change and ineffective communication, Distributed Spanning Tree (DST) interconnection technique can be used. DST technique guarantee network connectivity, efficient routing and maintain network performance in MANET. In MANET, nodes work for tasks of similar goal (common interest). So, most of the nodes try to access the similar data at different period. By using Data cache system (DCS), we can improve the efficiency of MANET to some extent [1]. In this paper, Ant Colony Optimization (ACO) technique is used to enhance the efficiency of data cache system in MANET at higher level. ACO technique improves the data transfer speed by finding optimal path between nodes of MANET in dynamic fashion. Analysis from simulation of our proposed work shows that data cache system efficiency can be improved using DST technique.

1 Introduction

Mobile ad hoc network (MANET) is an autonomous collection of mobile nodes that communicate over relatively bandwidth constrained wireless links. Mobile ad hoc networks have potential applications in civilian and military environments such as disaster recovery efforts, group conferences, mobile info-stations (in tourist centers, restaurants, and so on), and battlefield maneuvers, making them a focus of current research [2].Thus nodes in MANET probably work for tasks of like goal (common interest). So, most of the nodes try to access the similar data at different period. MANET applications should check for the existence of the desired data inside the network before attempting to connect to the external data source [3]. In MANET, performance of search and retrieval of Cached Data item relies on the efficiency of employed routing strategies [5]. An important problem in a MANET is finding and maintaining efficient routes since host mobility can cause topology changes [6]. The network should be able to adaptively alter the routing paths to alleviate any of these effects [7]. Nityananda Sarma et al [8] proposed that Most of the QoS routing algorithms proposed for Ad Hoc networks are based on on-demand routing [9,10].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 873–878.
springerlink.com

Agent based techniques in which each agent acts like a moderator has been studied by many researchers [14, 21-22]. An efficient DCS, which uses DST as interconnection technique and improve the data search application in MANET is proposed in [1]. This DST structures improves the DCS in MANET to some extend but it may lead bottleneck on overload. In this paper, a popular optimization technique, Ant Colony Optimization (ACO) is used to enhance the efficiency of DCM in MANET by finding optimal path between nodes in dynamic fashion. ACO [11, 12, 13], is a commanding heuristic approach to solve combinatorial optimization problems such as the TSP, Routing in telecommunication networks. So applying ACO approach can enhance the effective routing of message (at low cost) in the MANET which in-tern reduces the number of message pass required for communication to achieve high level efficiency in search applications.

2 Background Information Needed

2.1 Interconnection and Data Cache Technique

Distributed Spanning Tree (DST) [15-17] is the interconnection formation we follow as in [1, 18-20] which, improve the routing and reduce the number of message passes required for any communication in MANET. DST virtually convert the MANET into DST and each tree should have its root node we call it as Head Node (HN) and others are Leaf Node (LN). The details stored in HNs and LNs in the DST is used to enhance the efficient routing with minimum message pass.

A detailed mechanism for Data Cache System (DCS) and effective search and retrieval of cached data using DST as a communication structure in MANET is found in [1], in which the DCS algorithm is given as eight procedures, initialize(G) is the starting procedure of algorithm takes the graph G structure of MANET as argument. These procedures are executed in distributed fashion among the nodes of MANET which engage in data cache system. Thus each node performs the efficient data cache using DST as interconnection technique.

3 Proposed Ant Colony Optimization of DST Modeled MANET

ACO is a probabilistic technique which search for optimal path in a graph, which is based on behaviors of ants seeking a path between their colony and source of food. So, by applying the ACO on the DST, we can obtain optimal path for message pass among the nodes in the MANET. ACO is also capable to reform a new optimal path in case of any problem with the current optimal path which improve the effective routing in real-time. In this paper the Ant Colony Optimization procedure has been modified and proposed for finding an optimal path in DST of the MANET.

The Modified Ant Colony Optimization procedure uses four procedures. Procedure *Optimize()* is the entry point for the algorithm. It takes the MANET, Graph *G*, as its parameter and calls other procedures *broadcast(), structure(), daemonaction()* at some criteria.

Procedure *Optimize()* performs two operations, to find the dynamic optimal path between every HN and between HN and its every LNs in the DST formed MANET.

Operation 1: Let HN_i is a HN among $\{HN_1, HN_2 \ldots HN_n\}$ where 'n' is the number HNs in the Network. HN_i use Probe message *p* to find optimal path between HN_i and other HNs. First procedure *broadcast()* is called which takes Graph G, HN_i as v and probe message p as parameters. Probe *p* is flooded through all the possible paths from HN_i. Then procedure *structure()* is called by HN_i which takes Graph *G*, start HN *v*, specific end HN *x* and the measure concerned with edge between each Peers along the way between *v* and *x*, τ whose value is used to decide the optimal path. Each flooded probe *p*, count the value of τ along its way and save the value in a variable '*val*' and submit the value of τ on its path to *x* to take decision. The equation to find value of '*val*' can be given as,

$$val = \sum_{i=0}^{p} (\tau_i) \tag{1}$$

where,

- '*val*' is a variable to count the value of τ on each edge from v to x.
- '*p*' is the number of edges between HNs v and x in the network
- 'τ' is the value concerned with edge between the Peers

Procedure *DaemonAction(val)* is called by end HN *x*, which takes '*val*' as a parameter and decide the optimal path between the HNs *v* and *x* based on the value of τ along the path of each probe *p*. Every probe reaches *x* with its '*val*' then *x* decides the optimal path based on the '*val*' and the component type of τ. HN *x* inform the identified optimal path to the HN *v*. So, the operation of finding an optimal path between HNs has been successfully done.

Operation 2: In this operation optimal path between each HN and their LNs is identified.

Table 1. Comparison of various criteria measure between DST and ACO optimized DST MANET

S.No		DST	ACO optimized DST
1	No. of Messages created to formulate the technique	152	87
2	Time taken to formulate the technique (in sec)	2.36	1.01
3	No. of Local Read operations performed	126	185
4	No. of Nodes involved in Local Read Operation.	25	27
5	No. of External Read operations performed	22	22
6	No. of Nodes involved in External Read Operation	18	19

As said in operation 1, operation 2 is carried on such that the value of τ counted along the path between HN and all its LNs. Number of optimal paths identified in operation 2 and a best path has been chosen based on the message pass count. Communication through the best (optimal) path found using ACO needs low cost and improve overall efficiency of the Data cache system.

4 Simulation and Analysis

This section illustrates the simulation results obtained during the analysis and we used OMNeT++, which is an object-oriented modular discrete event network simulator. We simulated a mobile ad-hoc network by placing 30 nodes randomly within a region and propagation delay is set as 100ms.

In our simulation, Number of messages created to form DST in MANET is about 176 and the time taken is 2.69 seconds with propagation delay of 100 ms. To route message from source to destination Ant Colony Optimized routing technique is followed which improve the message pass efficiency of DST because of dynamically identified optimal route between every HN and LN Nodes. The various data observed from the simulation in first 50 seconds in DST and the ACO optimized DST are tabulated in Table 1. Thus by using DST as a interconnection structure and ACO as routing optimization technique we are reducing the message pass between the requester and servicing nodes at high level which makes the operation fast and consistent. Fig.1. shows Number of data items served from the cached data items in ACO optimized DST. From Table 2 we can clearly observe that increase in request served, request received and Hit Ratio. This is because the ACO optimization reduces the number of message passes required for an operation by finding best optimal path between the nodes.

On analyzing the various criteria measurements we obtained from simulation we conclude that ACO Optimized DST can perform nearly 50% faster than mere DST MANET. Our simulation is made with channel of propagation delay of 100ms. By decreasing the propagation delay we can improve the throughput of our DCS technique.

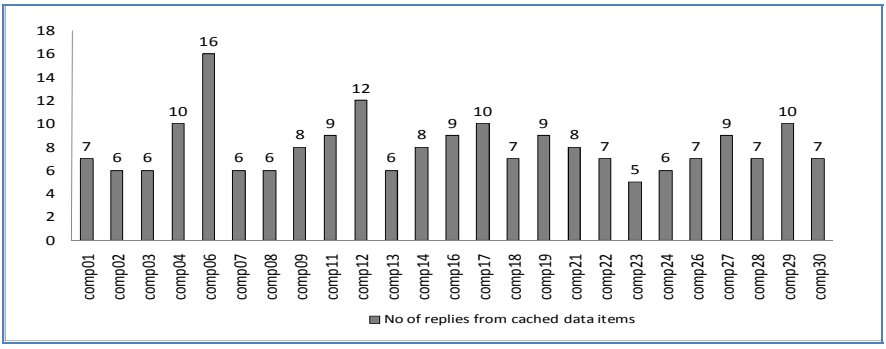


Fig. 1. No. of replies from the cached data items in ACO optimized DST

Table 2. Comparison Table for HNs involved serving the nodes in ACO optimized MANET

S. No	HN	Nearest HNs and distance in Hop(s)				No. of data items cached	No. of data item request served	No. of request received	Hit Ratio
		HN1	Distance	HN2	Distance				
1	comp05	comp10	2	comp25	4	7	39	118	35.3%
2	comp10	comp05	2	comp25	3	12	59	139	44.0%
5	comp25	comp05	4	comp10	3	4	35	91	39.6%

5 Conclusion and Future Work

The work we presented in this paper has described the Ant colony Optimization of Data Cache System (DCS) technique in MANET using DST Model. From the simulation analysis, it is shown that by employing ACO applied DST in MANET, with cost of few message pass we can assure improved data cache, retrieval of cached data and load balancing among the nodes. ACO increases the dynamic routing among the nodes under channel crashes and dynamic HN creation improves the fault tolerant capacity of the overall system. We are working for optimal cache replacement and cache admission control technique to improve the performance the DCS in search application.

References

1. Victer Paul, P., Vengattaraman, T., Dhavachelvan, P., Baskaran, R.: Improved Data Cache Scheme using Distributed Spanning Tree in Mobile Adhoc Network. International Journal of Computer Science and Communication (IJCSC) 1(2) (2010) ISSN: 0973-7391
2. Cao, G., Yin, L., Das, C.R.: Cooperative Cache- Based Data Access in Ad Hoc Networks. Pennsylvania State University. IEEE Computer Society (February 2004)
3. Artail, H., Mershad, K.: MDPF: Minimum Distance Packet Forwarding for search applications in mobile ad hoc networks. IEEE Transactions on Mobile Computing (2009)
4. Boukerche, A.: Algorithms and Protocols for Wireless and Mobile Ad Hoc Networks. Wiley Series on Parallel and Distributed Computing Copyright © 2009. John Wiley & Sons, Inc. (2009)
5. Broch, J., Maltz, D., Johnson, D., Hu, Y., Jetcheva, J.: A performance comparison of multi-hop wireless ad hoc network routing protocols Source. In: Proc. Fourth Annual ACM/IEEE Int'l Conf. on Mobile Computing and Networking, pp. 85–97 (1998)
6. Dahan, S., Philippe, L., Nicod, J.-M.: The Distributed Spanning Tree Structure. IEEE Trans. on Parallel and Distributed Systems 20(12) (December 2009)
7. Mobile Ad Hoc Networks (MANETs), Web site owner: The National Institute of Standards and Technology, `http://w3.antd.nist.gov/wahn_mahn.shtml`
8. Sarma, N., Nandi, S.: Route Stability Based QoS Routing in Mobile Ad Hoc Networks. © Springer Science+Business Media, LLC (2009)
9. Mohapatra, P., Li, J., Gui, C.: QoS in mobile ad hoc networks. IEEE Wireless Communications 10(3), 44–52 (2003)
10. Chakrabarti, S., Mishra, A.: QoS issues in ad hoc wireless networks. IEEE Communication Magazine, 142–148 (2001)
11. Neumann, F., Witt, C.: Runtime Analysis of a Simple Ant Colony Optimization Algorithm. Springer Science+Business Media, LLC (2007)
12. Dorigo, M., Maniezzo, V., Colorni, A.: The ant system: An autocatalytic optimizing Process. Tech.Rep. 91-016 Revised, Politecnico di Milano, Italy (1991)
13. Colorni, A., Dorigo, M., Maniezzo, V.: Distributed optimization by ant colonies. In: Proceedings of European Conference on Artificial Life, ECAL 1991, pp. 134–142. Elsevier Publishing, Amsterdam (1991)
14. Vengattaraman, T., Dhavachelvan, P.: An Agent-Based Personalized E-Learning Environment: Effort Prediction Perspective. In: IEEE International Conference on Intelligent Agent & Multi-Agent Systems, IAMA 2009 (2009) ISBN: 978 1-4 244-4710-7

15. Dahan, S.: Distributed Spanning Tree Algorithms for Large Scale Traversals. In: 11th International Conference on Parallel and Distributed Systems (ICPADS 2005). IEEE (2005)
16. Dahan, S., Nicod, J.-M., Philippe, L.: The Distributed Spanning Tree. IEEE Transactions on Parallel and Distributed Systems 20(12) (December 2009)
17. Rowstron, A., Druschel, P.: Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems. In: Guerraoui, R. (ed.) Middleware 2001. LNCS, vol. 2218, pp. 329–350. Springer, Heidelberg (2001)
18. Victer Paul, P., Vengattaraman, T., Saleem Basha, M.S., Dhavachelvan, P.: Distributed Spanning Trees for Effective Replica Management. In: IEEE International Conference on Advances in Communication, Network, and Computing – CNC 2010, India (2010)
19. Victer Paul, P., Saravanan, N., Jayakumar, S.K.V., Dhavachelvan, P., Baskaran, R.: QoS Enhancements for Global Replication Management in Peer to Peer networks. Future Generation Computer Systems (2011) 28(3), 573–582 (2012)
20. Victer Paul, P., Vengattaraman, T., Dhavachelvan, P., Baskaran, R.: Modeling of Mobile Adhoc Networks Using Distributed Spanning Tree Approach. International Journal of Engineering Science and Technology (IJEST) 2(6), 2241–2247 ISSN: 0975-5462
21. Dhavachelvan, P., Uma, G.V., Venkatachalapathy, V.S.K.: A New Approach in Development of Distributed Framework for Automated Software Testing Using Agents. International Journal on Knowledge –Based Systems 19(4), 235–247 (2006)
22. Venkatesan, S., Dhavachelvan, P., Chellapan, C.: Performance analysis of mobile agent failure recovery in e-service applications. International Journal of Computer Standards and Interfaces 32(1-2), 38–43 (2005) ISSN: 0920-5489

Automatic Extraction of Kannada Complex Predicates from Corpora

S. Parameswarappa and V.N. Narayana

Department of Computer Science & Engineering,
Malnad College of Engineering, Hassan
Affiliated to VTU, Belguam
param.phd@gmail.com, vnnarayana@yahoo.com

Abstract. Complex predicates (CP) are special kind of Multi Word Expressions (MWE), which are extracted with a special emphasis on Compound and Conjunct Verbs. Because of the free word order nature of Kannada, Automatic extractions of Kannada CP are quite challenging. Free word order languages have relatively unrestricted local word groups or phrase structures. This paper proposes a solution for automatic extraction of CP based on Shallow morphology and pattern matching technique. Lexicalization of CP has been done based on shallow morphological information. The proposed system uses Verbalizer for extracting CP. Experiments are conducted and the results obtained have been described. To the best of our knowledge and belief, there is no earlier research result in this direction for Kannada language. Hence, our work acts as a base line for further research and future comparison.

Keywords: Complex Predicates, Compound Verbs, Conjunct Verbs, Verbalizer, Pattern Matching, Kannada Corpora.

1 Introduction

Complex predicates are abound in South Asian languages primarily as compound verbs (verb + verb) combination or conjunct verbs (non-verb + verb) combination. It is equally true for Indian languages in general and Kannada in particular. Not all verb + verb sequence are true compound verbs. Hence, we propose a diagnostic test to extract only true compound verbs from Kannada Corpora. Further the compound verbs are classified into two types. Namely, derivationally (Syntax) constructed compound verbs and compound verbs formed by lexicalization (Lexicon). The non-verb in a conjunct verb may be noun or adjective. If non-verb is a noun then it is noun-verb (NV) compound otherwise if it is an adjective then it is an adjective-verb (AV) compound. Like compound verbs, not all non-verb + verb sequences are true conjunct verbs. The present paper proposes a diagnostic test to extract only true conjunct verbs from Kannada corpora. Identification and extraction of CP is a needful task for building lexical resources (dictionaries, Wordnet) and machine translation system. It is the motivation behind the present work.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 879–884.
springerlink.com

2 Related Work

The linguistic facts of complex predicate formation and the associated semantic roles are examined by Alsina [1]; they discussed general theory of complex predicate. In the present work, the proposed diagnostic tests required to identify the CP in a corpus has been inspired by [5]. The Verbalizer required for extracting the complex predicates has been prepared based on the work of Rajyarama [8]. Automatic extraction of V+V complex predicates from a corpus based on linguistic features was presented by [3]. The form and function of Conjunct verb construction in Hindi was examined by [6]. The issues related to the representation of Telugu complex N+V constructions in Wordnet were discussed by [9]. The mechanism to extract Bangla complex predicates automatically from a corpus was proposed by [4]. A Shallow morphology based complex predicates extraction mechanism for Oriya language was proposed by [2]. The solution to identify the conjunct verbs in Hindi was proposed by [7]. They also show the effect of conjunct verb identification on parsing accuracy.

3 Kannada Complex Predicates

The complex predicates are combination of two lexical items. The first and second lexical items of the complex predicates are called polar and vector respectively. The way how these two lexical items come together and forms a CP is quite interesting to examine. Consider a CP ಪಾಠ ಮಾಡು [paaTa maaDu] 'teach lesson'. In the example, the first constituent ಪಾಠ [paaTa] 'lesson' is a polar and the second constituent ಮಾಡು [maaDu] 'do' is a vector.

3.1 Semantics of Complex Predicates

It seems that, polar dominates the whole meaning of the CP. The meaning of the second vector seems to be de-lexicalized, grammaticalized or bleached. That means it does not retain its original meaning that is attested elsewhere in the language. Consider a CP ಸುಖ ಪಡು [sukha paDu] 'happiness'. Now, the polar ಸುಖ [sukha] 'happy' dominates the whole meaning of the CP. The meaning of the vector ಪಡು [paDu] 'experience' is bleached.

3.2 Morphology of Complex Predicates

The polar occurs in bare stem form. What seems to be morphologically uniform is the function of the vector. It takes the load of all kinds of inflectional markers of sentence. These markers are markers of tense, aspect and agreement morphology. The following examples illustrate the morphology of CP.

ಮದುವೆ ಆಗು [maduve aagu] 'to get married' .

maduve (polar stem) + aagu (vector - verbalizer).

ಊಟ ಮಾಡುವರು [uuTa maaDuvaru] 'they will eat'

uuTa(stem)+maaD(verbalizer)+uvaru(inflection).

3.3 Compound Verbs (CompV)

Compound verbs are composed of two words namely polar and vector, both are verbs. They retain the meaning of polar (V1) and the vector (V2) which semantically bleached adds semantic nuances to the meaning of compound verbs. Therefore compound verbs are considered as lexical variant of their polar (V1) component. On the surface, the constituent verbs enjoy a considerable amount of freedom of moment. Two kinds of compound verbs formations are possible in Kannada. One class of compound verbs consists of a past participle verb and a Verbalizer. These usually have idiomatic senses. The following examples illustrate this class. ಕಂಡು ಹಿಡಿ [kanDu hiDi] 'discover', ಹೇಳಿ ಕೊಡು [heeLi koDu] 'teach', ಕಂಡು ಬರು [kanDu baru] 'appear'. This class of compound verbs is also used extensively in indicating aspectual distinctions. An example of this kind is ತಿಂದು ಬಿಡು [tindu biDu] 'eat up'. Another class of compound verbs consists of a main verb (polar) in an infinitive form followed by a verbalizer. The senses are again idiomatic. The examples for them are ಬರ ಮಾಡು [bara maaDu] 'welcome', ತಿಳಿ ಹೇಳು [tiLi heeLu] 'teach, advice'. These two classes are the diagnostics tests used by the proposed system to extract true compound verbs from a corpus.

3.4 Conjunct Verbs (ConjV)

Conjunct verbs are composed of two words namely polar and vector, where the polar component is a non-verb and the vector component is a verb. The polar can be noun or an adjective. Based on the grammatical category of the polar component, there are two types of conjunct verbs. Namely, Noun-Verb (NV) compounds and Adjective-Verb (AV) compounds. Here also, the vector verb is semantically bleached; it adds semantic nuances to the meaning of conjunct verbs. The conjunct verb retains the meaning of its polar component; it can be noun or adjective. Hence the conjunct verbs are considered as lexical variants of their polar component. The examples for conjunct verbs of NV form are ಮಜಾ ಮಾಡು [majaa maaDu] 'enjoy', ಮೋಸ ಹೋಗು [moosa hoogu] 'be deceived'. An example for conjunct verb of AV form is ಒಳಹೊಕ್ಕು [Olahokku] 'penetrate'. The present system uses bare stem form of the polar with verbalizer as a diagnostic test to extract conjunct verbs.

3.5 Verbalizer

Kannada contains set of lexical verbs, they can be added to other constituents to make verb out of them. These are called Verbalizer. Table 1 lists the potential Verbalizer used in the present work. It includes Verbalizer required to construct both compound and conjunct verbs. During CP construction the Verbalizer acts as a vector (V2).

Table 1. Kannada Verbalizer and their Lexical Meaning (LM)

Verbalizer	LM	Verbalizer	LM	Verbalizer	LM
ಮಾಡು [maaDu]	Make/Do	ಬಿಡು [biDu]	Leave	ಹೊಡೆ [hoDe]	Beat
ಪಡು [paDu]	Experience	ಬೀಳು [biiLu]	Fall	ಹುಟ್ಟು [huTTu]	Being born

4 Methodology

Manual observation of the shallow parsed Kannada sentence shows that the complex predicates contain the lexical pattern {[XXX] (v) [YYY] (v)} where XXX and YYY represent any word. But, the lexical categories of root forms of both XXX and YYY must be verb. Then it is a probable candidate for compound verb. Otherwise, if complex predicates contain the lexical pattern {[XXX] (n/adj) [YYY] (v)} where XXX and YYY represents any word. But, the lexical category of root form of XXX is either noun (n) or adjective (adj) and the lexical category of the root form of YYY is verb (v) then it is a probable candidate for conjunct verb. Using the Verbalizer and by applying the diagnostic tests depicted in section 3, extract the true complex predicates from the probable candidates.

Fig. 1 shows the proposed architecture for Kannada Complex Predicates extractor. The architecture consists of the following modules based on their functionality. They are Sentence Extractor, Kannada Shallow Parser and Complex Predicates extractor.

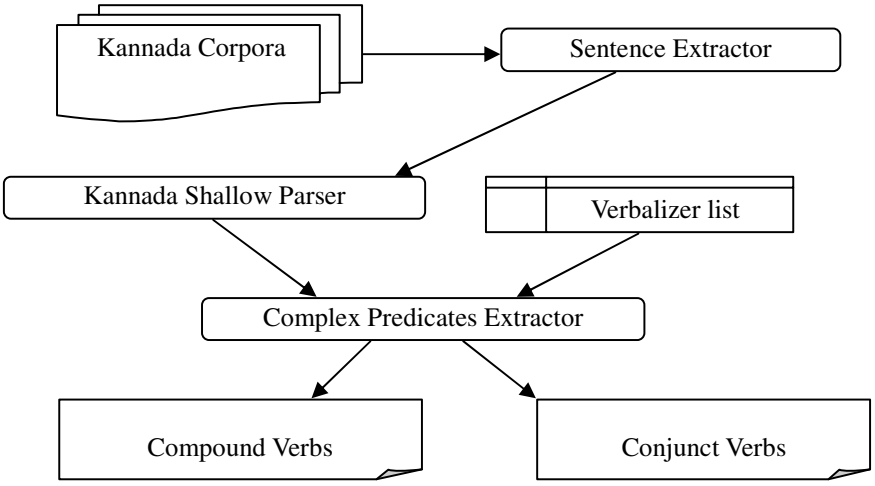


Fig. 1. Proposed system architecture

The input to the proposed system is a Kannada Corpora. The system uses verbalizer list along with Kannada Corpora for extracting the Complex predicates then it produces two output files namely Compound Verbs and Conjunct Verbs they contains extracted compound and conjunct verbs respectively. The functionality of each of these modules is explained briefly as follows.

The input for sentence extractor module is Kannada raw corpora. It extracts randomly selected sentences from corpora required for Complex predicates extraction task.

The morphological information of a given input sentence is obtained by using freely available Kannada Shallow Parser. It parses the given sentence at surface level and produces eight stages of intermediate outputs. The morphological information of

each word in a sentence required for complex predicates extraction task are extracted using this module.

The complex predicates extractor is the core module of the system. The input for the module is shallow parser output. The module scans the parser output for lexical pattern sequence such as verb+verb and (noun/adjective) + verb. If it finds the matched pattern, it will extract the pattern and it will conduct the following diagnostic tests to decide whether extracted pattern is a valid CP or not. For compound verbs, if the polar component (main verb – V1) in a verb(V1)+verb(V2) pattern is in stem form or it may be a past participle or in infinitive form and the light verb (V2) is in verbalizer list (Table 1) then the module decides it is a valid Compound verb and stores it in a Compound Verb file. Likewise for conjunct verbs, if the polar component (main – noun/adj) in a noun/adjective+verb(vector) pattern is in stem form and the vector verb is in verbalizer list (Table 1) then the module decides it is a valid Conjunct verb and stores it in a Conjunct Verb file.

5 Evaluation

The system is tested on randomly selected 500 sentences from both CIIL & web corpora. The extracted complex predicates are in turn given to 10 native speakers of Kannada and asked them to construct sentences using them. If they are able to build correct sentence, then it is a true CP. The precision of the system is calculated as the ratio of the actual CP arrived at through manual validation to the total number of anticipated CP identified by the system. The results of this calculation are shown in Table 2, with a precision rate of 75% for CompV and 87% for ConjV for web corpora. Likewise 73% for (CompV) and 83% (ConjV) for CIIL corpora.

Table 2. System performance

Corpora	CP	System detection	POS ambiguities	Manual validation	Precision
Web Corpora	CompV	75	15	56	0.75
	ConjV	103	21	90	0.87
CIIL Corpora	CompV	86	19	63	0.73
	ConjV	119	33	99	0.83

6 Discussion

The following observations were made during the system evaluation.

Conjunct verbs with loan words as a noun constituent (polar) were extracted.

The selection of the verbalizer is the key factor for extraction process. If we select the potential Verbalizer, then the system will extract complex predicates exhaustively in the corpora.

The loss in precision was caused by POS ambiguities, idiomatic usage and missing data in Kannada shallow parser dictionary.

7 Conclusion and Future Work

In this paper, we presented a study of Kannada Complex Predicates with a special focus on Compound and Conjunct Verbs. The solution for extracting the CP from Kannada corpora is proposed. The diagnostic tests required to validate the CP are specified. Native speaker verified the accuracy of the proposed system. The obtained results are described. The efficiency of the system is proved to be reliable and extendable. The loss in precision in the present system is due to POS ambiguities, Idiomatic usage and missing data in the dictionary. Hence in future, we would like to enhance the efficiency of the system by addressing all the issues encountered in the present system.

References

1. Alsina, A.: Complex Predicates: Structure and Theory. CSLI Publications (1996)
2. Balabantaray, R.C., Jena, M.K., Mohanty, S.: Shallow morphology based complex predicates extraction in Oriya. IJCA (0975 – 8887) 16(1) (2011)
3. Chakrabarti, D., Mandalia, H., Priya, R., Sarma, V., Bhattacharyya, P.: Hindi Compound Verbs and their Automatic Extraction. In: Coling 2008: Companion volume – Posters and Demonstrations, Manchester, pp. 27–30 (2008)
4. Das, D., Pal, S., Mondal, T., Chakraborty, T., Bandyopadhyay, S.: Automatic Extraction of Complex Predicates in Bengali. In: Proceedings of the Multiword Expressions: From Theory to Applications, Beijing (2010)
5. Paul, S.: An HPSG Account of Bangla Compound Verbs with LKB Implementation. Ph.D dissertation,University of Hyderabad, Hyderabad (2004)
6. Das, P.K.: The form and function of Conjunct verb construction in Hindi. Global Association of Indo-ASEAN Studies. Conference, Daejeon, South Korea (2009)
7. Begum, R., Jindal, K., Jain, A., Husain, S., Misra Sharma, D.: Identification of Conjunct Verbs in Hindi and Its Effect on Parsing Accuracy. In: Gelbukh, A.F. (ed.) CICLing 2011, Part I. LNCS, vol. 6608, pp. 29–40. Springer, Heidelberg (2011)
8. Rajyarama: A Study On Some Aspects Of Derivational Morphology In Telugu With Special Reference To Compounds. Ph.d. Thesis, University of Hyderabad (1998)
9. Uma Maheshwar Rao, G., Rajyarama, K.: Representation of Complex Predicates in Wordnet. In: Proceeding of the 5th Global Wordnet Conference, IIT Bombay, India (2010)

Texture Image Retrieval Using Greedy Method

Pushpa B. Patil[1] and Manesh B. Kokare[2]

[1] BLDEA's Dr. P.G.H. CET, Bijapur, Karnataka, India
pushpa_ms@rediffmail.com
[2] SGGS Institute of Tech. & Engg., Nanded, Maharashtra, India
mbkokare@yahoo.com

Abstract. There is a huge amount of methods for extracting image descriptors and defining the similarity measures. In this paper, we try to improve texture image retrieval performance with post processing based on the greedy technique called Prims algorithm. In the proposed method feature database is represented using distance matrix, which is the distance between every image of the database. Due to symmetric property of a matrix, we can improve the efficiency and effectiveness of the proposed retrieval system. However for large database the size of the matrix is large. The proposed system is tested with three different image descriptors, namely combined rotated complex wavelet filters (RCWF) and dual tree complex wavelets (DT-CWT), Contourlet Transform (CT), and Discrete Wavelet Transforms (DWT).

1 Introduction

Due to drastic growth of multimedia and digital technology in recent years, there is a need of effective and efficient management of digital image libraries and other multimedia databases. Hence, storage and retrieval of images in such libraries become a real demand in industrial, medical, crime prevention, biometric systems, and other applications. Content-Based Image Indexing and Retrieval (CBIR) is considered as a solution. In such systems, important features are extracted from every picture and stored as a feature vector. Content-based image retrieval has attracted substantial interests in the last decade [1,5,8,9,10]. Generally image retrieval system can be divided in to two different steps. First image descriptors have to be extracted by analyzing color, texture, shape or context. Second a similarity measure analyzing variations in images features has to be defined**.** Then given a query image, all other images of the database are sorted based on their similarity rank to the query image. Finally, high ranking images are returned to the users.

There are several methods available to perform these two steps in Content-based image retrieval. Recently some effort was also put on post-processing by using the obtained similarities between all given images [11]. In this paper, for post processing step we used greedy algorithm called Prims. Using this algorithm we can find minimum cost spanning tree.

1.1 Related Work

In 2004, Zhou et al. [12,13,14] proposed a novel semi-supervised learning algorithm named manifold-ranking. In this algorithm, a connected graph is created first, in

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 885–891.
springerlink.com

which vertex represents the data and edge represents the similarity between vertices, and then the score diffuses from the vertices to their neighbors. After several rounds, all the vertices get stable scores. Finally, all the data point are ranked by the scores of the corresponding vertices in the graph. The assumption of the algorithm is that all the data points are distributed in a low dimension manifold, which is embedded in the high dimension features space. Compared to the pairwise method, the main difference of the manifold ranking based method is, it ranks all the data according to the manifold structures represented by the labeled and unlabeled data, which means that the algorithm ranks the data by considering local and global consistency simultaneously. In [3,4], a pair wise graph based manifold ranking algorithm [12] is adopted to build image retrieval. These graph based methods motivates us to do work on the image retrieval using greedy algorithm.

The main contribution of this paper is summarized as, we have proposed novel texture image retrieval using greedy method called Prims algorithm. The experimental results of proposed method perform better compared with earlier approach. The rest of paper is organized as follows. In section 2, we discuss the image descriptors in brief. In section 3, we discuss the proposed greedy algorithm for image retrieval. In section 4, the experimental results are given and finally section 5 concludes the work.

2 Image Descriptors

For image feature extraction we have used three different methods, namely combined rotated complex wavelet filters (RCWF) and dual tree complex wavelets (DT-CWT), Contourlet Transform (CT), and Discrete Wavelet Transforms (DWT). DT-CWT, DT-RCWF, and CT are explained below in short.

2.1 DT-CWT

Real DWT has poor directional selectivity and it lacks shift invariance. Drawbacks of the DWT are overcome by the complex wavelet transform (CWT) by introducing limited redundancy into the transform. But still it suffer from problem like no perfect reconstruction is possible using CWT decomposition beyond level 1, when input to each level becomes complex. To overcome this, Kingsbury [7] proposed a new transform, which provides perfect reconstruction along with providing the other advantages of complex wavelet, which is DT-CWT. The DT-CWT uses a dual tree of real part of wavelet transform instead using complex coefficients. This introduces a limited amount of redundancy and provides perfect reconstruction along with providing the other advantages of complex wavelets. The DT-CWT is implemented using separable transforms and by combining subband signals appropriately. Even though it is non-separable yet it inherits the computational efficiency of separable transforms. A complex valued $\psi(t)$ can be obtained as

$$\psi(x) = \psi_h(x) + j\,\psi_g(x) \tag{1}$$

where $\psi_h(x)$ and $\psi_g(x)$ are both real valued wavelets.

2.2 DT-RCWF

Recently, Kokare et.al.[6] have designed 2D- rotated complex wavelet transform. Directional 2D RCWF are obtained by rotating the 2D DT-CWT filters by 45^0 so that decomposition is performed along new direction, which is 45^0apart from decomposition of DT-CWT. The size of a newly obtained filter is $(2N-1)X(2N-1)$, where N is the length of the 1-D filter. The decomposition of input image with 2-D DT-RCWF followed by 2-D downsampling operation is performed up to the desired level. The computational complexity associated with DT-RCWF decomposition is the same as that of standard 2-D DT-CWT, if both are implemented in the frequency domain. The set of DT-RCWFs retains the orthogonality property. The six subbands of 2D DT-RCWF gives information strongly oriented at $(30^\circ, 0^\circ, -30^\circ, 60^\circ, 90^\circ, 120^\circ)$.The mechanism of the DT-RCWF is explained in detail in [6]. The 2D DT-CWT and RCWF provide us more directional selectivity in the directions $\left\{\left(+15^\circ,\ +45^\circ,\ +75^\circ,\ -15^\circ,\ -45^\circ, -75^\circ\right), \left(0^\circ, +30^\circ,\ +60^\circ,\ +90^\circ,\ 120^\circ,\ -30^\circ\right)\right\}$.

2.3 Contourlet Transform

Multiscale and time frequency localization of an image is offered by wavelets. But, wavelets are not effective in representing the images with smooth contours in different directions. The Contourlet provides a much richer set of directions and shapes. Hence they are more effective in capturing smooth contours and geometric structures in images [2]. Contourlet transform is a multiscale and directional image representation that uses first a wavelet like structure for edge detection, and then a local directional transform for contour segment detection.

3 Proposed Greedy Algorithm

In this paper, we proposed the new content-based image retrieval using Prims algorithm. Prim's algorithm is a greedy algorithm that finds a minimum cost spanning tree for a connected weighted undirected graph. So we represented the feature database in the form of distance matrix. Due to symmetric matrix, we considered only upper main diagonal elements in order to retrieve the images. The size of matrix is large for large image database. The proposed method is tested using the different image descriptors namely combined dual tree rotated complex wavelet filters(DT-RCWF) and dual tree complex wavelet transform(DT-CWT)[6], Contourlet Transform(CT)[2], and Discrete Wavelet Transform(DWT)[6] separately.

Let $G = (V, E)$ be connected weighted undirected graph, V is the set of vertices represents images and E is the set of edges, which represent the similarity between images Cabrera distance d_{ij} between the image i and j represents the weight of each edge. Fig. 2 shows sample example for the connected weighted undirected graph, and its minimum cost spanning tree.

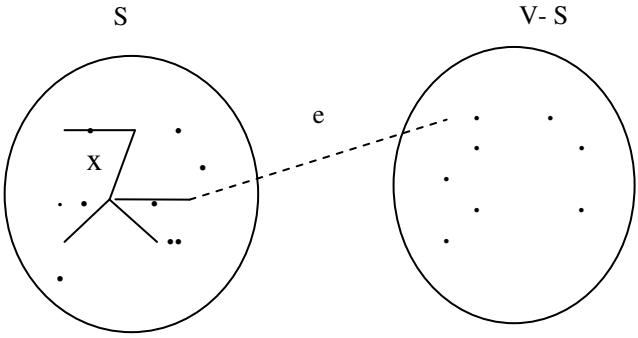
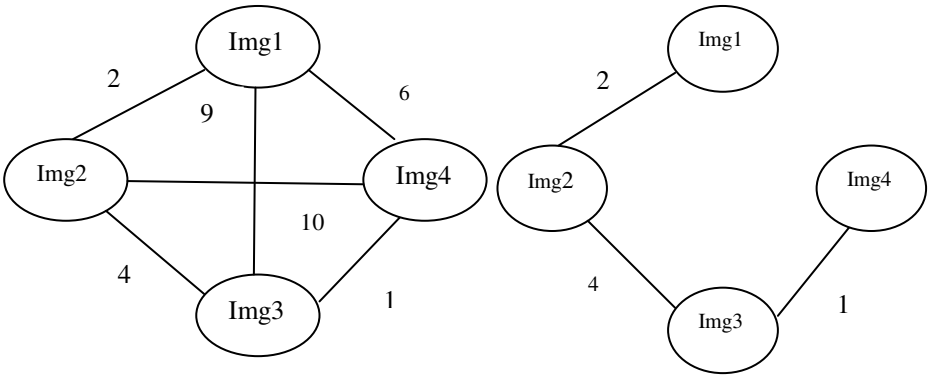


Fig. 1. Prim's algorithm: the edges X form a tree, and S consists of its vertices



a) Sample graph example **b) its minimum cost spanning tree**

Fig. 2. Sample graph and its spanning tree

Fig 1 shows, the algorithm continuously increases the size of a tree, one edge at a time, starting with a tree consisting of a single vertex (query image q), until it spans all vertices (all images). An algorithm 1 describes the proposed method.

Algorithm 1: Greedy Algorithm for Image Retrieval

Input: Image Database DB, Query image q, distance matrix

Output: Retrieved Images

Begin

1 $V_T = \{q\}$

2 $E_T = \emptyset$

3 $i = 1$

4 While $t \leq |V| - 1$ do
5 Begin
6 Find a minimum distance edge $e^* = (v^*, u^*)$ among all the edges
7 Such that v is in V_T and u is in $V - V_T$
8 $V_T = V_T \cup \{u^*\}$
9 $E_T = E_T \cup e^*$
10 End
11 Sort the edges in E_T
12 Display first Top N images
End

A query is considered as the starting vertex in the Prims algortim(line 1). In every iteration algorithm finds minimum distance edge $e^* = (v^*, u^*)$ among all the edges such that v is in V_T and u is in $V - V_T$ (line 6 and 7) and then that vertex(image) u^* and edge e^* is added to the minimum cost spanning tree vertices set V_T and set of edges E_T respectively(line 8 and 9). This procedure is repeated by number of images (vertices) minus one time(line 4 to line 10) . Finally tree edges are sorted and top most N similar images displayed (line 11 and 12). Fig 2 shows the formation of minimum cost tree.

A simple implementation using an adjacency matrix graph representation and searching an array of weights to find the minimum weight edge to add requires $O(|V^2|)$running time. Using a simple binary heap data structure and an adjacency list representation, Prim's algorithm can be shown to run in time $O(E \log V)$, where $|E|$ is the number of edges and $|V|$ is the number of vertices.

4 Experimental Results

To test the efficiency of proposed graph based CBIR, we employed the Brodatz texture database [6]. It consists of 116 different textures. We used 108 textures from Brodatz texture photographic album, seven textures from USC database and one artificial texture. Size of each texture image is 512×512. Each 512×512 image is divided into sixteen 128×128 non overlapping subimages, thus creating a database of 1856 texture images. We used combined dual tree rotated complex wavelet filters (DT-RCWF) and dual tree complex wavelet transform (DT-CWT)[6], Contourlet Transform(CT)[2], and Discrete Wavelet Transform(DWT)[11] to extract image features separately.

For each experiment, one image was selected at random as the query image from each category and thus retrieved images were obtained. For performance evaluation of the image retrieval system, it is significant to define a suitable metric. We employed accuracy, which is defined as follows

$$Accuracy = \frac{Number\ of\ relevant\ images\ retrieved}{Number\ of\ relevant\ images\ in\ database} \quad (2)$$

The comparative retrieval performance of the proposed system is shown in Table 1. Fig. 3 shows the comparison results for image retrieval using different image descriptors and improvement of performance with proposed method.

Table 1. Percentage Average Retrieval Accuracy for Brodatz texture Database

Image descriptors	%Average retrieval accuracy of earlier methods	%Average retrieval accuracy of proposed method
DWT[11]	69.61	72.31
CT[12]	76.13	78.61
DT-CWT+DT-RCWF[11]	78.5	81.61

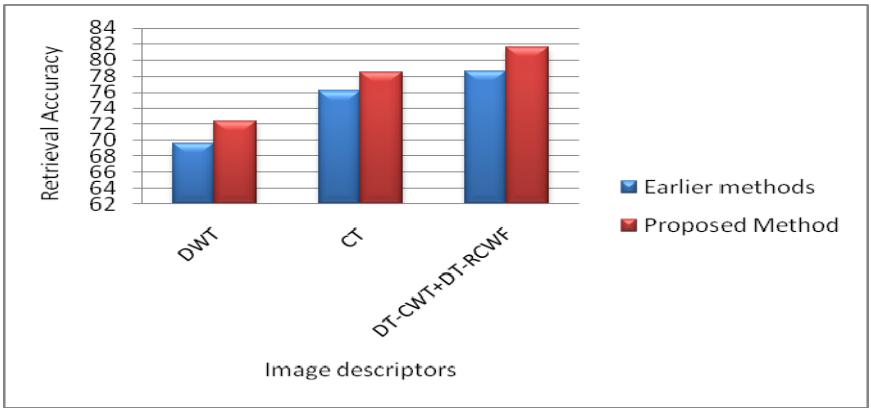


Fig. 3. Retrieval performance

5 Conclusions

In this paper, we have introduced a novel Content-Based Image Retrieval framework based on greedy technique, which uses the connected undirected weighted graph structure. It is used to represent the relationship among the vertices (images). We have tested proposed system using three different texture features. Experimental results indicate that the proposed method using combined DT-CWT and DT-RCWF features retrieval rate increases from 78.5% to 81.61%, from 76.13% to 78.61% using CT features and 69.61% to 72.31% on texture database. Hence experimental results of the proposed method are satisfactory compared to the existing methods.

References

1. Ritendra, D., Dhiraj, J., Jia, L., Wang, Z.J.: Image Retrieval: Ideas, Influ-ences, and Trends of the New Age. ACM Computing Surveys 40(2), article 5, 5:1–5:60 (2008)
2. Duncan, D.Y., Minh, N.D.: Directional Multiscale Modeling of Image using the Contourlet Transform. IEEE Transactions on Image Processing 15(6), 1610–1620 (2006)
3. He, J., Li, M., Zhang, H.J., Tong, H., Zhang, C.: Generalized Manifold-Ranking Based Image Retrieval. IEEE Transactions on Image Processing 15(10), 3170–3177 (2006)
4. He, J., Li, M., Zhang, H.J., Tong, H., Zhang, C.: Manifold-Ranking Based Image Retrieval. In: ACM Multimedia (2004)
5. Kokare, M., Chatterji, B.N., Biswas, P.K.: A Survey on Current Content-based Image Retrieval Methods. IETE J. Res. 48(3&4), 261–271 (2002)
6. Kokare, M., Chatterji, B.N., Biswas, P.K.: Texture Image Retrieval using New Rotated Complex Wavelet Filters. IEEE Trans. on Systems, Man, and Cybernetics-Part B: Cybernetics 35(6), 1168–1178 (2005)
7. Kingsbury, N.G.: Image processing with complex wavelet. Phil. Trans. Roy. Soc. London A 357, 2543–2560 (1999)
8. Liua, Y., Zhang, D., Lua, G., Mab, W.Y.: A Survey of Content-Based Image Retrieval with high-level semantics. In: Proceedings of the Pattern Recognition, pp. 262–282 (2007)
9. Rui, Y., Hung, T.S., Chang, S.F.: Image retrieval: Current Techniques, Promising Directions and Open Issues. J. Visual Comm. and Image Representation 10, 39–62 (1999)
10. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content–Based Image Retrieval at the End of the Early Years. IEEE Trans. Pattern Anal. Machine Intell. 22(12), 1349–1380 (2000)
11. Yang, X., Bai, X., Latecki, L.J., Tu, Z.: Improving Shape Retrieval by Learning Graph Transduction. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part IV. LNCS, vol. 5305, pp. 788–801. Springer, Heidelberg (2008)
12. Zhou, D., Bousquer, O., Lal, T., Weston, J., Schölkopf, B.: Learning with Local and Global Consistency. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 321–328 (2004)
13. Zhou, D., Weston, J., Agretton: Ranking on Data Manifold. In: Proceedings of the Advances in Neural Information Processing Systems, pp. 169–176 (2004)
14. Zhou, D., Schölkopf, B.: Learning from Labeled and Unlabeled Data Using Random Walks. In: Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B., Giese, M.A. (eds.) DAGM 2004. LNCS, vol. 3175, pp. 237–244. Springer, Heidelberg (2004)

Multi-lingual Speaker Identification with the Constraint of Limited Data

B.G. Nagaraja and H.S. Jayanna

Department of Information Science and Engineering,
Siddaganga Institute of Technology, Tumkur - 03
{nagarajbg,jayannahs}@gmail.com

Abstract. State-of-the-art Speaker recognition system uses Gaussian Mixture Model-Universal Background Model (GMM-UBM) as a modeling technique. This work describes a closed-set text-independent speaker identification system using GMM-UBM in the context of Mono, Cross and Multi-lingual with the constraint of limited data. To study the significance of GMM-UBM, experiments are conducted on three different languages (English, Hindi and Kannada) using an evaluation set of 30 speakers. Experiments are conducted with Not Including Evaluation set (NIE) and Including Evaluation set (IE) in UBM training. The results show that the GMM-UBM-IE, in comparison with GMM-UBM-NIE yields improved identification performance in all the speaker identification experiments.

Keywords: Speaker identification (SI), MFCC, GMM-UBM-IE, GMM-UBM-NIE.

1 Introduction

Speaker recognition aims at recognizing the Speakers from their voice [1]. Depending on the mode of operation, Speaker identification (SI) system can be either text-dependent (constraint on what is spoken) or text-independent (no constraint on what is spoken) [2]. Reynolds *et al.* [3, 4] presented GMM and GMM-UBM modeling techniques for a text-independent speaker identification/Verification system. The GMM based speaker recognition systems utilize a universal background model (UBM), which requires extensive resources [5]. In Limited data speaker recognition systems, speech is pooled from many speakers to train a single independent model, known as UBM [6]. Individual speakers are then adapted from the UBM using the maximum a posterior (MAP) adaptation algorithm [6].

In Mono-lingual SI, training and testing languages for a speaker are the same whereas in Cross-lingual SI, training is done in one language (say A) and testing is done in another language (say B). In Multi-lingual SI, some speakers in database are trained and tested in language A, some speakers are in language B and so on [7] [8], i.e., Speaker Models are trained in one language and tested with multiple languages of different speakers. Most of state-of-the-art SI systems work on Mono-lingual (preferably English) using sufficient data. The use of SI system in Multi-lingual context is a requirement in a country like India where there is a coexistence of large number of

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 893–898.
springerlink.com

languages. Sufficient data refers to the case of having few minutes (> 1 min) of speech data and Limited data refers to the case of having few seconds (≤ 15 sec) of speech data [8].

A novel Multi-lingual text-independent based speaker identification algorithm was proposed by Geoffrey Duron in [9] and investigated 2 facets of speaker recognition: cross-language speaker identification and the same language non-native text independent SI. The results indicated that how SI performance will be affected when speakers do not use the same language during the training and testing or when the population is composed of native speakers.

In our previous work [8], we have made an attempt to identify speaker in the context of Mono and Cross-lingual with the constraint of limited data using Mel-Frequency Cepstral Coefficients (MFCC) as feature vectors and Vector Quantization (VQ) as modeling technique. We observed that SI system with English language provides good performance in Mono-lingual study. Further, it was observed in Cross-lingual study that the use of English language either in training or testing gives better identification performance.

The paper is organized as follows: Section 2 describes the database used for the experiments. Feature extraction using MFCC and speaker modeling using GMM-UBM technique are presented in Section 3. Section 4 gives experimental results. Finally, Summary and conclusions of this study and scope for the future work are mentioned in Section 5.

2 Speech Database for the Study

Since the standard Multi-lingual database is not available, experiments are carried out on an our own created database of 30 speakers who can speak the three different languages. The database includes 17-males and 13-females speakers. The voice recording was done in the Engineering college laboratory. The speakers were undergraduate students and faculties in an engineering college. The age of the speakers varied from 18-35 years. The speakers were asked to read small stories in three different languages. The training and testing data were recorded in different sessions with a minimum gap of two days. The approximate training and testing data length is two minutes. Recording was done using free downloadable wave surfer 1.8.8p3 software and Beetel Head phone-250 with a frequency range 20-20 kHz. The speech files are stored in **.**wav format.

3 Feature Extraction and Modeling

The purpose of feature extraction stage is to extract the speaker-specific information in the form of feature vectors at reduced data rate [1]. In this work, features are extracted using MFCC technique. Speech recordings were sampled at the rate of 8 kHz and pre-emphasized (factor 0.97). Frame duration of 20 msec (160 samples) and a 10 msec (80 samples) of overlapping durations are considered. After framing, windowing (Hamming) method is carried out to minimize the spectral distortion. 35 triangular

band pass filters are considered. These filters are equally spaced along the Mel-frequency scale. First 13 coefficients are considered as feature vectors.

The GMM uses multi-modal Gaussian distribution to represent the speaker's voice and vocal tract configurations [10]. Recently, the GMM employing a UBM with MAP Speaker adaptation has become the dominant approach in text-independent SI [4]. The expectation maximization (EM) algorithm was used to estimate the parameters (mean vectors, covariance matrices and mixture weights) of the GMM models. The k-means algorithm was used to obtain the initial estimate for each cluster [13]. In GMM-UBM system, speech data collected from large number of speakers is pooled and the UBM is trained which acts as a speaker independent model. The speaker dependent model (GMM) can be created by performing MAP adaptation technique from the UBM using speaker-specific training speech.

The UBM training can be done in two ways [12]: 1) Speech data pooled from the other database, not used for the speaker recognition study, provided speech data is collected from the same environment known as Not Including Evaluation set (NIE). 2) Same speech data for both UBM training and evaluation, provided the speakers set used for recognition is not included in UBM training known as Including Evaluation set (IE). In [11] and [12], it was mentioned that there are no criteria to select number of speakers and amount of data to train the UBM. We trained UBM with roughly one hour of data.

4 Experiments

For NIE experiments UBM is trained using the first 30 speakers of YOHO [11] database of approximately one hour of speech data and for IE experiments UBM is trained using the 30 speakers of our own database of approximately one hour of speech data. The Mono-lingual experimental results for the 30 speakers of our own database for 15 sec of training and testing data and for different Gaussian mixtures are given in Table 1. Note: A/B indicates training with language A and testing with language B.

The SI system trained and tested with English language (E/E) gives the highest performance of 83.33% and 93.33% with 256 Gaussian Mixtures for NIE and IE respectively. The highest performance may be due to the Speaker's considered for the study. The SI system trained and tested with Hindi language (H/H) gives the highest performance of 80.00% and 93.33% with 128 and 256 Gaussian Mixtures for NIE and IE respectively. This performance is better than Kannada language. This is because almost all the speakers had taken additional time to practice the Hindi Story which was given to read out in the different sessions and thus their fluency was significantly improved. The performance of SI system trained and tested with Kannada language (K/K) is 73.33% and 86.66% with 256 and 128 Gaussian Mixtures for NIE and IE respectively. The poor performance may be due to the speaker's difficulty in reading Kannada language since they had just studied this language as one of the languages subject in school days.

Table 1. Mono-lingual Speaker identification performance (%). P_i represents the maximum identification performance among the number of Gaussian Mixtures.

Train/Test language	Modeling Technique	Gaussian Mixtures					
		16	32	64	128	256	P_i
E/E	GMM-UBM-NIE	40.00	43.33	63.33	70.00	83.33	83.33
	GMM-UBM-IE	76.66	76.66	83.33	86.66	93.33	93.33
H/H	GMM-UBM-NIE	53.33	56.66	63.33	80.00	76.66	80.00
	GMM-UBM-IE	76.66	86.66	90.00	90.00	93.33	93.33
K/K	GMM-UBM-NIE	50.00	53.33	63.33	63.33	73.33	73.33
	GMM-UBM-IE	80.00	83.33	83.33	86.66	83.33	86.66

The Cross-lingual experimental results for the 30 speakers of our own database for 15 sec of training and testing data and for different Gaussian mixtures are given in Table 2. The SI system trained with Hindi and tested with English language (H/E) yields a highest performance of 73.33% and 90.00% with 256 Gaussian Mixtures for NIE and IE respectively. The performance of SI system trained with Kannada language and tested with English language (K/E) is 73.33% and 86.66% with 256 Gaussian Mixtures for NIE and IE respectively. With English as a testing language, no much difference in identification performance was observed in comparison with Hindi and Kannada as training languages.

The SI system trained with English language and tested with Hindi language (E/H) yields a highest performance of 63.33% and 80.00% with Gaussian Mixtures 256 and 128 for NIE and IE respectively. The performance of SI system trained with Kannada language and tested with Hindi language (K/H) is 66.66% and 80.00% with 256 Gaussian Mixtures for NIE and IE respectively. The SI system trained with English language and tested with Kannada language (E/K) yields identification performance of 63.33% and 83.33% with 256 Gaussian Mixtures for NIE and IE respectively. The performance of SI system trained with Hindi language and tested with Kannada language (H/K) is 56.66% and 80.00% with 256 Gaussian Mixtures for NIE and IE respectively. In Comparison with the Mono-lingual SI, Cross-lingual SI performance decreases drastically. This may be due to the variation in fluency and word stress when same speaker speaks different languages and due to different phonetic and prosodic patterns of the languages [13].

The Multi-lingual experimental results for the 30 speakers of our own database for 15 sec of training and testing data and for different Gaussian mixtures are given in Table 3. The Multi-lingual SI system yields a highest performance of 83.33% and 96.66% with 256 Gaussian Mixtures for NIE and IE respectively. The Multi-lingual results are better than the Mono-lingual and Cross-lingual experiments. This may be due to the better discrimination between the trained and testing models (multiple languages) in Multi-lingual scenario. The high performance of GMM-UBM-IE modeling technique in all the experiments may be due to including the Evaluation set in building the UBM. Hence there is bias in the UBM towards each of the speakers [12].

Table 2. Cross-lingual Speaker identification performance (%). Pi represents the maximum identification performance among the number of Gaussian Mixtures.

Train/Test language	Modeling Technique	Gaussian Mixtures					
		16	32	64	128	256	P_i
H/E	GMM-UBM-NIE	36.66	50.00	56.66	56.66	73.33	73.33
	GMM-UBM-IE	73.33	80.00	83.33	86.66	90.00	90.00
K/E	GMM-UBM-NIE	30.00	46.66	53.33	63.33	73.33	73.33
	GMM-UBM-IE	63.33	76.66	80.00	80.00	86.66	86.66
E/H	GMM-UBM-NIE	40.00	40.00	56.66	56.66	63.33	63.33
	GMM-UBM-IE	66.66	66.66	76.66	80.00	76.66	80.00
K/H	GMM-UBM-NIE	26.66	50.00	50.00	60.00	66.66	66.66
	GMM-UBM-IE	66.66	66.66	76.66	76.66	80.00	80.00
E/K	GMM-UBM-NIE	36.66	43.33	63.33	56.66	63.33	63.33
	GMM-UBM-IE	60.00	70.00	70.00	66.66	83.33	83.33
H/K	GMM-UBM-NIE	26.66	53.33	46.66	53.33	56.66	56.66
	GMM-UBM-IE	63.33	73.33	73.33	76.66	80.00	80.00

Table 3. Multi-lingual Speaker identification performance (%). Pi represents the maximum identification performance among the number of Gaussian Mixtures.

Modeling Technique	Gaussian Mixtures					
	16	32	64	128	256	P_i
GMM-UBM-NIE	40.00	60.00	70.00	76.66	83.33	83.33
GMM-UBM-IE	86.66	90.00	93.33	93.33	96.66	96.66

5 Conclusions

In this paper we have compared the performance of GMM-UBM-NIE and GMM-UBM-IE for Mono, Cross and Multi-lingual SI with the constraint of limited data. The speaker independent UBM was trained by Including Evaluation set (IE) and Not Including Evaluation set (NIE). The speaker dependent model was built by MAP adaptation. The results indicate that GMM-UBM can be used for Speaker identification with the Constraint of limited data. We also pointed out and partly justified the reason for degradation of performance in Cross-lingual SI. In order to study the robustness of the system, needs to be verified with different languages (more than 3), different data sizes and more number of speakers.

Acknowledgments. This work is supported by Visvesvraya Technological University, Belgaum-590018, Karnataka, India.

References

1. Jayanna, H.S., Mahadeva Prasanna, S.R.: Analysis, Feature Extraction, Modeling and Testing techniques for Speaker Recognition. IETE Technical Review 26, 181–190 (2009)
2. Lawrence Rabiner, R., Ronald Schafer, W.: Digital Processing of Speech Signals, 1st edn. Prentice Hall (1978)
3. Reynolds, D.A., Rose, R.C.: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. IEEE Transactions on Speech and Audio Processing 3, 72–83 (1995)
4. Reynolds, D.A., Quatieri, T.F., Dunn, R.B.: Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing 10, 19–41 (2010)
5. Hasan, T., John Hansen, H.L.: A study on Universal Background Model training in Speaker Verification. Proc. IEEE 14(1), 277–288 (2010)
6. Ku, J.M.K., Ambikairajan, E., Epps, J., Togneri, R.: Speaker Verification using Sparse Representation Classification. In: Proc. IEEE ICASSP, pp. 4548–4551 (2011)
7. Arjun, P.H.: Speaker Recognition in Indian Languages: A Feature Based Approach. Indian Institute of Technology Kharagpur, India (July 2005)
8. Nagaraja, B.G., Jayanna, H.S.: Mono and Cross lingual Speaker Identification with the Constraint of Limited Data. In: Proc. IEEE, PRIME-2012, pp. 457–461 (March 2012)
9. Durou, G.: Multilingual text independent Speaker Identification, pp. 115–118
10. Chen, C.-C.T., Chen, C.-T., Hou, C.-K.: Speaker Identification using hybrid Karhunen-Loeve transform and GMM approach, pp. 1073–1075. Elsevier (2004)
11. Jayanna, H.S.: Limited data Speaker Recognition. Indian Institute of Technology, Guwahati, India (November 2009)
12. Reynolds, D.: Universal Background Models
13. Cucchiarini, C., Strik, H., Boves, L.: Evaluation of Dutch pronunciation by using Speech Recognition technology. In: Proc. IEEE ASRU, Santa Barbara (December 1997)

Improving Performance of K-Means Clustering by Initializing Cluster Centers Using Genetic Algorithm and Entropy Based Fuzzy Clustering for Categorization of Diabetic Patients

Asha Gowda Karegowda, Vidya T., Shama, M.A. Jayaram, and A.S. Manjunath

Department of Master of Computer Applications,
Siddaganga Institute of Technology, Tumkur - 03
{ashagksit,vidu.tr,shama.ammu,jayaramdps,asmanju}@gmail.com

Abstract. Medical Data mining is the process of extracting hidden patterns from medical data. Among the various clustering algorithms, k-means is the one of most widely used clustering technique. The performance of k-means clustering depends on the initial cluster centers and might converge to local optimum. K-Means does not guarantee unique clustering because it generates different results with randomly chosen initial clusters for different runs of k-means. This paper investigates the use of two methods namely Genetic Algorithm (GA) and Entropy based fuzzy clustering (EFC) to assign k-means initial cluster centers for clustering PIMA Indian diabetic dataset. Experimental results show markable improvement of 3.06% reduction in the classification error and execution time of k-means clustering initialized by GA and EFC when compared to k-means clustering with random cluster centers.

Keywords: k-means clustering, cluster center initialization, Genetic algorithm, Entropy based fuzzy clustering, Pima Indian Diabetics.

1 Introduction

The data mining functionalities mainly include association rule mining, classification, prediction & clustering. Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters [1]. One of most common clustering method is a k-means clustering. The performance of k-means clustering mainly depends on the initial cluster centers and might converge to a local optimum. Several methods proposed have been proposed to solve the cluster initialization for k-means algorithm. Bradley and Fayyad [2] proposed the method which forms a set of small random sub-samples of the data, and then apply k-means to each of sub-samples. The centroids of each sub-samples is given a initial center for k-means for all centroids of all subsamples. The centers of the final clusters that give minimum clustering error are to be used as the initial centers for clustering the original set of data using k-means algorithm.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 899–904.
springerlink.com

Bashar Al-Shbool [3] have used GA to initialize the k-means cluster centers. Jimenez [4] have used GA to determine the best initial cluster centers and find the number of clusters using binary encoding. Hao-Jun sun [5] has used GA for feature subspace selection of clustering. They have used binary encoding to represent feature subspace and cluster centers. Mohammad F [6] has used statistical information from the data set to initialize the k-means prototypes. Murat Erisoglu [7] has used two principal variables based on maximum coefficient of variation and minimum absolute value of the correlation. The reduced dataset is partitioned one at a time till the desired number of clusters is obtained. The cluster membership for each point is determined according to candidate initial cluster centers and selected two axis. Vidyut Dey [8] has used entropy based fuzzy c-means to initialize the fuzzy c-mean initial cluster centers. Maulik [9] have applied GA to find the k-means cluster centers using floating point representation of clustering three datasets: iris, crude oil and vowels dataset.

2 Diabetic Data Set

Diabetes mellitus is a disease in which the body is unable to produce or unable to properly use and store glucose (a form of sugar). Glucose backs up in the bloodstream causing one's blood glucose or "sugar" to rise too high. There are two major types of diabetes. World Health Organization (WHO) report had shown a marked increase in the number of diabetics and this trend is expected to grow in the next couple of decades. In the International Diabetes Federation Conference 2003 held in Paris, India was labeled, as "Diabetes Capital of the World," as of about 190 million diabetics worldwide, more than 33 million are Indians. The worldwide figure is expected to rise to 330 million, 52 million of them Indians by 2025, largely due to population growth, ageing, urbanization, unhealthy eating habits and a sedentary lifestyle [9,10]. By 2030, India's diabetes burden is expected to cross the 100 million mark as against 87 million earlier estimated.The PIMA diabetic dataset is availed from UCI Machine Learning Repository. The database consist of two categories in the data set (i.e. Tested positive, Tested Negative) each having 8 features: Number of times pregnant, Plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure (mm Hg), Triceps skin fold thickness (mm), 2-Hour serum insulin (mu U/ml), Body mass index (weight in kg/ (height in m) ^2), Diabetes pedigree function and Age (years). A total of 768 cases are available in PIDD. 5 patients had a glucose of 0, 11 patients had a body mass index of 0, 28 others had a diastolic blood pressure of 0, 192 others had skin fold thickness readings of 0, and 140 others had serum insulin levels of 0. After deleting these cases there were 392 cases with no missing values (130 tested positive cases and 262 tested negative) [11].

3 K-Means Clustering

K-means [12] is one of the simplest unsupervised learning algorithms and follows partitioning method for clustering. K-means algorithm takes the input parameter, k as number of clusters and partitions a dataset of n objects into k clusters, so that the resulting objects of one cluster are dissimilar to that of other cluster and similar to

objects of the same cluster. In k-means algorithms begins with randomly selected k objects, representing the k initial cluster center or mean. Next each object is assigned to one the cluster based on the closeness of the object with cluster center. To assign the object to the closest center, a proximity measure namely Euclidean distance is used that quantifies the notion of closest. After all the objects are distributed to k clusters, the new k cluster centers are found by taking the mean of objects of k clusters respectively. The process is repeated till there is no change in k cluster centers. K-means algorithm aims at minimizing an objective function namely sum of squared error (SSE).

SSE is defined as $$E = \sum_{i=1}^{k} \sum_{p \in C_i} |p - m_i|^2 \quad (1)$$

where E is sum of the square error of objects with cluster means for k cluster. p is the object belong to a cluster Ci and mi is the mean of cluster Ci .The time complexity of K-means is O (t*k*n) where t is the number of iterations, k is number of clusters and n is the total number of records in dataset.

K-means partitioning algorithm:

Input is k is the number of clusters, D is input data set .Output is k clusters.

1. Randomly choose k objects from D as the initial cluster centers.
2. Repeat.
3. Assign each object from D to one of k clusters to which the object is most similar based on the mean value of the objects in the cluster.
4. Update the cluster means by taking the mean value of the objects for each of k cluster.
5. Until no change in cluster means/ min error E is reached.

4 GA Based K-Means Initial Cluster Centers

Genetic algorithm (GA) [13] is an optimization techniques inspired by natural selection and natural genetics. Unlike many search algorithms, which perform a local, greedy search, GA is a stochastic general search method, capable of effectively exploring large search spaces. A genetic algorithm is mainly composed of three operators: reproduction, crossover, and mutation. As a first step of GA, an initial population of individuals is generated at random or heuristically. The individuals in the genetic space are called chromosome. The chromosome is a collection of genes where genes can generally be represented by different methods like binary encoding, value encoding, permutation encoding and tree encoding. In each generation, the population is evaluated using fitness function. Next comes the selection process, where in the high fitness chromosomes are used to eliminate low fitness chromosomes. But selection alone does not produce any new individuals into the population. Hence selection is followed by crossover and mutation operations. Crossover is the process by which two-selected chromosome with high fitness values exchange part of the genes to generate new pair of chromosomes. The crossover tends to facilitate the evolutionary process to progress toward potential regions of the solution space. Mutation is the random change of the value of a gene, which is used to prevent premature convergence to local optima. The new population generated undergoes the further selection,

crossover and mutation till the termination criterion is not satisfied. Convergence of the genetic algorithm depends on the various criterions like fitness value achieved or number of generations [14-15].

GA has been used in this paper identify initial k-means cluster centers. Chromosomes are encoded using binary encoding where 1 represents the sample selected as initial cluster center and 0 represents the sample is not selected as initial cluster center. The length of the chromosome is equal to number of samples. Each of the chromosomes has exactly k number of ones, where k represents the number of cluster. Once the terminating condition is reached, the highest fittest chromosome (with the least sum of square error (SSE)) decides which samples will be k-means initial cluster centers.

The working of GA for finding the initial k-means cluster:

Step1. Initialize the chromosome population randomly using binary encoding (Where one's represent the sample number as cluster center)

Setp2. Repeat the steps a-e following till terminating condition is reached

a) Apply k –means clustering to individual chromosome and find the SSE
b) Replace the low fit chromosome by highest fit chromosome (with least SSE).
c) Select any two chromosomes randomly and apply crossover operation
d) Apply mutation operation by randomly selecting any one chromosome and randomly Change the bit 1 to 0 and bit 0 to 1.

Setp3. The position of 1 bit in the best-fit chromosome decides the samples, are selected as initial k –means cluster centers.

5 EFC Based K-Means Initial Cluster Centers

Entropy based fuzzy clustering [16], identifies the number of clusters and initial cluster prototypes by itself. The entropy is calculated for each sample using equation (2).

$$E_i = \sum_{k \in x}^{j \neq i} (S_{ij} \log_2 S_{ij} + (1 - S_{ij}) \log_2 (1 - S_{ij})) \quad (2)$$

where $S_{ij} = e^{-\alpha d_{ij}}$ is the similarity between two data points (i, j) and dij is the Euclidean distance between points (i, j)

The algorithm for entropy based fuzzy clustering is as follows. The inputs for the algorithm are dataset D with N samples,β the threshold value, can be viewed as a threshold of similarity among the data points in the same cluster, an constant α is which is computed as (ln 0.5/ ($\bar{D}$)), where $\bar{D}$ is the mean distance among the pairs of data points in a hyper-space and is usually set to 0.5.

Step 1. Compute entropy Ei for each sample xi from dataset D for I = 1 to N
Step 2. Identify xi that has the minimum Ei value as the cluster center.
Step 3. Remove xi and data points having similarity xi > than some threshold β from D
Step 4. If D is not empty then go to step 2.

The k centroids identified by EFC, are selected as k-means initial cluster centers.

6 Experimental Results

GA and EFC have been used to identify initial k-means cluster centers. Compared to floating point encoding, binary encoding requires more space, but the crossover and mutation can be easily implemented with binary encoding. Hence binary encoding has been used in this paper. The GA was experimented with population's size of 40-120 chromosomes, number of generations with 15 to 30 and with both one point and two-point crossover. After mutation and crossover operation the number of ones in the chromosomes must be checked for not exceeding the number of required clusters. The best results with GA are found with populating size of 100. The one point and two-point crossovers generate almost the same results. The terminating condition is 80 % of the chromosomes represent the same initial cluster centers. The performance of EFC depends o the parameters β and α. Different values of β ranging from 0.4 to 0.7 and was experimented. The β with 0.6 gave the best results with constant α of 0.5. The k-means clustering performance for the standard medical data set: PIMA Indian diabetic dataset is improved by using the initial cluster centers identified by GA and EFC. The clustering performance of K-means in terms of classification error, number of iterations, sensitivity, specificity, Recall, Precision, f-measure and SSE with random cluster centers, EFC and GA is shown in Table 1 The time taken in milliseconds and improved classification accuracy of K-means with random, GA and EFC initialized cluster centers is shown in Fig 1. The classification accuracy of k-means is found to be 69.14%, 71.70% and 72.80% with initial centers initialized by random, EFC and GA method.

Table1. Performance of K-means clustering by initializing centers using random, GA and EFC

Method for K-means initial cluster center	TP	FP	TN	FN	# Iterations	Sensitivity	Specificity	Recall	F-measure	Precision	SSE	Classification Error %
Random	74	56	197	65	7	0.53	0.78	0.53	0.55	0.57	0.0087	30.86
EFC	89	41	192	70	5	0.56	0.82	0.56	0.61	0.68	0.0083	28.30
GA	86	44	197	65	5	0.57	0.82	0.57	0.61	0.66	0.0033	27.80

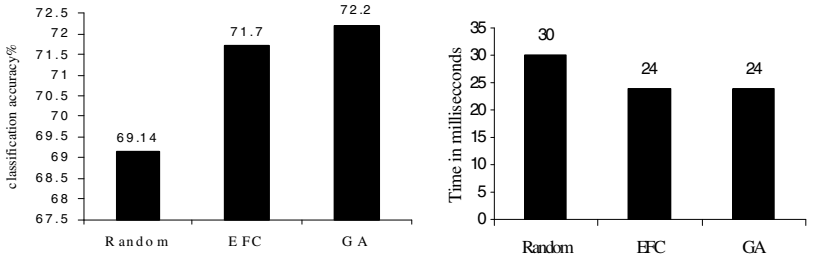


Fig. 1. a) Classification accuracy (b)Execution time in milliseconds for K-means clustering by initializing cluster centers using Random, GA and EFC

7 Conclusions

The performance of K-means clustering depends on the initial cluster centers. Entropy based fuzzy clustering identifies the number of clusters and initial cluster prototypes by itself which is given as initial cluster centers for k-means clustering. GA is a stochastic general search method, capable of effectively exploring large search spaces and has been used to identify the initial cluster centroids for k-means clustering using binary encoding of chromosomes. This paper illustrates the improvement in classification accuracy and reduction in execution time for K-means clustering by initializing the cluster centers using GA and EFC.

References

1. Han, Kamber, M.: Data Mining: Concepts and Techniques. Morgan Kauffmann Publishers, San Francisco (2001)
2. Bradley, P.S., Fayyad, U.M.: Refining initial points for k-means algorithm. In: Proceedings of the 15th International Conference on Machine Learning (1998)
3. Al-Shbour, B., Myaeng, S.-H.: Initializing K-means using Genetic Algorithm. World Academy of Science, Engineering and Technlogy 54, 114–118 (2009)
4. Jimenez, J.F., Cuevas, F.J., Carpio, J.M.: Genetic Algorithms applied to Clustering Problem and Data Mining. In: Proceedings of the 7th WSEAS International Conference on Simulation, Modeling and Optimization, pp. 219–224 (2007)
5. Hao-Jun, Lang-Haun: Genetic Algorithm-based High-dimensional Data Clustering Technique. In: Sixth International Conference on Fuzzy Systems and Knowledge Discovery, pp. 485–489 (2009)
6. Eltibi, M.F., Ashour, W.M.: Initializing K-means Clustering Algorithm using Statistical Information. International Journal of Computer Applications 29, 51–55 (2011)
7. Erisogly, M., Calis, N., Sakalliglu, S.: A new algorithms for initial cluster centers in k-means algorithm. Pattern Recognition Letters 32, 1701–1705 (2011)
8. Dey, V., Pratihar, D.K., Datta, G.L.: Genetic algorithm-tuned entropy-based fuzzy C-means algorithm for obtaining distinct and compact clusters. Fuzzy Optimization Decision Making 10, 153–166 (2011)
9. Editorial, Diagnosis and Classification of Diabetes Mellitus, American Diabetes Association, Diabetes Care 27(suppl.1) (January 2004)
10. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus: Follow up report on the Diagnosis of Diabetes Mellitus. Diabetic Care 26, 3160–3167 (2003)
11. Breault, J.L.: Data Mining Diabetic Databases: Are rough Sets a Useful Addition? (2001), `http://www.galaxy.gmu.edu/interface/I01/I2001Proceedings/Jbreault`
12. Mac Queen, J.: Some methods for the classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probility, pp. 281–297. University of California Press, Berkeley (1967)
13. Goldberg, D.: Genetic Algorithms in Search, Optimization, and Machine learning. Addison Wesley (1989)
14. Rajasekaran, S., Vijayalakshmi Pai, G.A.: Genetic Algorithm based Weight Determination for Backpropogation Networks. In: Proc. of the Fourth Int. Conf. on Advanced Computing, pp. 73–79 (1996)
15. Maulik, U., Bandopadhyay, S.: Genetic Algorithm-Based Clustering Technique. Pattern Recognition 33, 1455–1465 (1999)
16. Yao, J., Dash, M., Tan, S.T., Liu, H.: Entropy based fuzzy clustering and fuzzy modeling. Fuzzy Sets and Systems 113, 381–388 (2000)

Hindi and English Off-line Signature Identification and Verification

Srikanta Pal[1], Umapada Pal[2], and Michael Blumenstein[1]

[1] School of Information and Communication Technology,
Griffith University, Gold Coast, Australia
srikanta.pal@griffithuni.edu.au
m.blumenstein@griffith.edu.au
[2] Computer Vision and Pattern Recognition Unit,
Indian Statistical Institute, Kolkata, India
umapada@isical.ac.in

Abstract. Biometric systems play a significant role in the field of information security as they are extremely required for user authentication. Signature identification and verification have a great importance for authentication intention. The purpose of this paper is to present an empirical contribution towards the understanding of multi-script (Hindi and English) signature verification. This system will identify whether a claimed signature belongs to the group of English signatures or Hindi signatures from a combined Hindi and English signature datasets and then it will verify signatures using these two resultant signature datasets (Hindi script signature and English script signatures) separately. The modified gradient feature and SVM classifier were employed for identification and verification purposes. To the best of authors' knowledge, the multi-script signature identification and verification has never been used for the task of signature verification and this is the first report of using Hindi and English signatures in this area. Two different results for identification and verification are calculated and analysed. The accuracy of 98.05% is obtained for the identification of signature script using 2160 (1080 Hindi + 1080 English) samples for training and 1080 (540 Hindi + 540 English) samples for testing. The resultant data sets obtained in script identification of signatures were used for verification purpose. The FRR, FAR for Hindi and English was obtained 8.0%, 4.0% and 12.0%, 10.0% respectively.

Keywords: Signature verification, biometrics, SVMs, Gradient Feature.

1 Introduction

Signature verification has been a topic of intensive research during the past several years [1-8] due to an important role it plays in numerous areas including the financial system. The verification of human signatures is particularly concerned with the improvement of the interface between human beings and computers [2]. A signature verification system and the associated techniques used to solve the inherent problems of authentication can be divided into two classes [3]: (a) on-line methods [4] to measure temporal and sequential data by utilizing intelligent algorithms [5] and (b)

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 905–910.
springerlink.com

off-line methods [6] that use an optical scanner to obtain handwriting data written on paper. Off-line signature verification deals with the verification of signatures, which appear in a static format [7]. On-line signature verification has been shown to achieve much higher verification rates than off-line verification [6], since a considerable amount of dynamic information is lost in the off-line mode. However off-line systems have a significant advantage as they do not require access to special processing devices when the signatures are produced. Moreover, the off-line group has many more practical application areas than that of its on-line counterpart.

2 Database Preparation and Pre-processing

A database of 1620 Hindi signatures and 1620 English signatures are used for identification purpose. English signatures from GPDS were used in our experimentation. Each Hindi and English signature set consists of 24 genuine signatures and 30 skilled forgeries. A total number of 720 genuine Hindi signatures from 30 individuals were collected. For each contributor, all genuine specimens were collected in a single day's writing session. In order to produce the forgeries, the imitators were allowed to practice their forgeries as long as they wished with static images of genuine specimens. A total number of 900 Hindi skilled forgeries were collected from the writers.

3 Modified Gradient Feature

The gray-scale local-orientation histogram of the component is used for 576 dimensional feature extractions. To obtain 576-dimensional gradient-based feature vector, the following steps are executed.

Step 1: A 2 x 2 mean filtering is applied 5 times on the input image.

Step 2: The gray-scale image obtained in Step 1 is normalized so that the mean gray scale becomes zero with maximum value 1.

Step 3: The normalized image is then segmented into 17x7 blocks. Compromising trade-off between accuracy and complexity, this block size is decided experimentally. To get the bounding box of the grey-scale image, the image is converted into two-tone using Otsu's thresholding algorithm [9]. This will exclude unnecessary background information from the image.

Step 4: A Roberts filter is then applied on the image to obtain the gradient image. The arc tangent of the gradient (direction of gradient) is quantized into 32 directions and the strength of the gradient is accumulated with each of the quantized direction. The strength of the Gradient $f(x, y)$ is defined as follows:

$f(x, y) = \sqrt{(\Delta u)^2 + (\Delta v)^2}$ and the direction of gradient $(\theta(x, y))$ is: $\theta(x, y) = \tan^{-1} \frac{\Delta v}{\Delta u}$ Where $\Delta u = g(x+1, y+1) - g(x, y)$ and $\Delta v = g(x+1, y) - g(x, y+1)$ and $g(x, y)$ is the gray level of (x, y) point.

Step 5: Histograms of the values of 32 quantized directions are computed for each of the 17 x 7 blocks.

Step 6: The directional histogram of the 17 x 7 blocks is down sampled into 9 x 4 blocks and 16 directions using Gaussian filters. Finally, a 9 x 4 x 16 = 576 dimensional feature vector is obtained.

4 Classifier and Experimental Settings

In our experiments, we have used Support Vector Machines (SVM) as classifiers. SVMs have been originally defined for two-class problems and they look for the optimal hyper plane, which maximizes the distance and the margin between the nearest examples of both classes, namely support vectors (SVs). Given a training database of M data: $\{x_m | m=1,..., M\}$, the linear SVM classifier is then defined as:

$$f(x)=\sum_j \alpha_j x_j \cdot x+b$$

where $\{x_j\}$ are the set of support vectors and the parameters α_j and b have been determined by solving a quadratic problem [7]. The linear SVM can be extended to various non-linear variants; details can be found in [7, 8]. In our experiments, the RBF kernel SVM outperformed other non-linear SVM kernels, hence we are reporting our recognition results based on the RBF kernel only. The experimental settings we used are described below.

4.1 Settings for Script Identification

For the experiments in the proposed research, our developed Hindi signature database described in section 4 was used. A numbers of 60 set of signatures (30 Hindi dataset and 30 English dataset) were used for identification of signature script. A signature samples of 1080(20x54) Hindi and 1080(20x54) English were used for training phase whereas 540 (10x54) Hindi and 540(10x54) English signature samples were used for testing purpose for identification of signature script. The number of samples for training and testing for experimentation of identification are shown in Table 1.

4.2 Settings for Signature Verification

The accuracy of 98.05% is obtained for the identification of signature script using 1080 (540 Hindi + 540 English) samples for testing. SVMs classifier misidentified 21signatures, i.e 1.95% (100.00-98.05) of 1080 samples. The number of errors occurred in testing dataset for identification is shown in Table 2. The signature verification was done using 1059(1080-21 samples) correctly identified script of signatures. For verification, the database was split in two parts, to perform the training and testing components. The signature samples of 466 genuine signatures (226 Hindi+ 240 English) and 595 skilled forgeries (299 Hindi + 296 English) were used for verification purpose from 10 set of Hindi signatures and 10 set of English signatures, respectively. For each signature set, an SVM was trained with 14

randomly chosen genuine signatures. The negative samples for training were the 20 skilled forgeries of signatures. For testing, the remaining genuine signatures and remaining skilled forgeries were used. The Hindi and English signature samples used for verification with each signature set are shown in Table 3 and Table 4.

Table 1. Number of Signature Samples Used for Identification of Signature Script

	Hindi Signature		English Signature	
	Genuine	**Forged**	**Genuine**	**Forged**
Training	480	600	480	600
Testing	240	300	240	300

Table 2. Number of Signature Script Identification Errors Occurred in Different Datasets

No. of Errors in Test Datasets Obtained in Identification Part				
	Hindi Test Samples		**English Test Samples**	
Datasets	**Genuine Signatures**	**Forged signatures**	**Genuine Signatures**	**Forged signatures**
Set-1	0	0	0	0
Set-2	2	0	0	0
Set-3	1	0	0	0
Set-4	0	0	0	0
Set-5	0	0	0	0
Set-6	4	1	0	4
Set-7	7	0	0	1
Set-8	0	0	0	1
Set-9	0	0	0	0
Set-10	0	0	0	0
Total Errors	14	1	0	6

5 Results and Discussion

As mentioned earlier, the accuracy of 98.05% is obtained for the identification of signature script. Using the Gradient feature, an FAR (False Acceptance Rate) and FRR (False Rejection Rate) was computed. At this operational point, the FRR, FAR for Hindi were 8.0%, 4.0% and FRR, FAR for English were 12.0 %, 10.0% respectively. The FRR, FAR and AER (Average Error Rate) obtained from our experiments are shown in Table 5. The AER obtained in this research is 6.0% for Hindi and 11.0 % for English.

Confusion matrix of signature identification obtained from SVM classifiers and gradient features are shown in Table 6. It is noted that only 6 English signatures were misidentified as Hindi signatures and 15 Hindi signatures were misidentified as English. Two samples of signature script identification errors (English and Hindi signature treated as Hindi and English respectively) are shown in Figure 1 and figure 2.

Some verification errors (Hindi and English genuine signature treated as Hindi and English forged signature and Hindi and English forged signature treated as Hindi and English genuine signature) are shown in Figure 3, Figure 4 and Figure 5, Figure 6, respectively.

Table 3. Hindi Samples Used for Verification

Hindi datasets used for verification		
Hindi Datasets	**Genuine Signatures**	**Forged signatures**
Set-1	24	30
Set-2	22	30
Set-3	23	30
Set-4	24	30
Set-5	24	30
Set-6	20	29
Set-7	17	30
Set-8	24	30
Set-9	24	30
Set-10	24	30
Total samples	**226**	**299**

Table 4. English Samples Used for Verification

English datasets used for verification		
English Datasets	**Genuine Signatures**	**Forged signatures**
Set-1	24	30
Set-2	24	30
Set-3	24	30
Set-4	24	30
Set-5	24	30
Set-6	24	26
Set-7	24	29
Set-8	24	29
Set-9	24	30
Set-10	24	30
Total samples	**240**	**294**

Table 5. Results of FRR, FAR and EER

	FRR	**FAR**	**AER**
Hindi	8.0 %	4.0 %	6.0 %
English	12.0%	10.0%	11.0 %

Table 6. Confusion Metric for Identification of Signature Script

	Hindi	**English**
Hindi	525	15
English	6	534

Fig. 1. English Signature Sample Treated as Hindi Signature

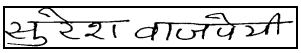


Fig. 2. Hindi Signature Sample Treated as English Signature

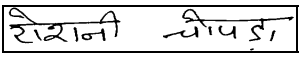


Fig. 3. Hindi Genuine Signature Sample Treated as Hindi Forged Signature

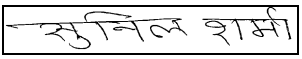


Fig. 4. Hindi Forged Signature Treated as Hindi Genuine Signature

Fig. 5. English Genuine Signature Sample Treated as English Forged Signature

Fig. 6. English Forged Signature Sample Treated as English Genuine Signature

6 Conclusions and Future Work

This paper presents a signature identification and verification scheme of bi-script off-line signatures. To the best of our knowledge, bi-script signatures have never been used for the task of signature verification and this is the first report in this area. This scheme of bi-script off-line signature identification is a novel contribution to the field of signature verification. In near future, we plan to extend our work for multi-script off-line signature identification and verification.

References

[1] Chen, S., Srihari, S.: Use of Exterior Contour and Shape Features in Off-line Signature Verification. In: 8th ICDAR, pp. 1280–1284 (2005)
[2] Ferrer, M.A., Alonso, J.B., Travieso, C.M.: Off-line Geometric Parameters for Automatic SignatureVerification Using Fixed-Point Arithmetic. IEEE PAMI 27(6), 993–997 (2005)
[3] Madabusi, S., Srinivas, V., Bhaskaran, S., Balasubramanian, M.: On-line and off-line signature verification using relative slope algorithm. In: International Workshop on Measurement Systems for Homeland Security, pp. 11–15 (2005)
[4] Emerich, S., Lupu, E., Rusu, C.: On-line Signature Recognition Approach Based on Wavelets and Support Vector Machines. In: Intl Conf. on Automation Quality and Testing Robotics, pp. 1–4 (2010)
[5] Kholmatov, A., Yanikoglu, B.: Identity Authentication using improved online signature verification method. PRL 26, 2400–2408 (2005)
[6] Kalera, M., Srihari, S., Xu, A.: Offline signature verification and identification using distance statistics. In: IJPRAI, pp. 1339–1360 (2004)
[7] Vapnik, V.: The Nature of Statistical Learning Theory. Springer (1995)
[8] Burges, C.: A Tutorial on support Vector machines for pattern recognition. In: Data Mining and Knowledge Discovery, pp. 1–43 (1998)
[9] Otsu, N.: A threshold selection method from gray-level histogram. IEEE Trans. on SMC 9, 62–66 (1979)

A Robust Method of Image Based Coin Recognition

B.V. Chetan and P.A. Vijaya

Dept. of E&C Engg., Malnad College of Engineering, Hassan-573201, Karnataka, India
chetanbv019@gmail.com, pav@mcehassan.in

Abstract. An image based approach which detects Indian coins of different denomination has been proposed in this paper. This consists of matching an input coin image with a database of coin images (templates) in two phases. The first phase involves, identifying matching radius database coins. In the second phase, template matching is performed by correlating the edges of input and matching radius database coin images. Template matching phase involves two parts, coarse matching and fine matching. This provides rotation invariance and does away with the requirement of placing the front face of the coin up. If the correlation coefficient obtained by template matching satisfies the threshold, then coin stands recognized. The algorithm has been developed in MATLAB 7.9.0 and the obtained results are recorded. The proposed method has been compared with existing methods (Difference, LBP, FFT) and comparison results have been recorded.

1 Introduction

Currently major coin recognition machines rely on physical properties of coins. Most of the coin testers in slot machines, work by testing physical properties of coins such as size, weight and materials. However, if physical similarities exist between coins of different currencies, then the traditional coin testers would fail to distinguish the different coins [1-8]. So there is a need of a robust real time system which can actually visualise the coin and recognise it.

As the coins are in frequent usage in daily life the surfaces of same pattern coins would not be same. Fig. 1 shows an example of this problem. The above image based methods [1-8] of coin recognition would not yield proper result in this case, as the intensity values change from image to image of same denomination or pattern.

To counter this problem the proposed method makes use of edge detection and correlating the edge detected images during template matching. After edge detection the image matrix consists of 1s or 0s, instead of 0-255 values, which gives better matching results. Proposed method also involves radius matching due to the fact that different Indian coins have different radius as shown Fig. 2.

Fig. 1. Comparison of two coins of same pattern

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 911–918.
springerlink.com

Fig. 2. Front and back faces of different Indian coins

2 Proposed Rotation-Invariant and Side-Invariant Coin Recognition Method

Fig. 3 shows the block diagram of overall recognition process. First a database of coin images is created and recognition is carried out. Following are the stages involved in the overall process.

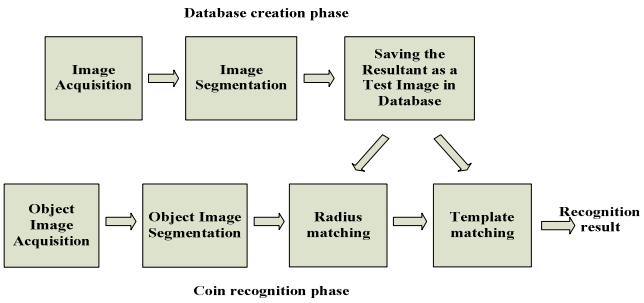


Fig. 3. Block diagram of the recognition process

2.1 Image Acquisition and Segmentation

A digital camera is used for image acquisition. The next step of the coin recognition system would be image segmentation, i.e. separating the coin image from the background. Color-based segmentation using K-means clustering method is employed, to get the binary image of the coin. Metric calculation is performed to confirm the input is a coin. Metric close to 1 indicate a round object. Obtaining the position of the binary image of coin, the parent image is re-fitted into it.

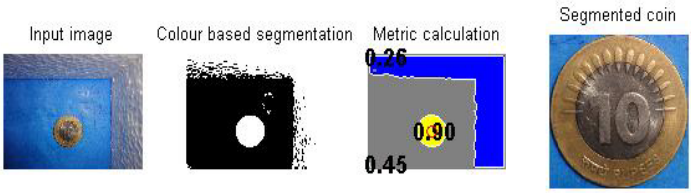


Fig. 4. Steps involved in image acquisition and segmentation

2.2 Radius Matching

After segmentation the diameter of the resultant image is found by taking average of the number of pixels in column-wise and row-wise. The half of the diameter gives the radius in terms of pixels. The database coins which have same radius as input coin are noted for further phase of template matching.

2.3 Template Matching

Fig. 5 describes the entire process of template matching. This involves two parts: coarse matching and fine matching. The matching in both parts is performed by correlating the edges of input and database coins. The correlation coefficient is given by,

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}} \quad (1)$$

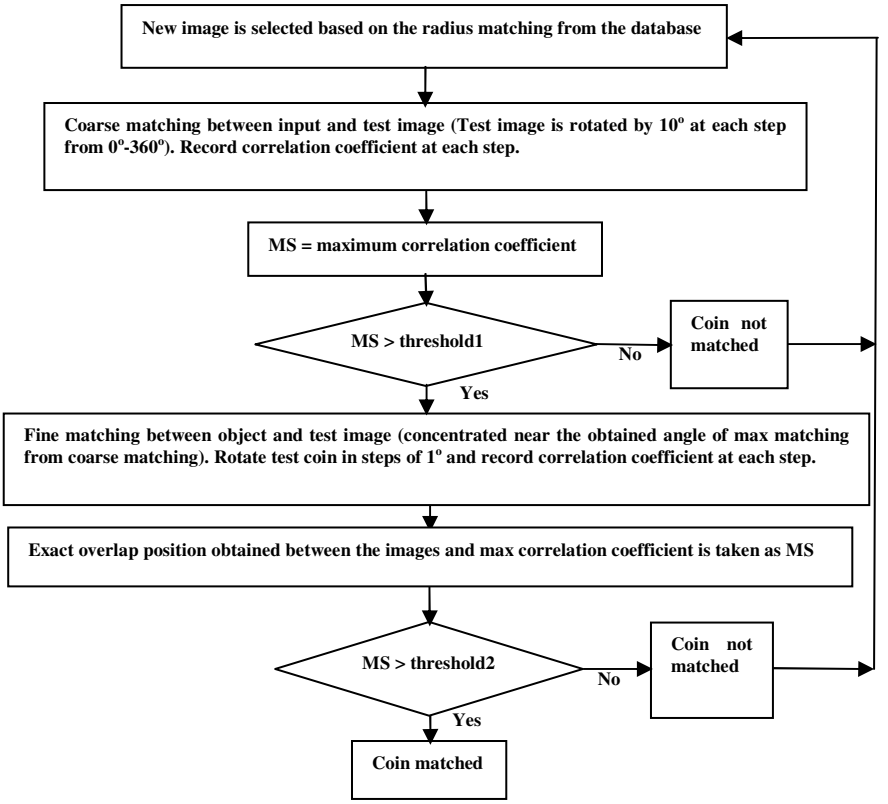


Fig. 5. Flow chart of the designed template matching algorithm

where, r is correlation coefficient, A & B are two image matrices, m & n are number of rows & columns respectively in both A & B, $\overline{A}$ & $\overline{B}$ are the means of matrices A & B respectively.

Equation 1 is used to obtain the matching score (MS) in both coarse and fine matching parts. Edge detection is the preliminary step of matching. Multi-scale edge detection is used for edge detection of the input and database coins which are about to be matched. Fig. 6 shows an example result of edge detection obtained by the proposed method.

Fig. 6. Result of edge detection

2.4 Decision Making

The database coin image giving maximum MS with input coin is the matching database coin. Hence the denomination of input coin is detected. Fig. 7 shows the final result of matching of two coins. Fig. 8 shows the resultant graph of correlation coefficients versus angles for coarse matching. Fig. 9 shows the resultant graph of correlation coefficients versus angles for fine matching.

Fig. 7. Input coin and its matching database coin

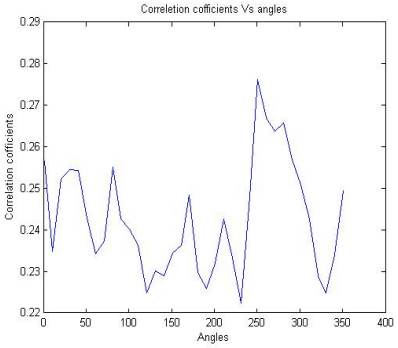


Fig. 8. Coarse matching resultant graph

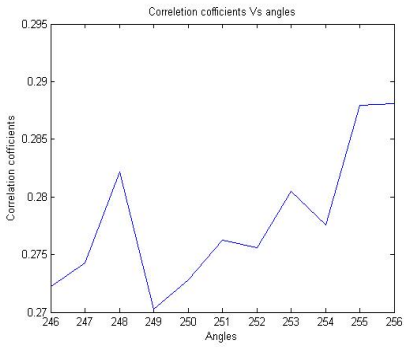


Fig. 9. Fine matching resultant graph

The correlation coefficient is 2.762570e-001 at 251° during coarse matching. The exact match is found at 256° with correlation coefficient 2.881234e-001 by fine matching. Hence the final matching score is 2.881234e-001 at 256°.

3 Experiments and Results

The parameters such as: lighting condition, distance between camera and coin, position of camera and coin are kept constant during image acquisition. Care is taken to keep the surface of the coin clean.

3.1 Recognition Capacity of the Method

A coin has two faces and hence the recognition method must be capable of matching the input with correct pattern of database coin. Two test sets consisting of 20 patterns each are formed. The patterns of first set are matched with patterns of same radius in second set. Fig. 10 shows the outcome of this process. From Fig. 10, by considering 0.28 as threshold, the following parameters have been found:

True acceptance ratio (TAR) = 0.9
True rejection ratio (TRR) = 0.84
False acceptance ratio (FAR) = 0.16
False rejection ratio (FRR) = 0.1.

Hence, the proposed method yields accuracy of 90% in recognizing input coins.

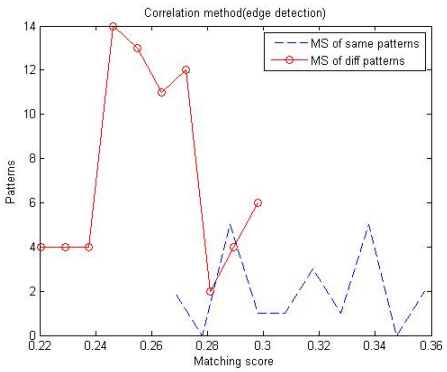


Fig. 10. Graph of recognition capacity of the proposed method

3.2 Comparison of the Proposed Method with other Existing Coin Recognition Methods

Four input sets of 20 patterns each are formed. Each input set consists of coins different from other sets. The input sets consist of both old/used coins and new coins. The database includes all 20 patterns which are in input sets, but the coins of input sets and database are different.

Table 1. Comparison of proposed method with other 3 existing methods

Method	Input set1	Input set2	Input set3	Input set4	Overall accuracy
Proposed	95%	90%	85%	100%	92.5%
Difference	25%	40%	45%	45%	38.75%
LBP	30%	35%	40%	35%	35%
FFT	30%	30%	35%	40%	33.75%

Better comparison of proposed technique with other 3 existing methods can be achieved by plotting the graphs for other 3 methods similar to Fig. 10.

Figures 11, 12 and 13 show that the existing methods such as, difference, LBP and FFT have very less recognition capacity. The chance of false acceptance or false rejection of input coins is always more in these three methods. Hence, the proposed method is more efficient than these three methods.

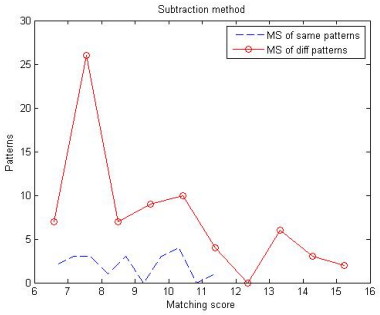


Fig. 11. Recognition capacity graph for difference method of coin recognition

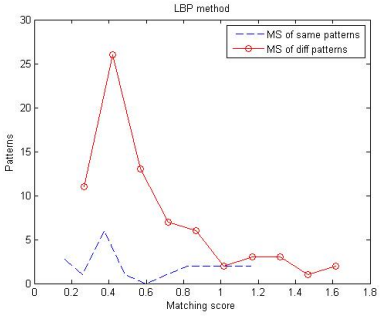


Fig. 12. Recognition capacity graph for local binary pattern (LBP) method of coin recognition

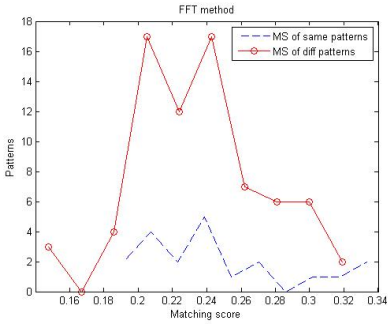


Fig. 13. Recognition capacity graph for FFT method of coin recognition

4 Conclusion

The proposed method for coin recognition is found to be simple and accurate. This method yields an accuracy of almost 90%, which is much greater than accuracies of other existing methods. The method provides rotation invariance and coin side invariance for recognition process. The method also avoids the dependence on constant light factor during image acquisition up to certain extent. Also two subsequent template matching steps are provided to give precise results. This solves a real life problem where physical similarities between these coins.

Future works will include modifications of the technique and also merging of other image processing techniques, such as, neural networks training using edge detection which would completely extricate the process from the dependency over standard light intensity acquisition adding on to the accuracy of the process.

Acknowledgments. Thanks to Mr. Hiren B. Parmar, Senior Engineer, VED Labs, Bangalore, for his suggestions and support.

References

1. Chen, C.-M., Zhang, S.-Q., Chen, Y.-F.: A Coin Recognition System with Rotation Invariance. In: MVHI, April 24-25, pp. 755–757 (2010), doi:10.1109/MVHI.2010.60
2. Chen, H.: Chinese Coin Recognition Based on Unwrapped Image and Rotation Invariant Template Matching. In: ICINIS 3, November 1-3, pp. 5–7 (2010), doi:10.1109/ICINIS.2010.10
3. Shen, L., Jia, S., Ji, Z., Chen, W.-S.: Statictics of Gabor Features for Coin Recognition. In: IST 2009, May 11-12, pp. 295–298 (2009), doi:10.1109/IST.2009.5071653
4. Shen, L., Jia, S., Ji, Z., Chen, W.-S.: Extracting local texture features for image Based Coin Recognition. IET 5(5), 394–401 (2011), doi:10.1049/iet-ipr.2009.0251
5. Fukumi, M., Omatu, S., Takeda, F., Kosaka, T.: Rotation-Invariant Neural Pattem Recognition System with Application to Coin Recognition. IEEE Transactions on Neural Networks 3(2), 272–279 (1992), doi:10.1109/72.125868

6. Thumwarin, P., Malila, S., Janthawong, P., Pibulwej, W., Matsuura, T.: A Robust Coin Recognition Method with Rotation Invariance. In: ICCCAS, June 25-28, pp. 520–523 (2006), doi:10.1109/ICCCAS.2006.284690
7. Bremananth, R., Balaji, B., Sankari, M., Chitra, A.: A New Approach to Coin Recognition using Neural Pattern Analysis. In: Indicon, December 11-13, pp. 366–370 (2005), doi:10.1109/INDCON.2005.1590191
8. Gupta, V., Puri, R., Verma, M.: Prompt Indian Coin Recognition with Rotation Invariance using Image Subtraction Technique. In: ICDeCom, February 24-25, pp. 1–5 (2011), doi:10.1109/ICDECOM.2011.5738496

N-Gram Based Approach to Automatic Tamil Lyric Generation by Identifying Emotion

Rajeswari Sridhar, Jalin Gladis D., Ganga K., and Dhivya Prabha G.

Department of Computer Science and Engineering,
Anna University, Chennai, India
{rajisridhar,jalingladis18,kss.ganga,be.cooldivi}@gmail.com

Abstract. This paper discusses a tri-gram approach to automatic Tamil lyric generation. The approach is based on identifying the emotion from a given scenario and uses this emotion as a seed word to interpret the context of the scenario. A lyric model based on tri-gram is constructed which is referred using the identified seed word to generate lyrics. The lyric model, tri-gram of words, Tamil sentence rules and suffixes are used by a Morphological generator to generate lyrics. Using this approach we achieved an average accuracy of 74.17% with respect to exact emotion being conveyed in the generated lyrics.

1 Introduction

Poetry is an art and a natural talent of creativity of an individual [1]. Unorganized poetry can be thought of as lyric which is defined as a collection of words that together convey an emotion or feeling. Lyrics are part and parcel of Indian film industry, to enhance and describe the story of a movie. Writing lyrics for a movie is challenging as it should blend with the story of the movie, and therefore requires proper understanding of the story. Hence, lyrics can be thought of as a poetry incorporating the innovation, creativity and expressing the underlying story of a movie.

In today's scenario, lyricists are given a situation from the movie along with a tune for which they have to write lyrics [2] [3]. They alter the words in the lyrics not only to the given situation but mostly to the tune. In case of discrepancies of the lyrics with the tune the parameters of the lyrics like words, rhyming words, and vocative words are adjusted to resolve the conflicts by maintaining the tune as such. This restricts the imagination and the innovation of the lyricist. In the earlier generation lyricist were given a scenario from the movie for which they write lyrics and tune is made in accordance for the lyrics [4]. In fact lyricist refuse to write lyrics for a pre-define tune. Hence, the job of writing lyrics by understanding the context of the scene is more challenging for automating and hence we have chosen this method to automatically generate lyrics. In this work, we discuss the algorithm that we have proposed for automatic lyric generation for Tamil language movies.

This paper is organized as follows: Section 2 discusses some existing work on lyric generation, Section 3 discusses the system architecture in detail, Section 4 elaborates on the results and findings of our work and Section 5 discusses the future work in this domain.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 919–926.
springerlink.com

2 Literature Survey

One of the works for generating lyrics for Tamil [2], focuses on creating meaningful lyrics given a melody in ABC which is the 'asai' pattern specific to Tamil language [5]. This is based on the grammar representation of Tamil language namely KNM which indicates the syllables as 'Kuril', 'Nedil', 'Muttru'. The melody is analyzed and a series of possible syllable patterns is generated in KNM representation scheme. The central idea lies in choosing a selection restriction rule which has the following specifications in it – A verb along with the characteristics of the subject and object that is normally used with that verb. Using this, sentences are framed which are matched and validated against the given ABC pattern. If it not confining to the given ABC pattern the same process is repeated until an expected outcome is achieved.

The lines generated using the technique mentioned were not confining to a particular context. The lyrics generated were not domain specific and conveyed incongruent thoughts. Hence to get an idea of how to generate context specific lyrics we referred to the work of Burr Settles [6], which aims at providing a set of words that are related to the given seed word. The author has generated automatic word suggestion in English given a seed word. The system's Lyric cloud is the heart of the tool. Lyric cloud takes a seed word and provides up to 25 related words. The suggestions of alternate words for a given seed can be useful in generating rhyming poetry.

After generating words, we need to combine them to a meaningful poetry sentence. We need some rules for generating sentences in Tamil. We refer to the work of S. Lakshmana Pandian and T.V.Geetha [7] which uses a machine learning technique based on bigram and semantic roles for Tamil sentence generation. This module identifies a best sequence of words that form a sentence, for the given set of semantic roles.

From these work on lyric generation based on melody and word generation we acquired ideas for collecting words and assembling them to a meaningful sentence to form lyrics. However rather than using melody as an input we considered using the scene from a movie as input for generating lyrics which is discussed in the next section.

3 System Overview

The system aims at meaningful lyric generation given a description about the scene. The initial step of the process is to extract the epitome of the scene. The epitome is then analyzed to obtain the emotion conveyed in the description. The identification of the emotion gives us a hint to trace the domain. Based on the domain knowledge we identify the words that lie inside the context of the given scenario. With these words as seed, sentences are generated using a n-gram based approach. The generated lyrics are assured to be pragmatically and semantically correct as it is based on the grammar rules of Tamil. The block diagram of the system is shown in figure 1.

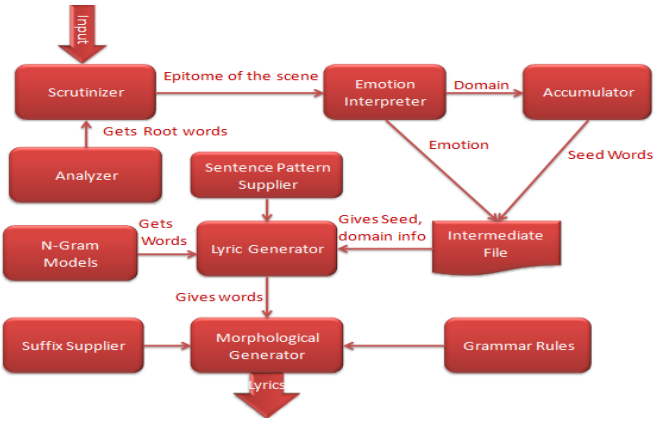


Fig. 1. Architectural Design for Lyric Generation process

The following section describes the individual modules of the proposed lyric generation system.

3.1 Scrutinizer

The scrutinizer is the one that removes the stop words in the scenario. The input to the stop words remover is a text document which has the vernacular description of the scene. The words which are identified as preposition, conjunction, interjection etc gets removed based on a look-up approach. After stop word removal process, this module calls the morphological analyzer to get the root words to form the epitome that will essentially be needed for the lyric generation process.

3.2 Morphological Analyzer

The morphological analyzer gets the words from the scrutinizer and returns the root words for the same. We essentially use an analyzer tool from TACOLA lab of the Department of Computer Science and Engineering, Anna University [8]. However we modified the analyzer to convey the actual meaning of the root word based on the suffix. For example, the words 'kaadhalikkiraan' and 'kaadhalikkavillai' will have the root word identified as 'kaadhali'. However the first word conveyed a positive meaning while the second word a negative meaning. We added this information to the intermediate file in addition to the root which is used in the process of generating lyrics without affecting the meaning.

3.3 Emotion Interpreter

The system gets as input, the epitome of the scenario from the scrutinizer. The system involves finding the emotion of the scene by picking the words in the scene which best convey the actual mood of the situation. The emotions include nostalgia, betrayal, love, friendship, melancholy, sadness and the list is endless. We essentially

maintain a list of words conveying a particular emotion which is used to identify the emotion of a scene. In case of mixture of emotions existing in the same scenario we adopt a prioritization technique to resolve the conflict of identifying the prime emotion. This component is very much crucial to convey the actual mood of the given situation.

3.4 Accumulator

After obtaining the emotion, the domain information is obtained from the emotion interpreter which is fed to the accumulator. The accumulator plays a vital role in identifying the core word which is chosen at random for a domain. This core word is identified after identifying the domain information from the emotion interpreter. Once the core word is obtained the accumulator is triggered to gather more words that occur frequently in this context. These words could be conveying similar meaning and phrases that are closely associated to the core word. These are written in the intermediate file and constitute the set of seed words for the actual lyric generation process.

3.5 Sentence Pattern Supplier

This module has all the available sentence patterns that are commonly used in Tamil poetry. The sentence pattern module gets the seed word from the intermediate file, finds the part of speech of the seed word and supplies a sentence pattern starting with that particular part of speech. The lyric generator then appropriately chooses words to fit these sentence patterns.

3.6 Lyric Generator

This module chooses the appropriate trigram models, based on the domain information which is discussed in the next section. The lyric generation process starts with picking up a seed word from the intermediate file and sentence pattern from the sentence pattern supplier module. With the seed word set as the starting word of the current sentence, this module calls the Lyric Model to get the subsequent words which form the sentence and are sent to the morphological generator. This process iterates for all the seed words available in the intermediate file thus gathering the words which form the lyrics.

3.7 Lyric Model

In this approach we use an n-gram model for generating words. We essentially restrict ‘n’ to 3 and bring the semantic correctness using this Tri-Gram model which we call the lyric model, where the words are chosen relative to the previous two words. For better performance we adopt a technique that makes use of blend of words and co-occurrences. The words which were collected from the existing lyrics were classified based on the domain and parts of speech. The trigram is constructed based on a probabilistic measure to ensure that most suited words form a sentence and the correlation

is maintained. A randomization technique is used to ensure that same words are not chosen always for similar situations. We therefore call this a randomized probabilistic trigram model.

This module receives a sentence pattern and a seed word from the lyric generation module. For example, the sentence pattern is N-V-AD (Noun Verb Adverb) and the seed word is 'Ni' which is a noun. Using the bigram model designed for N-V, we select the verb Vi, based on the probability of the occurrence of a particular verb following that noun and let this probability be Pi. We repeat the process of selecting verbs for this noun with different probabilities say Pi1,Pi2,Pi3.. Once we have chosen the verbs we select the corresponding ADis for each < Ni; Vi > pairs from on the trigram model N-V-AD. Thus we get a set < Ni; Vi; ADi > from which the most appropriate one chosen, that yields the maximum cumulative probability, which is returned to the lyric generator module for further processing.

3.8 Suffix Supplier

After identifying the sequence of root words from the Tri-gram lyric model, the suffix for each word need to be added based on the grammar rules of Tamil language. According to Tamil Grammar we have a set of suffixes which will suit a pair of words from which we choose a random suffix and supply it to the generator module.

3.9 Morphological Generator

The morphological generator [9] gets the trigram of root words constituting a sentence from the lyric generator module and the suffixes from the suffix supplier module. Using these two inputs, the generator, frames the sentence accordingly to the Tamil grammar rules. The form of the suffix varies based on the words, which fall under 16 paradigms. The words are analyzed to find the paradigms to which they belong and based on that suffixes are appended to the words appropriately.

4 Result Analysis

The system has been tested for 100 scenarios from various domains like nostalgia, betrayal, love, friendship, melancholy, sadness. In figure 2 the emotion is identified to be love. Five lines of lyrics have been generated for the given input using the seed words shown in the figure with the help of the lyric model.

அவன் அவளை ரயில் பயணத்தில் சந்திக்கிறான். அவளை பார்த்ததும்
காதலில் விழுகிறான். அவன் தன் காதலை எண்ணி பாடுகிறான்.

Fig. 2. Input for a scenario conveying 'Love'

The input given in figure 2 is analyzed and the emotion is identified using the highlighted words. The emotion identified is 'Love' and the seed words chosen from the 'Love' domain are shown in figure 3.

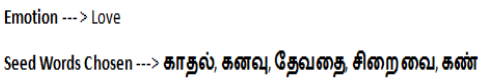


Fig. 3. Emotion and Seed words obtained for the input scene

Figure 4 shows the lyrics generated using the Tri gram lyric model.

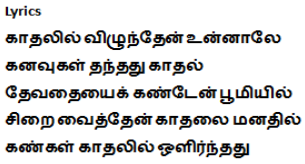


Fig. 4. Lyrics generated for the input scene

Likewise the outputs achieved for melancholy is shown in the subsequent figure 5.

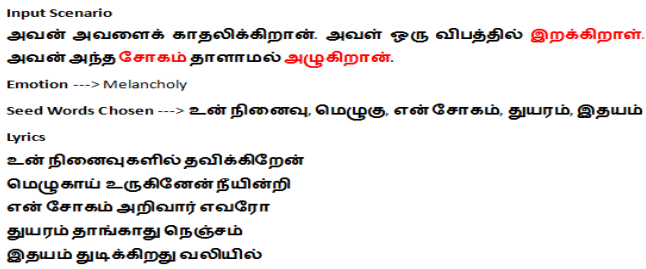


Fig. 5. Output for a scenario conveying ‘Melancholy’

It is obvious from the output that the emotion has been conveyed in the lyrics effectively and the context is also achieved to some extent and the manual analyses of the results are tabulated in Table 1.

On analyzing the outputs achieved so far it can be concluded that on an average, in 88.33% of cases the context described in the scenario is interpreter by the emotion interpreter module and in 74.17% of cases the same emotion is conveyed in the generated lyrics. The reason for the low percentage of emotion conveyed in the generated lyrics is due to the fact that we have chosen only some emotion conveying words from the input and not the overall context of the input. In addition, lyrics are generated based on the sentence pattern selected by the sentence pattern supplier and hence lacks free-word nature of the language. Hence the word set that could be used to convey the context gets narrowed down. This serves as a restriction to generate poetry.

Table 1. Manual analysis of the context preserved in the generated lyrics

S.No	Emotion conveyed in the scene	Percentage of cases the same emotion is identified by the Emotion interpreter	Percentage of cases the same Emotion is conveyed in the lyrics generated
1	Sadness	80%	70%
2	Love	90%	85%
3	Friendship	100%	90%
4	Betrayal	95%	70%
5	Nostalgia	85%	70%
6	Melancholy	80%	60%

When there are conflicting emotions prioritizing them became tedious. In such cases the emotion with which the scenario has ended or the emotion that has occurred more frequently in the scene were chosen for the lyric generation. In few cases this technique led to misinterpretation of the scenario. At times the emotion conveying words in the scenario may be misleading. For example "Aval sandhoshadhdhil azhudhaaL" (She cried with joy) will be interpreted as 'Sadness' because the scenario has ended with the word 'azhudhaaL' – indicating cried which is a misinterpretation.

The system developed was emotion centric .Hence the complete semantics of the situation could not be conveyed. Hence we tried identifying the semantics by picking a core word from the input. This core word was chosen such that it gives the essence of the scene. But identifying it was a tedious task because the essence could be in any of the words used in the input. With the current approach the lyrics generated are not correlated. In addition from the generated lyrics it could be observed the absence of logical coherence between successive lines though they collectively convey an emotion. Hence we tried establishing a link between the lines by randomly taking a word from a previous line that was generated, and take a synonymous word for it and generate a second line. This approach gave adjacent lines that convey the same statement using different set of words. We also tried an alternate strategy to link adjacent lines, by taking antonym of a word from the current sentence and tried to establish coherence. This approach however resulted in having adjacent lines that vary in the domain. Hence we ended up with adjacent lines independent of each other but conveyed the correct emotion.

5 Conclusion and Future Work

The system proposed was successful in identifying the emotion but bringing the complete context of the scenario is what we are concentrating on. Solving this will resolve various drawbacks stated above. As approaches like choosing synonymous phrases to link two lines, taking antonyms to bring a link failed and hence our parallel focus is on linking lines in the lyrics to convey logical coherence.

Poetry features [10] like rhyming, alliteration, simile, etc., are yet to be considered which primary components of any poem are. This would be not just repetition of words but a focus on the Tamil language specific, 'edhugai'(rhyme), 'monai'(alliteration) and 'eyaibu'(rhyme) need to be considered. Vocative words could be added to the lyrics to make it effective. Sentence generation could be done using some alternative mechanism so that the restriction caused by the sentence pattern adoption technique is eradicated that brings out the beauty of the free-word order nature of Tamil language.

Acknowledgments. We kindly thank Tamil Computing Lab, DCSE, Anna University Chennai for providing the morphological analyzer module.

References

[1] Manurung, H.M.: An evolutionary algorithm approach to poetry generation. Institute for Communicating and Collaborative Systems, School of Informatics, University of Edinburgh (2003)
[2] Ananth Ramakrishnan, A., Devi, S.L.: An alternate approach to meaningful lyric generation in Tamil. In: Workshop on Computational Approches to Linguistic Creativity, Los Angeles, California (June 2010)
[3] Suriyah, M., Karky, M., Geetha, T.V., Parthasarathi, R.: Special Indices for LaaLaLaa Lyric Analysis & Generation Framework. Tamil Internet (2011)
[4] http://msvtimes.net/fanclub/andy1.html
[5] http://ta.wikipedia.org/wiki/அசை_(யாப்பிலக்கணம்)
[6] Settles, B.: Computational creativity tools for song writers. In: Workshop on Computational Approches to Linguistic Creativity, Los Angeles, California (June 2010)
[7] Lakshmana Pandian, S., Geetha, T.V.: Semantic Role based Tamil Sentence Generator. In: International Conference on Asian Languages Processing (2009)
[8] Anandan, P., Saravanan, K., Parthasarathi, R., Geetha, T.V.: Morphological analyser for Tamil. In: Proceedings of ICON 2002, RCILTS-Tamil Anna University Chennai (2002)
[9] Anandan, P., Parthasarathi, R., Geetha, T.V.: Morphological analyser for Tamil. In: Proceedings of Tamil Inayam Conference, Malaysia, pp. 46–50 (2001)
[10] Ranganathan, K., Geetha, T.V., Parthasarathi, R., Karky, M.: Lyric Mining: Word, Rhyme & Concept Co-occurrence Analysis. Tamil Internet (2011)

Texture Analysis and Defect Classification for Fabric Images Using Regular Bands and Quadratic Programming

R. Obula Konda Reddy[1], B. Eswara Reddy[2], and E. Keshava Reddy[2]

[1] SSITS, Rayachoti
rkondaredy@gmail.com
[2] JNTUA, Anantapur
eswarcsejntu@gmail.com
keshava_e@rediffmail.com

Abstract. Defect detection is a key problem in quality control for many industrial fields like wallpaper scanning, ceramic flow detection and fabric inspection. For a long time the fabric defects inspection process is still carried out with human visual inspection, and thus, insufficient and costly. Therefore, automatic fabric defect inspection is required to reduce the cost and time waste caused by defects. Many techniques have been developed for detection of defects for fabrics through the years using neural networks, Fourier transform. However, most of the methods mentioned above are mainly designed for unpatterned fabric inspection. In this paper, the work is concentrated on the patterned texture inspection of the fabrics, using regular bands and enhancement of these images using linear quadratic programming.

Keywords: Texture analysis, Regular bands, Defect images, Linear quadratic programming.

1 Introduction

Raising quality requirements for manufactured products has led quality control procedures and inspection procedures to an outstanding place in production processes. Industrial sectors related to materials with periodic textured surface (metallic nets, plastic, paper, films, fabric, etc.) are aware of it and are devoting great efforts to this field. Textile fabric is a representative manufactured product of this kind that presents high quality requirements and challenges for quality control and inspection. Most defects arising in the production process of a textile material are still detected by human inspection. The work of inspectors is very tedious and time consuming. They have to detect small details that can be located in a wide area that is moving through their visual field. The identification rate is about 70%. In addition, the effectiveness of visual inspection decreases quickly with fatigue. The technological development has introduced automation in production processes, increasing their productivity and requiring quality control procedures to be automated too. In general, fabric analysis is performed on the basis of digital images of the fabric. Alternatively, there are some

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 927–934.
springerlink.com

works based on the optical Fourier transform directly obtained from the fabric with optical devices and a laser beam.

Digital image processing techniques have been increasingly applied to textured samples analysis over the last ten years. Several authors have considered defect detection on textile materials. Kang et al. [1, 2] analyzed fabric samples from the images obtained from transmission and reflection of light to determine its interlacing pattern. Tsai and Hu [3] used Fourier transforms of solid plane fabric images as the inputs to an artificial neural network for fabric defect detection. They trained the neural network to identify our types of defects: missing pick, missing end, oil fabric stains and broken fabric. In a recent paper, Hu and Tsai [4] have also used wavelet packet bases and an artificial neural network for the stated goals. Wavelets had been previously applied to fabric analysis by Jasper et al. [5].Escofet et al. [6, 7] have applied Gabor filters (wavelets) to the automatic segmentation of defects on non solid fabric images for a wide variety of interlacing patterns.

Regularity is one of the most important features in many textures, patterned texture – like fabric is built on a repetitive unit of a pattern. Many traditional approaches such as co-occurrence matrices, auto correlation, traditional image subtraction and hash function are based on the concept of periodicity. These approaches have been applied for image retrieval, image synthesis, and defect detection of patterned texture. The above approaches were not impressive in terms of sensitivity to noise and inability to outline the shape of the defect after detection. The main contribution of the present work is to propose a new approach based on the classical statistical method of moving average and standard deviation, which has been applied, patterned texture inspection.

2 Regular Bands for Texture Analysis

Regularity of a patterned texture can be defined as the spatial relationship between the pixel intensities and the repeat distance of repetitive units, spatial relationship means that one pixel in an image should have dependencies and steady changes with its surrounding neighbors on a patterned texture. The repeat distance of a repetitive unit is a measurement that can monitor whether the pattern distorts and overlaps within its placement rule for the construction of the whole image. The structural characteristic is obtained by using the repetitive unit as a convolution filter sliding on the test signal. The numerical values of an abnormal part (defective region) would exceed the normal range of the signal. Therefore, by designing a suitable transformation, any numerical values of the abnormal part is significant enough to be segmented out using thresholding and the shape of any defective region can be outlined [10]. The regular bands method is novel since it only requires the determination of one parameter, the length of period of the repetitive unit in the patterned texture. The RB consists of two subbands, the light regular band (LRB) and the dark regular band (DRB).

The methodology for the fabric defect detection can be followed as

Step 1: Read an image and enhance it with the quadratic programming (QP) as mentioned in (image enhancement using quadratic programming by Tzu-Cheng Jen et al;)

In a typical environment, the dynamic range of a luminance capture device is usually smaller than that of the real scene. Hence, some portions of the captured image may appear either too bright or too dark. To tackle this problem, we try to suppress image

gradients to narrow down the dynamic range while enlarging small image gradients to enhance low-contrast details using the method mentioned in [8]. The local gradients of an image indicate the detail variations of image contents. If we can properly manipulate the magnitudes of local gradients, we may be able to change the perceived visual quality accordingly. In this scenario we considered the 5th order polynomial for the transfer function. Earlier Henry Y.T et al used histogram equalization (HE) for the enhancement of the defective images [10,11].

Step 2: Calculation of the Regular bands

For a particular row in a preprocessed image from the renewed database X=x(i,j)of size MxN.

The light regular band (LRB) is defined as $L_{rn} = \left| u_{rn} - \sigma_{rn} \right| + u_{rn}$ (1)

The dark regular band (DRB) is defined as $D_{rn} = \left| u_{rn} + v_{rn} \right| - u_{rn}$ (2)

Where the moving average is u_{rn} is defined as $n_{rn} = \frac{\sum_{j-r1}^{rn} x_{ij}}{n}$ (3)

Where σ_{rn} is the standard deviation of the n[th] row.

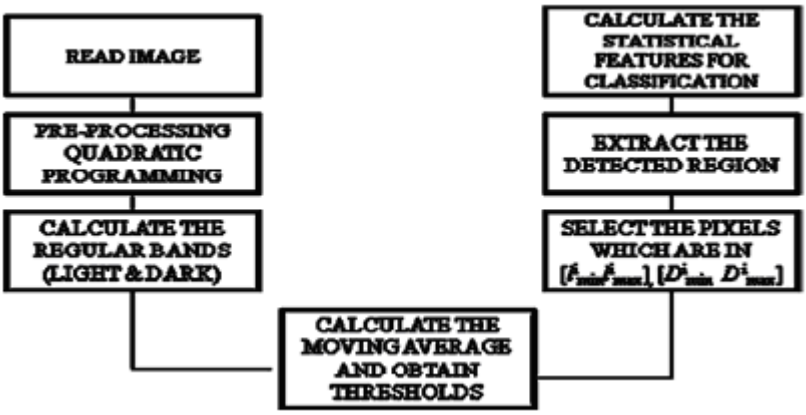


Fig. 1. Block diagram of the proposed approach

Step 3: Obtain the threshold values

Calculate the light regular band and dark regular band on all rows and columns, respectively, for every image. Obtain the upper bound and lower bound of the light regular band $\left[l^{i}{}_{\min} l^{i}{}_{\max} \right]$, and then obtain those of the dark regular band, after the calculations of the regular bands on row $\left[D^{i}_{\min} \; D^{i}{}_{\max} \right]$.so we obtain four threshold values in two sets, $\left[l^{i}{}_{\min} l^{i}{}_{\max} \right], \left[D^{i}{}_{\min} D^{i}{}_{\max} \right]$.[9]

3 Experimental Analysis

In this paper for the training, box patterned and star patterned fabric defect free images are taken. As for testing, the same patterned images with defects like broken end, hole ,netting multiple, thick end and thin end. For training 50 defect free images against 5 defect images for each defect (like broken end, hole etc) of each pattern are taken. A statistical analysis has been made for the above said images which are shown in tables 1&2.

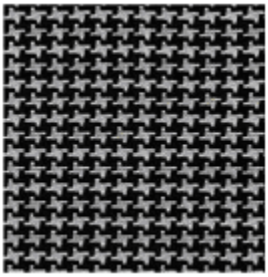
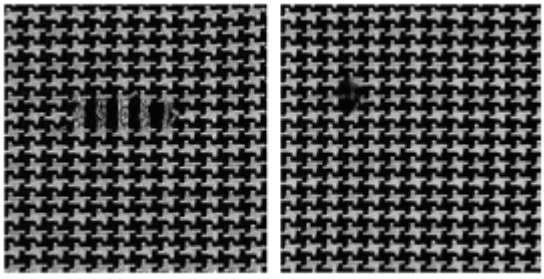
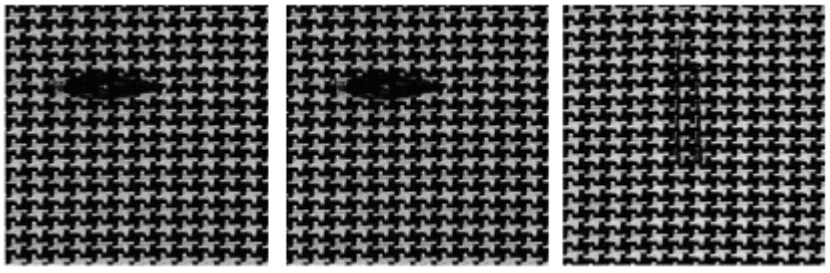


Fig. 2. Star Patterned defect free image



a) b)



c) d) e)

Fig. 3. Star Patterned defect images. a) Broken end b) Hole c) Netting Multiple d) Thick Bar e) Thin Bar

The fig 2, star patterned 25 defect free fabric images are taken for training and for the testing 5 images of each defect are taken.

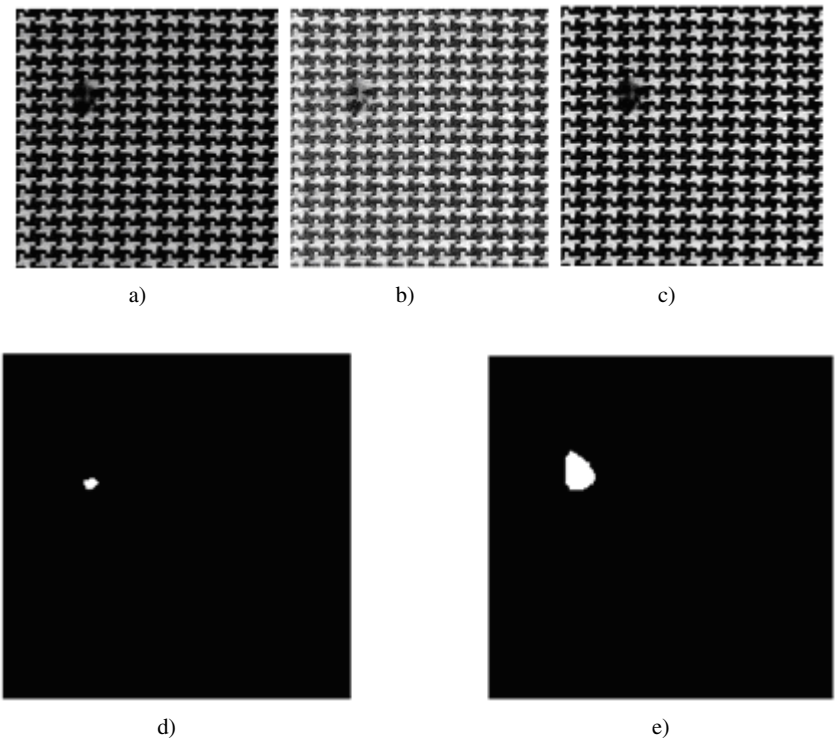


Fig. 4. a) Original image b) Histogram equalized c) Enhanced using quadratic programming d) Detected with RB &HE e)Detected with Proposed approach

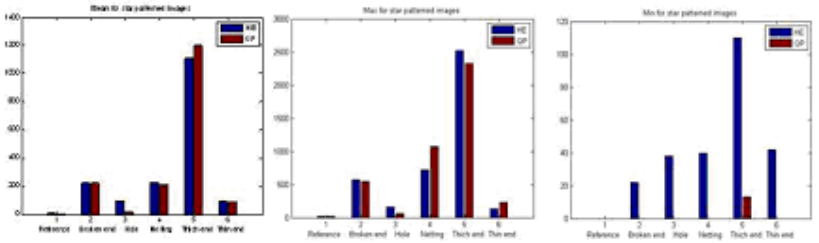
Table 1. For star pattered images for n=25 with HE

S.no	Type of image	Mean	Max	Min	Mean (%)	Max (%)	Min (%)
1	Refference	4.32	23	0	0.007	0.04	0
2	Broken End	224.8	572	22	0.39	0.9	0.038
3	Hole	96.6	160	38	0.17	0.27	0.065
4	Netting Multiple	227.8	720	40	0.4	1.25	0.069
5	Thick end	1111	2526	110	1.93	4.39	0.19
6	Thin end	94	138	42	0.16	0.24	0.072

Table 2. For Star pattered images for n=25 with QP

S.no	Type of image	Mean	Max	Min	Mean (%)	Max (%)	Min (%)
1	Refference	3.2	23	0	0.0056	0.04	0
2	Broken End	221.8	547	0	0.39	0.949	0
3	Hole	15.6	69	0	0.027	0.119	0
4	Netting Multiple	212.2	1078	0	0.37	1.8	0
5	Thick end	1201	2326	13	2.09	4.04	0.02
6	Thin end	90	233	0	0.16	0.4	0

A defective image is defined as an image with a considerable number of pixels exceeding the normal range defined by the regular bands. They appear as white pixels after thresholding in the RB method. So, a final threshold image is determined to be defective if it exceeds a certain amount of white pixels. Experiments were conducted using Mat lab R2006a on Pentium dual core Processor, 2 GB RAM. An experiment is conducted with the 50 said images for the both defective free and defective. These images are been tested with the algorithm and the statistical parameters line mean, max, min are evaluated which are tabulated .The tables include different types of images which are compared against the reference defect free images. These tables contain the values of the proposed and the conventional histogram based approach. From all these tabulated statistical results we conclude that the proposed approach leads to a very better way detect the defects.



a) b) c)

Fig. 5. Comparative results with star patterned images a) Mean b) Max c) Min

4 Conclusion

Using quadratic programming approach for the image enhancement and applying the regular bands for the defect detection proves to be effective and the overall detection percent is been increased . The RB method enhances the defective regions through the calculations of moving averages and standard deviations. In this analysis we have used 25 defect free reference images and considered the different types of the defective images like broken end, Hole, Thick end, thin end and Netting multiple. For all

these images we have considered the star patterned and box patterned images and the approach proved to be success for detection of about 96%.Using the QP(quadratic programming) it took us 6.46 seconds where as in the earlier method using HE(histogram equalization) it took 1.6 seconds with the above mentioned system configuration, this becomes a limitation for the current approach where as this approach proves to be excellent in the defect detection which makes it more advantageous than the earlier. Using quadratic programming approach for the image enhancement and applying the regular bands for the defect detection proves to be effective and the overall detection percent is been increased . The RB method enhances the defective regions through the calculations of moving averages and standard deviations. In this analysis we have used 25 defect free reference images and considered the different types of the defective images like broken end, Hole, Thick end, thin end and Netting multiple. For all these images we have considered the star patterned and box patterned images and the approach proved to be success for detection of about 96%.Using the QP(quadratic programming) it took us 6.46 seconds where as in the earlier method using HE(histogram equalization) it took 1.6 seconds with the above mentioned system configuration, this becomes a limitation for the current approach where as this approach proves to be excellent in the defect detection which makes it more advantageous than the earlier.

References

1. Amet, A.L., Ertuzun, A., Ercil, A.: An efficient method for texture defect detection: Subband domain co-occurrence matrices. Image Vis. Comput. 18, 543–553 (2000)
2. Wang, C., Ye, Z.: Brightness preserving histogram equalization with maximum entropy: A variational perspective. IEEE Trans. on Consumer Electronics 51(4), 1326–1334 (2005)
3. Escofet, J., Navarro, R., Millán, M.S., Pladellorens, J.: Detection of local defects in textile webs using Gabor filters. In: Refregier, P. (ed.) Vision Systems: New Image Processing Techniques. Proceedings SPIE, vol. 2785, pp. 163–170 (1996)
4. Escofet, J., Navarro, R., Millan, M.S., Pladellorens, J.: Detection of local defects in textile webs using Gabor filters. Opt. Eng. 37(8), 2297–2307 (1998)
5. Hu, M.C., Tsai, I.S.: Fabric Inspection Based on best Wavelet Packet Bases. Textile Res. J. 70(8), 662–670 (2000)
6. Ngan, H.Y.T., Pang, G.K.H.: Novel method for patterned fabric inspection using Bollinger bands. Optical Engineering, Society of Photo-Optical Engineers (August 2006)
7. Ngan, H.Y.T., Pang, G.K.H.: Regularity Analysis for Patterned Texture Inspection. IEEE 6(1), 131–144 (2009)
8. Ngan, H.Y.T., Pang, G.K.H., Yung, S.P., Ng, M.K.: Wavelet based methods on patterned fabric defect detection. Pattern Recognit. 38(4), 559–576 (2005)
9. Jasper, W.J., Garnier, S.J., Potlapalli, H.: Texture characterization and defect detection using adaptive wavelets. Opt. Eng. 35(11), 3140–3149 (1996)
10. Kang, T.J., et al.: Automatic recognition of Fabric Weave Patterns by Digital Image Analysis. Textile Res. J. 69(2), 77–83 (1999)

11. Kang, T.J., et al.: Automatic Structure Analysis and Objective Evaluation of Woven Fabric Using Image Analysis. Textile Res. J. 71(3), 261–270 (2001)
12. Chin, R.T., Harlow, C.A.: Automated visual inspection: A survey. IEEE Trans. Pattern Anal. Machine Intell. PAMI-4(6), 557–573 (1982)
13. Tsai, I.S., Hu, M.C.: Automatic Inspection of Fabric Defects Using an Artificial Neural Network Technique. Textile Res. J. 66(7), 474–482 (1996)
14. Jen, T.-C., Wang, S.-J.: Image Enhancement Based on Quadratic Programming. In: 15th IEEE Conference on ICIP 2008, pp. 3164–3167 (2008)

Exploring the Pattern of Customer Purchase with Web Usage Mining

Paresh Tanna[1] and Yogesh Ghodasara[2]

[1] School of Engineering – MCA Department, R.K. University, Rajkot
[2] College of Agricultural InfoTech, Anand Agricultural University, Anand
paresh.rkcet@gmail.com, yrghodasara77@yahoo.co.uk

Abstract. The purpose of this paper is to do an analysis of the sample / raw data to obtain a meaningful interpretation using some of the data mining algorithms like a vector quantization based clustering and then an 'Apriori' based Association rule mining algorithm. Web session clustering plays a key role to classify web visitors on the basis of user click history and similarity measure. An important application of chronological mining techniques is web usage mining, for mining web log accesses, where the sequences of web page accesses made by different web users over a period of time, through a server, are recorded. The experiment will be conducted base on the idea of Apriori algorithm along with VQ based clustering, which first stores the original web access sequence database for storing non-sequential data. The experimental result will be given with analysis on further refinement. This is aimed at a meaningful segregation of the various customers based on their RFM values, as well to find out relationships and patterns among the purchases made by the customer, over several transactions.

Keywords: Apriori, VQ, Chronological mining, Web Usage, Data Mining, RFM.

1 Literature Review

1.1 The Need for Exploring the Pattern with Web Mining

It is well know that users' online interactions with the website are recorded in server web log files that serve as a valuable pool of information. By applying the data mining techniques on web log file, we obtain good insights about the users' behaviors. We can analyze the web log files for various aspects of website enhancements[6]. Furthermore, proper analysis of web log unleashes useful information for webmaster or administrator for numerous advantages such as web personalization, website schema modification, user surfing behaviors, website structure modification and we can tackle the issue of web server performance as well[13]. Extraction process information rapidly from the log files of a web site can be used to identify patterns of access (usage patterns) and profile their users [4, 5]. Often associated as clickstream data because each insert according to the lawyer-click the mouse button[3]. The abundance of customer information enables marketers to take advantage of individual-level

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 935–941.
springerlink.com

purchase models for direct marketing and targeting decisions[7]. The major customer values or characteristics that are used to measure purchase behavior of customers include Regency, Frequency, and Monetary values (RFM)[11].

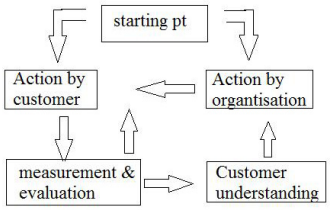


Fig. 1. The basic CRM cycle. [1]

1.2 Exploring the Pattern with Custom Algorithm

The concept of chronological pattern mining is as follows: given a database of successions where each succession is an ordered list of user access based on access time and each access consists of a collection of useful information, then searched all the access pattern with minimum support defined by the user, where support a number of database sequences that contain the pattern[6].

Algorithm: Chronological Pattern Algorithm
INPUT:
D = (C1, C2 ,..., Dk) / / Database of the session

OUTPUT: Chronological Pattern
Chronological Pattern Algorithm:
D = D sorting on User ID and time of reference on the first page in each session.
Find L in D; L = APriori (D) Find a maximal reference sequence of the L;

Above Algorithm [6] shows the steps needed to find chronological patterns to sort data. After doing the sorting, the next steps are the same as that performed by the algorithm a priori. Sorting steps to make a series of real users, which is a complete reference sequence of one user (inter-transaction). By following segmentation, customer buying pattern can be further improved for decision making.

2 The Customer Segmentation Approach

2.1 Clustering and Customer Segmentation

Customer segmentation is one of the most important area of knowledge-based marketing. In this paper we consider a clustering algorithm, which is based on a Vector Quantization based algorithm, and can be effectively used to automatically assign existing or new arriving customers into the respective clusters.

2.2 Quantization Based Clustering Algorithm for Enhancement in Pattern Discovery

It is an efficient algorithm designed by Linde, Buzo and Gray for the design of good block or vector quantizers with quite general distortion measurements is developed for use on either known probabilistic source descriptions or on a long training sequence of data, incorporated herein by reference, [1]. An N-level k-dimensional 'quantizer' is a mapping, q ; that assigns to each input vector, x = (x_0, -, X_{k-1}), a reproduction vector, $\hat{x}$ = q(x), drawn from a finite reproduction alphabet, A = bi; i = 1, -, N} [1]. The level N describes the number of times the division of the codebook occurs. The quantizer is completely described by the reproduction alphabet (or codebook) . Such quantizers are also called block quantizers, vector quantizers, and block source codes [1].

$$d(x, x\hat{}) = \sum_{i=0}^{k-1} |x_i - x\hat{}_i| \quad (1)$$

ALGORITHM : VQ :

1. Initialization: Given N = number of levels, a distortion threshold ε ≥ 0, and an initial N-level reproduction alphabet A_0, and a distribution F. Set m = 0 and $D_{-1} = \infty$.
2. Given A_m ={ y_i ; i = 1, …, N}, find its minimum distortion partition P(A_m) = {S_i; i = 1... N}: x S_i if d(x,y_i) d(x, y_j) for all j. Compute the resulting average distortion, D_m =D({A_m, P(A_m)}) = E min$_{EAm}$ d(X,y).
3. If (D_{m-1} - D_m)/D_m < , halt with A, and P(A_m) describing final quantizer. Otherwise continue.
4. Find the optimal reproduction alphabet d(P(A_m)) = {x^(s_i); i = 1,…, N} for P(A_m). set A_m x^(P(A_m)). Replace m by m + 1 and go to 1.

Earlier this algorithm was mainly used for image compression and other related works. But, I found it useful for my clustering approach.

3 The Association Rules Based Approach for Customer Purchase Predictions

3.1 A Description of the Association Rules Mining Model Proposed Here

'Apriori' is the most basic algorithm for learning association rules. Apriori is designed to operate on databases containing various kinds of transactions (for example, collections of items bought by customers, or details of a website surfing)[9]. The algorithm attempts to find subsets which are common to at least a minimum number K of the itemsets [12].

4 Implementation and Results

4.1 Sample Dataset and the Transformation of Data

In this study sample dataset from the sample database "Nwind.mdb" has been taken, which is a Microsoft Access sample database file, and used the "OrderDeatils" table, which has about 2155 entries, and can be sufficiently large enough for the requirement of the analysis.

OrderID	Product	Unit Pric	Quanti	Discou
10248	Queso Cabrales	$14.00	12	0%
10248	Singaporean Hokkien Fried Mee	$9.80	10	0%
10248	Mozzarella di Giovanni	$34.80	5	0%
10249	Tofu	$18.60	9	0%
10249	Manjimup Dried Apples	$42.40	40	0%
10250	Jack's New England Clam Chowder	$7.70	10	0%
10250	Manjimup Dried Apples	$42.40	35	15%

Fig. 2. The sample dataset for the Association rule mining approach (taken from the table 'Orderdetails' "nwind.mdb" MS access database)

The implementation of the customer segmentation used a different table for the VQ approach.

OrderID	Customer	Freight	Ship Name
10248	Vins et alcools Chevalier	$32.38	Vins et alcools Chevalier
10249	Toms Spezialitäten	$11.61	Toms Spezialitäten
10250	Hanari Carnes	$65.83	Hanari Carnes
10251	Victuailles en stock	$41.34	Victuailles en stock

Fig. 3. Sample dataset for the VQ based customer clustering approach. (taken from the Orders table of 'Nwind.mdb' MS access database)

4.2 Choice of Programming Language and Environment

The study has taken Java as the choice of programming language and Netbeans IDE as the programming environment. The nature of the algorithms that has been implemented, is to work upon the datasets, which can be viewed as objects.

4.3 Study of the VQ Based Clustering Algorithm. Results Found and Discussion

From the simulation of the algorithm, which had the data of 91 customers of a company, the algorithm calculated the no. of customers in a certain price range of expenditure, using the monetary values obtained from the freight values in the table.

These were the findings or observations:

```
Spending of customers (the data retrieved from the database, shown in Integer for lucidity):

1357 1394 822 493 2755 364 1403 983 448 1017

1353 5605 2134 58 125 724 821 367 1001 194

1259 6205 3 862 327 1678 623 67 558 469

475 322 1559 64 632 319 281 187 274 432
```

Fig. 4. The initial input data for VQ approach

```
Using the clustering algorithm, I found the following results :

The no. of people spending less than 100 :18

The no. of people spending between 100 and 500 :37

The no. of people spending between 500 and 1000 :18

The no. of people spending between 1000 and 1500 :9

The no. of people spending more than 1500 :8
```

Fig. 5(a). The output observed for e = 0.002 for the above VQ approach

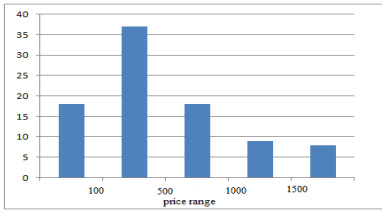


Fig. 5(b). The output shown in a graphical form. X-axis represents the price range and Y-axis represents the number of customers.

4.4 Study of the Association Rule Based Customer Behavior Mining Approach. Results and Discussions

Initially the study applied the Association rule based mining algorithm for 1-itemset, for which I obtained the following output.

```
Occurence of item 7 : 13.0
Occurence of item 8 : 5.0
Occurence of item 9 : 33.0
Occurence of item 10 : 38.0
Occurence of item 11 : 14.0
Occurence of item 12 : 40.0
Occurence of item 13 : 22.0
```

Fig. 6. The occurrences for 1- itemset

Then the study moved on to a 2-itemset approach and devised the rules for 2 – itemsets based on the same datasets. The output had hundreds of rules which then had to be pruned down on the basis of the "coverage values". Initially I took coverage value as 0.2, for which the following output was shown.

```
if 2 then 55
if 2 then 59
if 2 then 60
if 2 then 61
if 2 then 68
if 2 then 70
.....
```

Fig. 7. The final rules obtained after pruning with coverage = 0.2

```
The final list of rules :
if 8 then 4
if 8 then 30
if 8 then 60
if 8 then 75
if 14 then 6
if 14 then 19
```

Fig. 8. The final rules obtained after pruning with coverage = 0.5

5 Conclusions

From the results of experimental analysis, we can conclude that the greater the number of combinations is produced, the less likely the number of users who perform a combination of these, while the fewer number of combinations generated then it is likely the number of users who perform a combination of these will be even greater. Using Chronological Pattern Mining Web Logs can further explore the pattern of habits of users who access the website pages on the Internet that do a search ordered patterns that may be performed by the user in accessing the website addresses on the internet. We applied Chronological Pattern Algorithm and the Apriory and VQ based clustering algorithm then it was observed that the two approaches show varied resulting data that can be interpreted in different ways: VQ approach can be basically used to segment customers, according to any of the RFM values, or all of them together. It needs the initial vectors as its input for it to start creating the clusters. Also, the increase from a 1-item to a 2-itemset and then onto a 3-itemset, results in a 10-folds increase in the computation times of the algorithms.

References

1. Linde, Y., Buzo, A., Gray, R.M.: An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications, com-28(1), 84–86 (1980)
2. Zakrzewska, D., Murlewski, J.: Clustering Algorithms for Bank Cus tomer Segmentation. In: 5th International Conference on Intelligent Systems Design and Applications, pp. 1–2 (2005)
3. Berendt, B., Spiliopoulou, M.: Analyzing navigation behaviour in web sites integrating multiple information systems. VLDB Journal, Special Issue on Databases and the Web 9(1), 56–75 (2000)
4. Borges, J.A., Levene, M.: Data Mining of User Navigation Patterns. In: Masand, B., Spiliopoulou, M. (eds.) WebKDD 1999. LNCS (LNAI), vol. 1836, pp. 92–112. Springer, Heidelberg (2000)
5. Gaol, F.L., Widjaja, B.: CLS and CLS Close: Scalable Mining of Frequent Graph Structure Patterns in Large Graph Databases. In: Proceedings of International Online Conference on Systems, Computing Sciences and Software EngineeringInstitute of Electrical & Electronics Engineers, Bridgeport, USA, pp. 251–258. IEEE (December 2007)
6. Gaol, F.L.: Exploring The Pattern of Habits of Users Using Web Log Squential Pattern. In: 2010 Second International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp. 161–163 (2010) 978-0-7695- 4269-0/10 $26.00 ©, IEEE
7. Al-Mudimigh, A., Saleem, F., Ullah, Z.: Department of Information System: Efficient implementation of data mining: improve customer's behavior, pp. 7–10. IEEE (2009)
8. Ha, S.H., Park, S.C., Bae, S.M.: Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case. Computers & Industrial Engineering 43, 801–820, 801–806 (2002)
9. Suh, E., Lim, S., Hwang, H., Kim, S.: A prediction model for the purchase probability of anonymous customers to support real time web marketing: a case study. Expert Systems with Applications 27, 245–250 (2004)

10. Chen, M.-C., Chang, H.-H., Chiu, A.-L.: Mining changes in customer behavior in retail marketing. Expert Systems with Applications 28, 773–776 (2005)
11. Thirumalai, S., Sinha, K.K.: Customer satisfaction with order fulfill ment in retail supply chains: implications of product type in electronic B2C transactions. Journal of Operations Management 23, 291–296 (2005)
12. Witten, I.H., Frank, E.: Data Mining: Practical machime learning tools and techniques, pp. 112–118, 136–139. Morgan Kaufmann Publishers, San Francisco (2005)
13. Hussain, T., Asghar, S.: A Hierarchical Cluster Based Preprocess ing Methodology for Web Usage Mining. In: 2010 6th International Conference on Advanced Information Management and Service (IMS), pp. 472–477 (2010) E-ISBN: 978-89-88678-32-9 Print ISBN: 978-1-4244-8599-4

Information Fusion from Mammogram and Ultrasound Images for Better Classification of Breast Mass

Minavathi[1], Murali S.[2], and M.S. Dinesh[3]

[1] PES College of Engineering, Mandya
minavati@yahoo.com
[2] MIT, Mysore
[3] PET Research Center, Mandya

Abstract. Various medical modalities are used in all phases of cancer detection. Information extracted from these modalities reveals morphological, metabolic and functional information of tissues. Integrating this information in a meaningful way assists in clinical decision making. Sometimes using multimodal techniques supply complementary information for improved therapy planning. Proposed investigation is on classification of breast mass as benign or malignant for early detection of breast cancer using mammograms and Ultrasound modalities. The proposed approach is based on the fusion of information from two modalities at image feature level with different normalization techniques to improve the performance of breast mass classification. Gabor filters are used to retrieve texture features from mammograms, shape and structural features are retrieved from ultrasound images. Training of classifier is done using Support vector machine (SVM) classifiers to classify masses. Receiver operating characteristic curves (ROC) are used to evaluate the performance. Our method was validated on 20 set of images. Where each set consists of one mammogram and one ultrasound image of a same person out of which 9 sets were malignant and 11 were benign. SVM classifiers achieved 95.6% sensitivity in classifying the masses using the features retrieved from two modalities.

Keywords: Mammogram, Ultrasound, Dual modality, SVM, Feature level fusion, Z-score Normalization.

1 Introduction

Breast cancer is the most common, life-threatening cancer which has been reported to have the highest mortality rates of any women's cancer. It is the second leading cause of cancer deaths among women in United States and it is the leading cause of cancer deaths among women in the 40 – 55 age groups. Approximately 182,000 new cases of breast cancer are diagnosed and 46,000 women die of breast cancer each year in the United States. In 2009, about 40,610 women died from breast cancer in the United States [17, 19]. According to the recent statistics, one out of nine women will develop breast cancer during her lifetime. There is no effective way to prevent the occurrence of breast cancer. Therefore, early detection is the first crucial step towards treating breast cancer.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 943–953.
springerlink.com

Some of the important signs of breast cancer radiologists normally look for are: spiculated masses, micro calcifications, architectural distortions and bilateral asymmetry. Spiculated masses are characterized by radiating lines or spicules from a central mass of tissue. Spiculated masses carry a much higher risk of malignancy than calcifications or other types of masses [12]. Architectural distortion is the third most common finding of breast cancer According to our survey 81% of spiculated mass and 48-60% of AD is malignant and 12-45% of cancers missed in screening are spiculated masses with AD. The detection sensitivity of the current CAD systems for Spiculated masses with AD is low and there is a pressing need for improvements in their detection. Breast cancer does not always produce a visible mass, but it frequently disrupts the normal tissues in which it develops. This distortion of architecture may be the only visible evidence of the malignant process. The probability of malignancy increases as a lesion becomes more irregular in shape [3, 13].

Mammography and ultrasonography are currently the most sensitive noninvasive modalities for detecting breast cancer. A panel report issued from Institute of Medicine and National research council of National Academics says that Mammography though useful wasn't always enough and health practitioners needed to investigate other complementary screening methods like ultrasound [15]. It also says that mammography depicts about three to four cancers per 1000 women. But in women with dense breasts ultrasound depicts another three cancers per 1000 women. In addition, mammography produces a high false positive rate, and only about 525 of 1800 lesions that were sent to biopsy are malignant [15,17]. Mammography has limitations in cancer detection in the dense breast tissue of young patients. Most cancers arise in dense tissue, so lesion detection for women in this higher risk category is particularly challenging. The breast tissue of younger women tends to be dense and full of milk glands, making cancer detection with mammography problematic. In mammograms, glandular tissues look dense and white, much like cancerous tumor. The reasons for the high miss rate and low specificity in mammography are, low conspicuity of mammographic lesions, noisy nature of the images, overlying and underlying structures that obscure features of the mammographic images [11]. The cancers found on ultrasound are almost all small invasive cancers that have not yet spread to the lymph nodes and therefore have good prognoses. Ultrasonography is proved to be more effective for women younger than 35 years of age and is an important adjunct to mammography [4]. Literature suggests that denser the breast parenchyma, higher will be the accuracy of malignant tumors in ultrasound images. However ultrasound itself has some limitations: low resolution, low contrast, blurry edges and speckle noise. So it is very difficult for a radiologist to read and interpret an ultrasound image. Though Mammography and ultrasonography are currently the most sensitive noninvasive modalities for detecting breast cancer, they have their own limitations. The above argument justifies that features retrieved from one modality are not sufficient to detect the abnormalities of breast cancer in early stages. Integration of multimodalities has been widely used for generating more diagnostic and clinical values in medical imaging [9]. Proper multimodality fusion techniques need to be employed. Thus our proposed work concentrates on designing image processing algorithms to extract features from dual modalities (ultrasound and mammogram) and to fuse them to improve the performance of classification. Multimodal techniques supply complementary

information for improved therapy planning. As early detection of cancer is probably the major contributor to a reduction in mortality for certain cancers, images guided and targeted minimally invasive therapy has the promise to improve the outcome and reduce collateral effects.

2 Review of Related Works

Studies showing the advantages of dual modality and feature level fusion have been appeared in the literature. Fabio Rolia et al. [6] have proposed a serial scheme on well-known benchmark face datasets and fingerprint dataset which combines two serially matchers at which the performance of the serial model is higher than parallel. Brunelli and Falavigna [18] have experimented using tanh method for normalization and weighted geometric average for fusion of voice and face biometrics. Hierarchical combination scheme is also used by them for a multimodal identification system. Kittler et al. [10] has used various fusion techniques on face and voice biometrics. He has experimented on sum, product, minimum, median, and maximum rules and have found that the sum rule outperformed others. He finally concluded that the sum rule is not significantly affected by the probability estimation errors and this explains its superiority.

Hassan and A. S. Mohamed [8] presented a study of multimodal palm veins and signature identification by extracting the features of both modalities using morphological operations and Scale Invariant Features Transform (SIFT) algorithm. They have used simple sum rule to achieve feature level fusion for both modalities. They have applied discrete cosine transform (DCT) algorithm to reduce the feature vectors dimensionalities of feature extraction techniques. Finally they have stated that SIFT algorithm is more accurate and does not need more preprocessing steps to identify people.

Han-ling and Fan [7] proposed a hybrid optimization algorithm to deal with multimodal (CT and MRI) medical images. They used mutual information as a similarity measure and proved that subvoxel accuracy can be achieved for an efficient image registration and can avoid getting into local optimum. Andrzej Krol and Ioana [1] have investigated an approach for co-registration of PET images with MR images in image fusion level. They proved that it is an alternative to surgical breast biopsy.

Francis and Thomas [5] worked on fusion of data from mammography, ultrasound and non invasive infrared imaging modalities to improve early diagnosis. They concluded that data fusion will add early back into early detection of breast cancer.

It has been understood by the literature that, the works in medical image processing generally are not trying to perform fusion in feature level. But some of the work is been carried out in data level and image level fusion. In our proposed work we are introducing a new approach of fusing features of mammograms and ultrasound in order to improve the diagnosis by giving second opinion to the radiologist that hopefully reduce the rate of biopsy.

3 Materials and Methods

In this paper we have used Ultrasound and mammogram images of same person. Data set was created by collecting and getting the ground truth marked images from expert radiologists trained with those kinds of images. All images in our dataset contained only one abnormality. For each image, a rectangular region of interest (ROI) including mass and the area around it were determined by an experienced radiologist. The radiologist also depicted mass contours and has classified them as regular or irregular. Both modalities will output a collection of features. The fusion process fuses this collection of features into a single feature set.

Mammography whether film or digital is a best choice of screening for women who are less than age 40. But for younger women and dense breast women it is not an adequate choice. Though Ultrasound is a commonly used diagnostic tool, it is not FDA (Food and Drug Administration) approved for screening. Fusion of information can be done in feature level, data level or decision level. Feature level methods combine various features into single fused one which can be later used by conventional classifier. On the other hand decision level fusion combines several classifiers to make a stronger classifier which is also called as post classification fusion.

The purpose of our work is to demonstrate the fusion of features from these two modalities that is not way off in future, but one that can be utilized today. Multimodality is not a new concept but the inclusion of feature level fusion to mammogram and ultrasound modalities for classification of mass is rare. Our objective is to show how feature level fusion in multimodality helps in detecting breast cancer. The proposed dual modality system combines the structural and functional behavioural trait of mammogram and ultrasound modalities as shown in Fig.1. We have extracted texture features and directional features from mammograms using Gabor filters, gradient orientation and phase portraits. Using angle of curvature method and intensity based method functional features like acoustic shadow and shape features are retrieved from ultrasound. As Mammogram and Ultrasound are independent modalities, we need to normalize the features. Feature Normalization is done using Z-score for feature level fusion. Support vector machine (SVM) classifiers are used to classify the fused features.

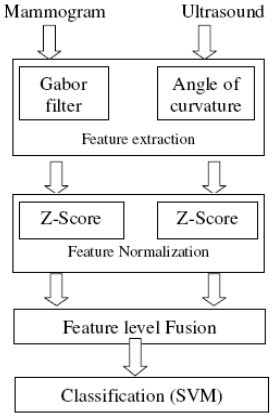


Fig. 1. Proposed Dual modality system

3.1 Feature Extraction from Mammograms

In mammograms we are retrieving the features mainly related to AD with spiculated mass. Architectural distortion (AD) with spiculation is very important finding for the early detection of breast cancer. Such distortions can be classified as spiculation, retraction, and distortion. AD with Spiculated masses have stellate appearance as shown in Fig.2(b). The size of the lesion ranges from few millimeters to centimeters. To extract features from mammograms we have developed a new computerised method to retrieve texture features. The method is based on the distribution of the mammary gland which is approximated to linear structures. In normal breast, the direction of the distribution tends toward the nipple and in an abnormal breast it tends toward suspected areas [13]. Based on the linear structure of mammary gland, we evaluate local structure of mammary ducts. We then focus on characterizing the degree of concentration of mammary ducts to a specific point. Additional features like denseness texture feature, standard deviation, entropy and homogeneity are also considered.

In the present work we have used Gabor filters as line detectors. In order to extract the texture orientation at each pixel of a mammogram, we filter the mammogram with a bank of Gabor filters of different orientations [13, 14].

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2} + \frac{y'^2}{\sigma_y^2}\right)\right] \tag{1}$$

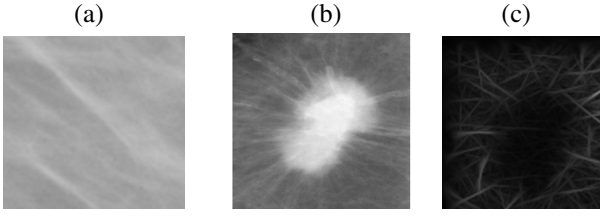


Fig. 2. (a) Mammogram ROI of Normal breast structure, (b) Mammogram ROI of AD with Spiculated mass breast structure, (c) Gabor filtered magnitude image of mammogram ROI of AD with Spiculated mass

Kernels at other angles can be obtained by rotating this kernel. We have used 180 kernels with angles spaced evenly. A Gabor filter can provide good detection accuracy for linear patterns with thickness up to 0.8 mm [13, 14]. It is desirable to reduce the influence of the low-frequency components of the mammographic image in the orientation field magnitude, since the low-frequency components are not related to the presence of oriented structures in the image. Therefore, the mammographic image is high pass filtered prior to the extraction of the orientation field. Gaussian filters are used for low pass filtering. The magnitude image which shows the line structures of the mass after Gabor filtering is shown in Fig. 2(c). In order to extract orientation at each pixel of magnitude image obtained after Gabor filtering, gradient based orientation extraction is used. Here gradient vectors are calculated by taking partial derivatives of image intensity at each pixel in Cartesian coordinates.

We proceed by searching for node or star like stellate structures in the image. Phase Planes provide an analytical tool to study systems of first-order differential equations. We have drawn node and star maps using phase planes (PPlanes) and it is compared with the gradient orientation image. As we need some measure of distance between two orientation fields we have applied flow field analysis using distance measure (nonlinear least squares) [13, 2] and the degree of distortion is calculated. The presence of stellate appearance (strong node or star) point indicates the sites of AD with spiculated mass. The features that constitute image texture extracted above are not sufficient for classification, thus some other characteristic features that are concerned with spatial organization of gray level primitives are considered. Additional features that we have extracted from gradient orientation image are: denseness texture feature, standard deviation, entropy and homogeneity.

3.2 Feature Extraction from Ultrasound

In mammograms we are retrieving the features mainly related to AD with spiculated mass. Architectural distortion (AD) with spiculation is very important finding for the early detection of breast cancer. Su Ultrasound (US) is an important adjunct to mammography in breast cancer detection as it doubles the rate of detection in dense breasts and also does dynamic analysis of moving structures in breast. Architectural distortions and spiculated masses with Architectural distortions on mammography are considered to be one of the most indicators of breast cancer, where distortion refers to presence of radiating structure concentrated at a point. But recently AD and AD with spiculated mass has been detected via ultrasonography also.

As the spatial resolution of ultrasound is not good detection of AD even in the absence of definite mass is difficult. Thus we are concentrating on retrieving the features of AD with spiculated mass in our work. Ultrasound images are first preprocessed using Gaussian smoothing to remove additive noise and anisotropic diffusion filters to remove multiplicative noise (speckle noise). For segmentation active contour method is used to extract a closed contour of filtered image which is the boundary of the spiculated mass. During feature extraction spiculations which make breast mass unstructured or irregular are marked by measuring the angle of curvature of each pixel at the boundary of mass. To classify the breast mass we have used the structure of mass in accordance with spiculations and elliptical shape.

Several features might be derived from an image. But not all of the features are suitable for classification. Too many irrelevant features not only make the classifier complicated, but also will reduce the accuracy of the classification. The most important issue is to select features that are able to represent the characteristics of spiculated masses in the breast ultrasound images. Spiculations are the small needle like structures found in malignant mass which shows uncontrollable multiplication of breast cells. These spiculations will make the breast masses unstructured and irregular [15].

The first feature retrieved is the spiculation feature of mass by finding angle of curvature at each pixel of contour. Most benign masses tend to be wider and roughly ellipse. Thus as a second feature we consider shape of the mass by fitting the contour to ellipse. Based on these features the spiculated malignant mass can be significantly discriminated from the benign masses by the classifier. In breast ultrasound images, spiculations and angular margins are the significant characteristics. Spiculations

produce the higher positive predictive value of malignancy. Also, the hyper-echogenicity, well-circumscribed lobulation, ellipsoid shape and a thin capsule are the significant characteristics of benign masses in breast ultrasound images [15, 21].

The angle of curvature of every pixel at that boundary of the mass is considered. At every pixel the angle of curvature is found by projecting lines from that pixel to some appropriate pixels and the angle between the lines are found and is as shown in Fig 3. Spiculated regions will be having lesser angle of curvature and thus the measured angle of curvature at each pixel is compared with certain range of angle, showing the spiculated region. Here we have considered spiculated angle range as 45° to 60° and if any pixels showing this feature are found, they are marked 'x' as shown in Fig. 4(d) and are considered for analysis.

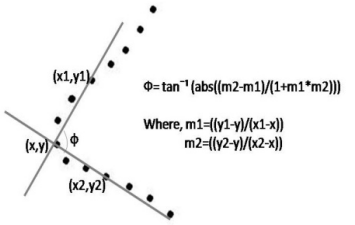


Fig. 3. Angle of curvature at pixel (x,y) found by projecting lines from that pixel

Shape of the mass is also one of the important features that can be considered for classification of mass as benign or malignant in ultrasound images. The proportion of width and height of the mass and its ellipsoid shape are some of the important features which help us to decide the mass as benign or malignant [16]. A mass with ellipsoidal shape shown in Fig. 5(a) will increases the probability of mass being benign. Most of the malignant masses will normally produce projections from the surface of the mass which extend towards nipple, thus they will be taller than wider as shown in Fig. 5(b).

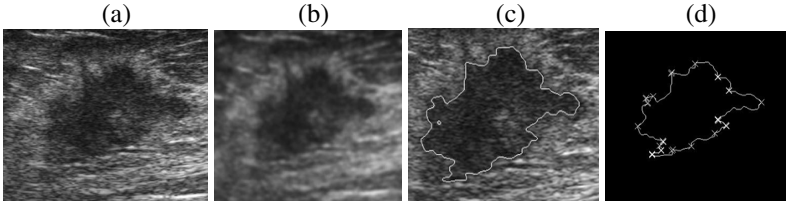


Fig. 4. (a) Ultrasound ROI showing Spiculated mass, (b) ROI after preprocessing, (c) ROI after segmentation, (d) Spiculations marked with "x"

A mass is said to produce acoustic shadow if the ultrasound is attenuated when crossing through it. If a mass generates acoustic shadow it is considered as malignant [16]. Appearance of mass with acoustic shadow and without acoustic shadow is shown in Fig. 5(c) and Fig. 5(d). Acoustic shadow can be determined by considering

intensity as a main factor. To find whether mass has generated shadow or not, we first calculated the mean intensity of the region under the mass and compare it with the mean intensity of the region at the same level which is not covered by the mass.

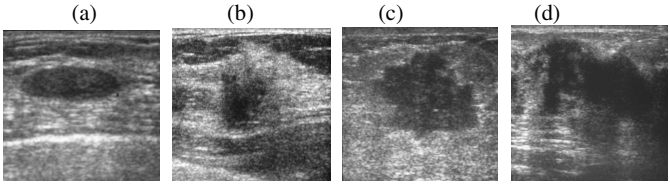


Fig. 5. (a) Mass with elliptical shape (benign), (b) Mass which is taller than wider (malignant), (c) Mass without Acoustic shadow, (d) Mass with acoustic shadow

The features we have considered are discriminative and effective in characterizing the mass in ultrasound images. Using a single feature as parameter to discriminate is always a tradeoff between the sensitivity and specificity. The tradeoff is due to that each feature parameter is mainly related to its nature. So we have considered the features like spiculation feature, acoustic shadowing, elliptical shape feature, entropy, homogeneity and standard deviation for discrimination.

3.3 Feature Level Fusion

The features retrieved from mammogram are structural features and that from ultrasound are functional features and they are dissimilar in terms of dimension. For fusion of features we needed coherent dataset from both modalities which belong to a same person. Data set was created by collecting and getting the ground truth marked images from expert radiologists trained with those kinds of images. All images in our dataset contained only one abnormality (AD with spiculated mass). As discussed in previous sections both modalities will output a collection of features. The fusion process fuses this collection of features into a single feature set. Feature level fusion is a medium level fusion strategy which performs well, if the features are homogenous. If the features are heterogeneous, then it requires normalization to convert them into a range that makes them more similar. We have used Z-score normalization which transforms the scores to a distribution with mean of '0' and standard deviation of '1' as shown below:

$$n = \frac{S_i - mean(S)}{std(S)} \tag{2}$$

4 Experimental Results and Discussion

The proposed method is applied on 20 set of images. Where each set consists of one mammogram and one ultrasound image of a same person out of which 9 set were malignant and 11 were benign. Data set was created by collecting and getting the ground truth marked images from expert radiologists trained with those kinds of images. All images in our dataset contained only one abnormality.

The feature level fusion is realized to be simply concatenating the feature points obtained from different sources of information. The concatenated feature vector has better discrimination power than the individual feature vectors. The main aim of this fusion is to test if the fusion of two sets of features can improve classification accuracy. A fair assessment should be based on the feature vectors with the same dimensionality. In this feature level fusion, proper normalization is required to address the difference in measurement scale because during fusion we augment features that are retrieved from different extraction methods. One major problem associated with this fusion scheme is that the same classifier has to be applied to the fused feature set. But the feature sets from both the modalities mammogram and ultrasound may have different utilities and may have their own individuality favored classifiers. Because it is known that classification performance depends mainly on the characteristics of the data. It is difficult to find a single classifier that works best on all given data sets.

Support vector machines (SVM) are a learning tool based on modern statistical learning method that classifies binary classes. SVM has been shown to perform better than many other classification algorithms due to several reasons [13, 20]. Thus we have used SVM classifiers to classify the fused feature vector. Implementation is done using MATLAB. For experimentation we have randomly partitioned the dataset training and testing data with the proportion of 70% and 30% respectively. We have used receiver operating characteristic curve (ROC) to evaluate the performance. ROC graphically represents the true positive rate as a function of false positives rate. The sensitivity achieved by SVM classifier in classifying breast mass using dual modality is 95.6%. Maximum sensitivity achieved in classifying breast mass using single modality mammogram and ultrasound was 93.52% and 92.7% respectively [13, 16].

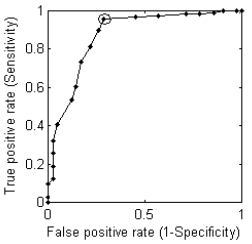


Fig. 6. ROC curve depicting the performance of SVM classifier for dual modality with sensitivity 95.6%

Table 1. Comparison of Results

Modality	Problem addressed	Sensitivity (%)
Ultrasound [16]	Classification of breast mass	92.7
Mammogram [13]	Classification of breast mass	93.52
Dual modality (Ultrasound and Mammogram)	Classification of breast mass	95.6

5 Conclusions

In this study information from multiple modalities such as mammogram and ultrasound was used to classify the breast mass as benign or malignant. We have performed feature level fusion by fusing the features retrieved from both the modalities. The features retrieved from mammogram are spiculation feature and denseness texture feature and that from ultrasound are spiculation feature, shape feature and shadowing feature. From both modalities we have retrieved some additional features like: standard deviation, entropy and homogeneity. Both modalities supply complementary information which is helpful in discriminating benign from malignant mass. The results for the fusion model were compared with individual modalities to understand the effects of more than one modality. Results show that the multimodal fusion improved the performance to classify breast mass more accurately.

References

1. Krol, A., Coman, I.L., Mandel, J.A., Baum, K., Luo, M., Feiglin, D.H., Lipson, E.D., Beaumont, J.: Inter-Modality Non-Rigid Breast Image Registration Using Finite-Element Method (2004) 0-7803-8257-9/04/$20.00 © 2004 IEEE
2. Ravishankar Rao, A., Ramesh, C.: Computerized Flow Field Analysis: Oriented Texture Fields, 0162-8828. IEEE (1992)
3. Kopans, D.: Breast Imaging. Lippincott-Raven Publishers, New York (1998)
4. Sickles, E.A., Filly, R.A., Callen, P.W.: Breast detection with sonography and mammography. AJR 140, 843–845 (1983)
5. Arena, F., DiCicco, T., Anand, A.: Multi-modality data fusion aids early detection of breast cancer Using conventional technology and advanced digital infrared Imaging. In: Proceedings of the 26th Annual International Conference of the IEEE EMBS, San Francisco, CA, USA, September 1-5 (2004) 0-7803-8439-3/04/$20.00© 2004 IEEE
6. Marcialis, G.L., Roli, F., Didaci, L.: Pattern Recognition 42(11), 2807–2817 (2009)
7. Zhang, H.-L., Yang, F.: Multimodality Medical Image Registration Using Hybrid Optimization Algorithm. In: 2008 International Conference on Bio Medical Engineering and Informatics (2008), 978-0-7695-3118-2/08 $25.00 © 2008 IEEE, doi:10.1109/BMEI.2008.108
8. Solimanv, H., Mohamed, A.S., Atwan, A.: Feature Level Fusion of Palm Veins and Signature Biometrics. International Journal of Video & Image Processing and Network Security IJVIPNS-IJENS 12(01), 28
9. Gholam Hosseini, H., Alizad, A., Fatemi, M.: Integration of Vibro-Acoustography Imaging Modality with the Traditional Mammography. International Journal of Biomedical Imaging 2007, Article ID 40980, 8 pages (2007), doi:10.1155/2007/40980
10. Kittler, J., Duin, R.P.W.: The combining classifier: to train or not to train. In: Proceedings of the International Conference on Pattern Recognition, vol. 16(2), pp. 765–770 (2002)
11. Jameson, M.: Ultrasound as a breast cancer test is becoming more accepted, Los Angeles Times, 000037057, June 14 (2004)
12. Sampat, M.P., Whitman, G.J., Bovik, A.C., Markey, M.K.: Comparison of Algorithms to Enhance Spicules of Spiculated Masses on Mammography. Journal of Digital Imaging, 1–8 (2007)

13. Minavathi, Murali, S., Dinesh, M.S.: Model based approach for Detection of Architectural Distortions and Spiculated Masses in Mammograms. International Journal on Computer Science and Engineering (IJCSE) 3(11), 3534 (2011) ISSN : 0975-3397
14. Minavathi, Murali, S., Dinesh, M.S.: Detection of Architectural Distortions with Spiculations in Mammograms by analyzing the structure of mammary glands. In: Proceedings of Fifth Indian International Conference on Artificial Intelligence (IICAI), Tumkur, pp. 218–230 (December 2011)
15. Minavathi, Murali, S., Dinesh, M.S.: Curvature and shape analysis for the detection of spiculated masses in breast ultrasound images. IJMI International Journal of Machine Intelligence 3(4), 333–339 (2011) ISSN: 0975–2927 & E-ISSN: 0975–9166
16. Minavathi, Murali, S., Dinesh, M.S.: Classification of Mass in Breast Ultrasound Images using Image Processing Techniques. International Journal of Computer Applications(IJCA) 42(10), 5120/5731-7801 (March 2012)
17. Wirth, M.A.: Nonrigid Approach to Medical Image Registration Matching Images of the Breast, Ph.D. Thesis, RMIT University, Elbourne, Australia (2000)
18. Brunelli, R., Falavigna, D.: Person identification using multiple cues. IEEE Transactions on Pattern Analysis and Machine Intelligence (1995)
19. Gupta, R., Undrill, P.E.: The use of texture analysis to delineate suspicious masses in mammography. Phys. Med. Biol. 40, 835–855 (1995)
20. Gunn, S.R.: Support Vector Machines for Classification and Regression, Technical Report, University of Southampton (1998)
21. Ikedoa, Y., Fukuokab, D., Haraa, T., Fujitaa, H., Takadac, E., Endod, T., Moritae, T.: Computerized mass detection in whole breast ultrasound images: Reduction of false positives using bilateral subtraction technique. In: Proc. of SPIE Medical Imaging 2007, vol. 6514, p. 65141T (2007)

Design of Fuzzy PD Controller for Inverted Pendulum in Real Time

Nidhi Patel and M.J. Nigam

Indian Institute of Technology,
Roorkee, Haridwar
nidhicoool@gmail.com,
mkndnfec@iitr.ernet.in

Abstract. This paper presents the control of 1-stage non-linear model of Inverted Pendulum (IP) in real time using conventional PID and Fuzzy PD controller. Fuzzy controller is non-linear controller in which it is cumbersome to tune the parameters of fuzzy membership function to acquire the desired results. To overcome this problem, linear fuzzy PD controller is used. Firstly controllers are designed in MATLAB Simulink environment and after that both are implemented in real time for controlling the IP. The simulation results reveals that the performance of Fuzzy PD controller is better, efficient and improved one compared to conventional PID controller.

Keywords: Inverted Pendulum, PID controller, Fuzzy controller.

1 Introduction

The Inverted Pendulum has the property of unstable, higher order, multivariable and highly coupled, which can be treated as a typical non-linear control problem [3]. The IP system provides an excellent experimental platform to test various control theories and techniques. Inverted pendulum can vividly simulate the flight control of rockets and the stabling control in walking robots etc. In conventional control theory, most of control problems are generally solved by mathematical tools based on system models. In practical world, it is not possible to derive exact mathematical model of complex system because of certain uncertainties. Various control techniques are present for which exact mathematical model of the system is not necessary such as fuzzy system, neural system, genetic algorithm etc.

In this paper, fuzzy PD control system is used to control the non-linear model of IP [5] in a better way. Fuzzy system [6] used linguistic variables to approximate the system. Two controllers are used to stabilize cart position and pendulum angle, one for cart position and another for pendulum angle. Firstly, PID controller has been developed by tuning the various gains of PID for stable the IP. Fuzzy controller is a non-linear controller; it is difficult to tune the parameters of membership function which gives better result. To reduce the difficulty, it is convenient to use the linear fuzzy controller [6]. Here, fuzzy PD controller is used which is one of the type of linear fuzzy controller.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 955–962.
springerlink.com

2 Mathematical Modeling of Inverted Pendulum

The 1-stage inverted pendulum is made up of cart onto which pendulum is hinged. The 1-stage inverted pendulum is shown in fig.1 [3]. The cart is constrained to move only in the horizontal x direction, while the pendulum can rotate in the x-y plane. The single inverted pendulum system has two degrees of freedom and can therefore be fully represented using two generalized coordinates: horizontal displacement of the cart, and rotational displacement of pendulum. The physical properties of the system are fixed and are shown in table 1.

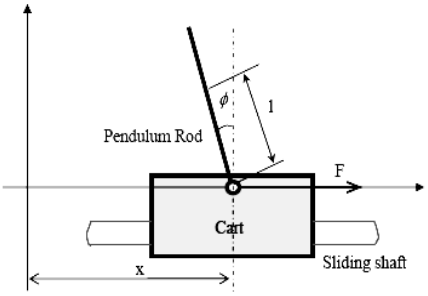


Fig. 1. 1- Stage Inverted Pendulum

In this section, the mathematical model of 1-stage inverted pendulum and dynamical equations will be given. After ignoring the air resistance and other frictions, 1-stage IP can be simplified as a system of cart and rod, as shown in Fig. 1. The two dynamic equations of Inverted Pendulum are:

$$(M + m)\ddot{x} + b\dot{x} + ml\ddot{\theta}\cos\theta - ml\dot{\theta}^2\sin\theta = F \tag{1}$$

$$(I + ml^2)\ddot{\theta} + mgl\sin\theta = -ml\ddot{x}\cos\theta \tag{2}$$

Non linear model of Inverted Pendulum is made with the help of above two equations [5].

Table 1. Physical parameters of Inverted Pendulum

Symbol	Definition	Value
M	Mass of the cart	1.096 Kg
m	Mass of the rod	0.1096 Kg
b	Friction coefficient of the cart	0.1 N/m/sec
I	Rod inertia	0.0034 Kg*m*m
l	Distance from the rod axis rotation center to the rod mass center	0.25 m
F	Force acting on the cart	
x	Cart position	
θ	Angle between the rod and the vertically downward direction	

3 Implementation of Controller

Controllers are used to stabilize the unstable system and make it robust to disturbances. The framework of the Inverted Pendulum-cart system controller is shown in fig.2. As in fig.2, Inverted Pendulum-cart system is controlled by two separate controller, pendulum angle controller and cart position controller. From the dynamic equations of this system, it is found that there are two dynamic objects in the inverted pendulum-cart system. One is the pendulum and the other is the cart. However, there is only one control action is allowed for the inverted pendulum–cart system.

Therefore, the control action Fp for the pendulum subsystem and the control action Fc for the cart subsystem need to be combined into one control action F for the inverted pendulum-cart system. It can be seen that to provide a control action to push the cart toward left-hand side will move the pendulum to the right-hand side. This instinctive knowledge indicates that the control actions to move the cart and pendulum to the same direction have opposite sign. Since the main purpose for the position control of the inverted pendulum-cart system is to balance the pendulum at the straight upward direction, the combination of Fp and Fc is defined as $F=F_p - F_c$ [1].

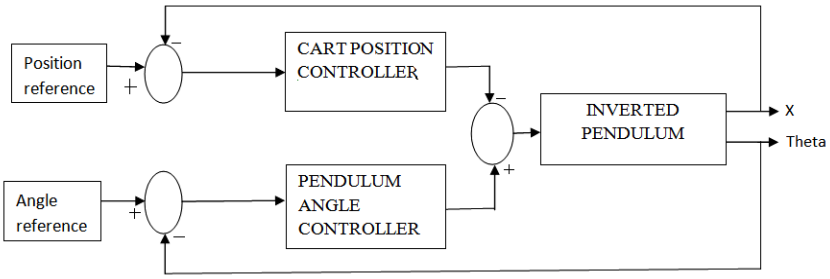


Fig. 2. Block diagram of inverted pendulum-cart controller system

3.1 Conventional PID Controller

Conventional PID controller is widely used in control applications for its simple structure, easy tuning and independence from system model. Inverted pendulum is an unstable system, and the main objective is to stable the pendulum rod in upright position with moving cart in particular position. For this two controllers are used, one for pendulum angle and another for cart position. For both PID controllers, the standard structure is as follows:

$$u = K_p e + K_I \int e dt + K_d \frac{de}{dt} \tag{3}$$

Where u is control action, e is error between desired output and the actual plant output, i.e. $e = y_{ref} - y_{output}$. PID controller trying to reduce the error between the

desired output and the actual output by varying the gains K_p, K_I, K_d. K_p, K_I, K_d are the Proportional, Integral and Derivative gain respectively. In this work, Zeigler-Nicholas tuning is used.

3.2 Fuzzy PD Controller

A fuzzy controller is an automatic, non-linear controller, a self-acting or self-regulating mechanism that controls an object in accordance with the desired behaviour. A fuzzy controller acts or regulates by means of natural language or linguistic language, with the distinguishing feature, fuzzy logic. The basic block diagram of fuzzy controller is shown in fig.3.The fuzzy controller is in between pre-processing block and post-processing block. The fuzzification block converts the crisp input into fuzzy sets. Rule base is in the if-then format, and the variables are error and change in error. This gives control output in fuzzy form. For applying to the plant, it is necessary to convert it into crisp output. To obtain the crisp output, defuzzification block is used.

Conventional fuzzy controller is a non-linear controller, thus it is very difficult to tune the parameters of fuzzy controller for obtain the desired behaviour. To overcome this problem, linear fuzzy controller is used. Linear fuzzy controller is similar to PID controller. In this paper, fuzzy PD controller is used [6], block diagram is shown in fig.4.

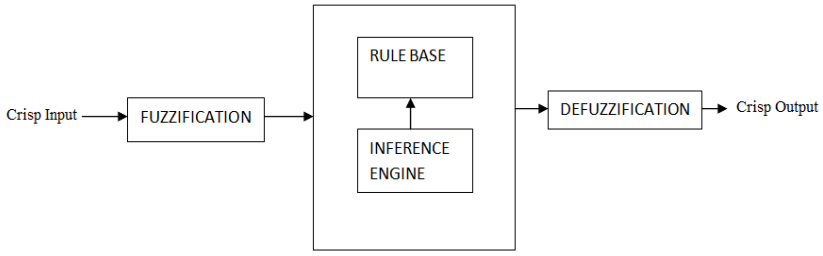


Fig. 3. Basic building block of fuzzy controller

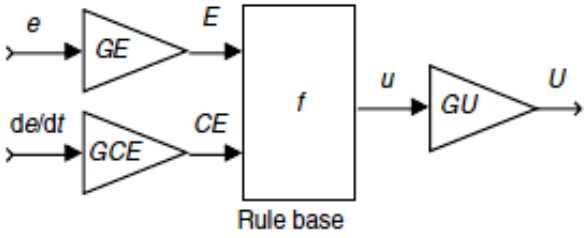


Fig. 4. Fuzzy PD controller

To control the whole inverted pendulum, two controllers are used. One is pendulum angle controller and cart position controller. Inputs to both the controllers are error and change in error. The fuzzy variables used for input variable error(e) , change in error(edot) and control output(u) for cart position controller are three. Therefore, total rules in rule base are 9 for position controller. The fuzzy variables used for input variable error(e), change in error(edot) and control output(u) for pendulum angle controller are five. The total rules in rule base for angle controller are 25.The triangular membership is used for all fuzzy variables and crossover point between membership function is at membership grade value 0.5, for linearize the fuzzy controller. At the time of real time simulation, large number of rules creates a problem; system will hang or shut down. Thus, it is necessary for real time simulation to reduce the rules. Reduction of rules takes place by the elimination of redundant rules, i.e rules which are not fire ever at any time of instant.

4 Simulation Results and Analysis

Inverted pendulum system has been real time simulated with two PID and two Fuzzy PD controllers. The reference values of cart position and pendulum angle are 0.2 and 0 respectively. Various gains of PID controllers are tuned by Zeigler- Nicholas criteria manually. It is necessary to choose the correct values of gains, otherwise system will unstable.

PID parameters of Cart position controller are:

$$K_p = 2.55, K_I = 0.005, K_d = 0.005$$

PID parameters of Pendulum angle controller are:

$$K_p = 50, K_I = 35, K_d = 8$$

Two fuzzy PD controllers are used. Triangular membership function is used for all fuzzy variables for both controllers. The membership function of inputs error, errordot and control output for Cart position controller are in the range: [-0.5 0.5], [-0.03 0.03] and [-10 10] respectively as shown in fig.5. The membership function of inputs error, errordot and control output for pendulum angle are in the range: [-3 3], [-3 3], and [-30 30] respectively as shown in fig.6. Gain of angle fuzzy PD controller: GE=4, GCE=0.6, GU=1. Gain of cart position fuzzy PD controller: GE=2, GCE=0.05, GU=1. The results are shown in fig.7 and fig.8.

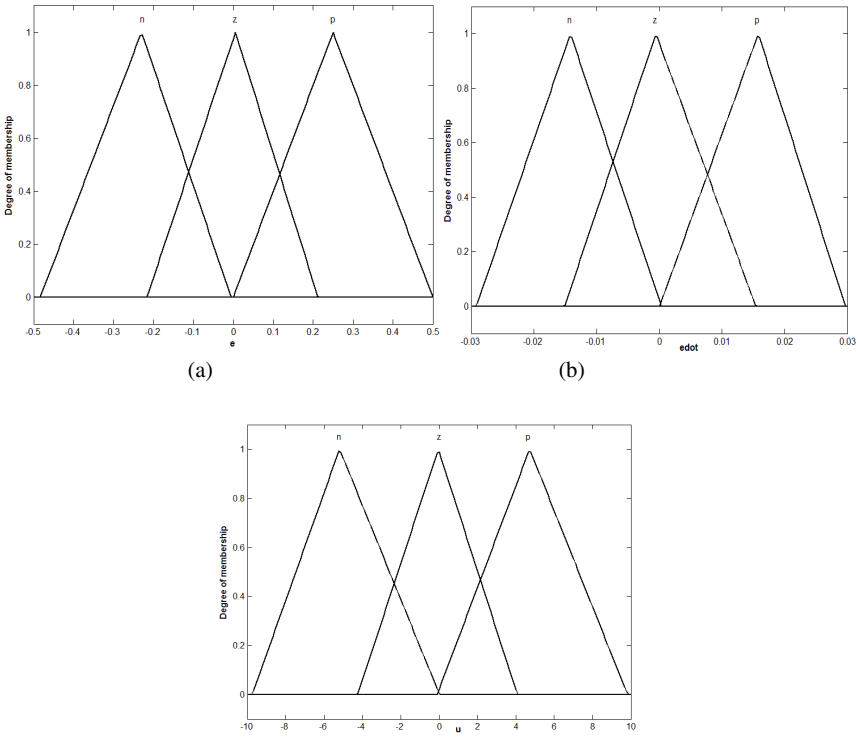


Fig. 5. Membership function of (a) error (b) errordot (c) control output of Cart position controller

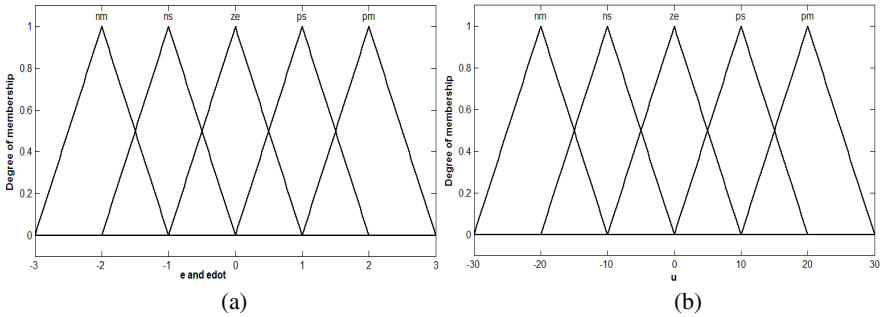


Fig. 6. Membership Function of (a) error and erroedot (b) control output of Angle controller

The rule base used in the Fuzzy- PD controller for pendulum angle and cart position controller is shown in table 2 and 3 respectively.

Table 2. Rule base for pendulum angle controller

e/ė	NM	NS	Z	PS	PM
NM	-	-	-	-	-
NS	-	-	NM	Z	-
Z	-	NM	-	PM	-
PS	NM	Z	PM	-	-
PM	Z	PM	-	-	-

Table 3. Rule base for cart position controller

$e/\dot{e}$	N	Z	P
N	N	N	Z
Z	N	Z	P
P	Z	P	P

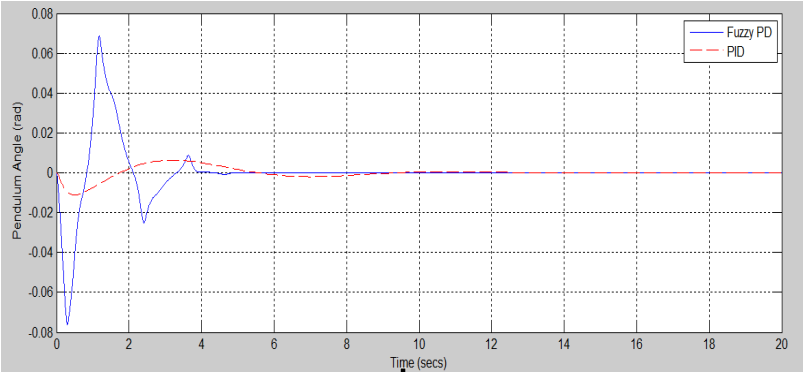


Fig. 7. Pendulum angle response of Inverted pendulum

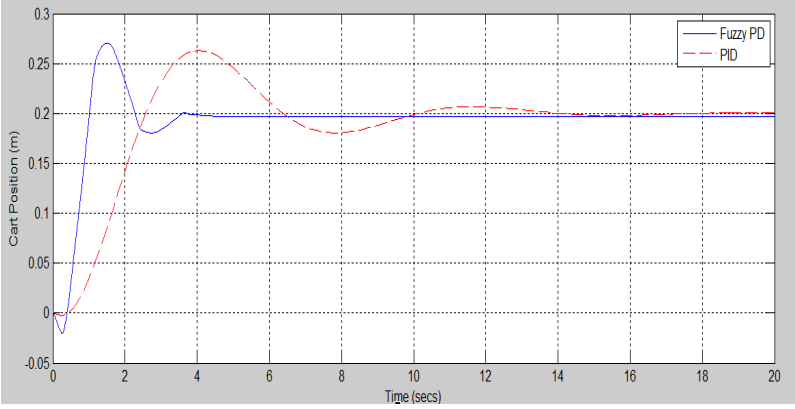


Fig. 8. Cart position response of Inverted pendulum

5 Conclusion

The objective of this work was to design a controller for stable the inverted pendulum which has been successfully achieved. The fuzzy PD controller is proved to be effective and feasible in both of the angular control of pendulum at upright position and position control of cart to the desired position. As shown in the results, that pendulum angle and cart position both acquire desired position in 3.5 secs. in case of Fuzzy PD controller while 14 secs in PID controller. It is clear from results that fuzzy PD controller stable the Inverted Pendulum more accurately and efficiently than the PID controller.

Acknowledgements. This work is supported by MHRD, Government of India, Indian Institute of Technology, Roorkee.

References

1. Akole, M., Tyagi, B.: Design of fuzzy logic controller for nonlinear model of inverted pendulumcart system. In: XXXII National Systems Conference, NSC, pp. 750–755 (2008)
2. Ji, C.W., Lei, F., Kin, K.: Fuzzy Logic Controller for An Inverted Pendulum System. In: IEEE International Conference on Intelligent Processing Systems, ICIPS, vol. 1, pp. 185–189 (1997)
3. Googol Technology, Inverted Pendulum Experimental Manual, 2nd edn., pp. 1, 25–34 (July 2006)
4. Liu, H., Duan, F., Gao, Y., Yu, H., Xu, J.: Study on Fuzzy Control of Inverted Pendulum system in the Simulink Environment. In: IEEE International Conference on Mechatronics and Automation, ICMA, August 5-8, pp. 937–942 (2007)
5. `http://www.library.cmu.edu/ctms/ctms/simulink/examples/pend/pendsim.htm`
6. Jantzen, J.: Foundations of Fuzzy control, pp. 71–90. John Wiley & Sons (2007)
7. Tanak, K., Sugeno, M.: Stability analysis and design of fuzzy Control systems. Fuzzy Sets and Systems 45, 135–156 (1992)
8. Ross, T.J.: Fuzzy Logic with Engineering Applications. McGraw-Hill, Inc. (1991)

Classification of Kannada Numerals Using Multi-layer Neural Network

Ravindra S. Hegadi

Department of Computer Science, Solapur University,
Solapur - 413255, India
ravindrahegadi@rediffmail.com

Abstract. A simple multilayer feed forward neural network based classification of handwritten as well as printed Kannada numerals is presented in this paper. A feed forward neural network is an artificial neural network where connections between the units do not form a directed cycle. Here four sets of Kannada numerals from 0 to 9 are used for training the network and one set is tested using the proposed algorithm. The input scanned document image containing Kannada numerals is binarized and a negative transformation is applied followed by noise elimination. Edge detection is carried out and then dilation is applied using 3 × 3 structuring element. The holes present in this image are filled. Every image is then segmented out forming 50 segmented images each containing one numeral, which is then resized. A multilayer feed forward neural network is created and this network is trained with 40 neural images. Then testing has been performed over ten numeral images. The proposed algorithm could perfectly able to classify and recognize the printed numerals with different fonts and hand written numerals.

1 Introduction

Kannada or Canarese, is a language spoken in India and it is official language of Karnataka state. Kannada is the 25th most spoken language of the world with about 60 million native speakers. It has got the status of the scheduled languages of India and the official and administrative language of the state of Karnataka. The Kannada script is an alphasyllabary (sometimes called an abugida) of the Brahmic family, [3] used primarily to write the Kannada language, one of the southern language in India and also Sanskrit in the past. Kannada language uses forty-nine phonemic letters, divided into three groups: swaragalu (vowels - thirteen letters); vyanjanagalu (consonants - thirty-four letters); and yogavaahakagalu (neither vowel nor consonant - two letters: the anusvara and the visarga).

Development of OCR for Kannada characters and numerals is still open problem. There were few works reported for the recognition of Kannada character [4]. A template matching based approach for numeral recognition was proposed by Hegadi R. S. [6], in which the resized numeral is compared with the stored templates. Based on the correlation coefficient between the two numerals, recognition is carried out. This method has reported reasonable accuracy. A neural network based classifier using wavelet transform coefficients as features for recognition was proposed by Kunte S. R.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 963–968.
springerlink.com

et. al. [7]. It could address the problems associated with the template matching approaches for character recognition. The work proposed by Ashwin T. V. et. al. [1] for printed Kannada characters works on template matching approach for recognition mechanism and uses SVM classifier. This methodology is highly sensitive to font changes, and pre-processing stage is not framed properly, due to which the problems are found in segmentation and resizing stages. Kannada character recognition based on k-means clustering is reported in [8]. The authors propose a segmentation technique to decompose each character into components from 3 base classes, thus reducing the magnitude of the problem. K-means provides a natural degree of font independence and this is used to reduce the size of the training database to about a tenth of those used in related work.

In this paper a multilayer back propagation neural network based classifier is proposed for classifying the Kannada numerals. Before classification, the numerals are pre-processed, which include binarization, edge detection, dilation and region filling. The proposed methodology is discussed in section 2, section 3 presents the results of this work and the conclusions are presented in section 4.

2 Proposed Methods

The image document containing printed Kannada numerals having 5 copies of each numeral, forming 50 numerals will be the input for the system. The proposed algorithm will use four set of numerals for training the network and testing will be performed on one set of numerals. The proposed method has two stages namely image preprocessing stage and segmentation and classification stage as described in the following sub sections.

2.1 Image Preprocessing

The hard copy of the document containing numerals is scanned using optical scanner. The scanned image will be a color image, which is converted to a binary image by applying thresholding. The pixels corresponding to the numerals will be black and the background pixels will be white. A negative transformation is applied on this image. The image may contain noise in the form of tiny dots, due to the optical sensors, which may mislead the detection process. These dots are eliminated by applying the morphological opening operation. Morphological opening operation removes small objects from the foreground (usually taken as the dark pixels) of an image, placing them in the background. Through this process all those regions which contain less than 20 pixels are removed. This image is subjected to edge detection using Sobel method. The Sobel method finds edges using the Sobel approximation to the derivative. It returns edges at those points where the gradient of image is maximum. Due to the edge detection process the numerals can have breaks. These breaks are filled by performing dilation over this image. Dilation is one of the basic morphological operations. Dilation operation adds pixels to the boundaries of objects in an image. A structuring element is used for performing the dilation. Depending on the size and shape of this structuring element, the number of pixels added from the objects in an image. In the morphological dilation operation, the state of any given pixel in the output image is determined by applying a rule to the corresponding pixel and its neighbors in the

input image. A structuring element of size 3 × 3 is chosen for dilating the image. There may be holes within many numerals. A hole is a set of background pixels that cannot be reached by filling in the background from the edge of the image. These holes are filled by performing the morphological region filling operation. This image contains 50 regions, forming one region for each numeral. Each region is segmented and a bounding box is applied over individual region. The drawing of bounding box will eliminate all the empty rows and columns (i.e. rows and columns with 0 values) on all four sides of each numeral. Each numeral is further resized to 7 × 5 pixels. Further the two dimensional data is converted to a one dimensional vector. Figure 1 shows the intermediate images in the pre-processing stages. In Figure 1 (a), the original scanned image containing hand written numerals are shown. It is converted into binary image and its noise is removed. Then edge is detected using Sobel operator as shown in Figure 1(b). The result of dilation operation is shown in Figure 1(c). The holes in the numerals are filled as shown in Figure 1(d). The next stage is to classify using neural networks.

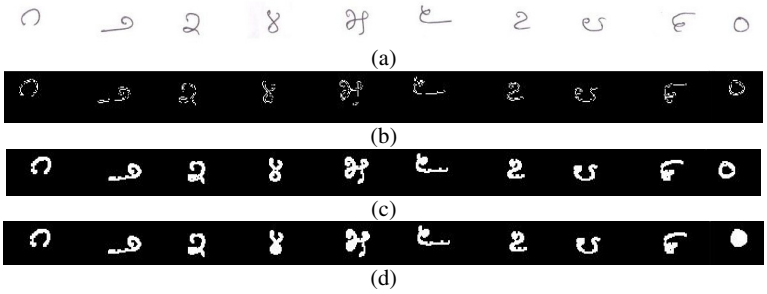


Fig. 1. Images in pre-processing stages, (a) Original handwritten numeral image, (b) edge image after binarization and inverse transform, (c) dilated image, and (d) image after filling of holes

2.2 Neural Network Based Classification

A general method of minimizing the objective function by training artificial neural networks is through backpropogation. It was described as a multi-stage dynamic system optimization method by Bryson A. E. [2]. It is a supervised learning method, and is a generalization of the delta rule. It requires a dataset of the desired output for many inputs, making up the training set. It is most useful for feed-forward networks which do have feedback or do not have looping connections. The activation function used by the artificial neurons (or "nodes") should be differentiable for Backpropagation.

A feed-forward network has a multi-layered structure. Each layer consists of units which receive their input from units from a layer directly below and send their output to units in a layer directly above the unit. There are no connections within a layer. The N_i inputs are fed into the first layer of $N_{h,1}$ hidden units. The input units are merely 'fan-out' units; no processing takes place in these units. The activation of a hidden unit is a function f_k of the weighted inputs plus a bias, as given in in equation

$$y_k(t+1) = f_k(s_k(t)) = f_k\left(\sum_j w_{jk}(t) y_j(t) + \theta_k(t)\right) \quad (1)$$

The output of the hidden units is distributed over the next layer of $N_{h,2}$ hidden units, until the last layer of hidden units, of which the outputs are fed into a layer of N_o output units.

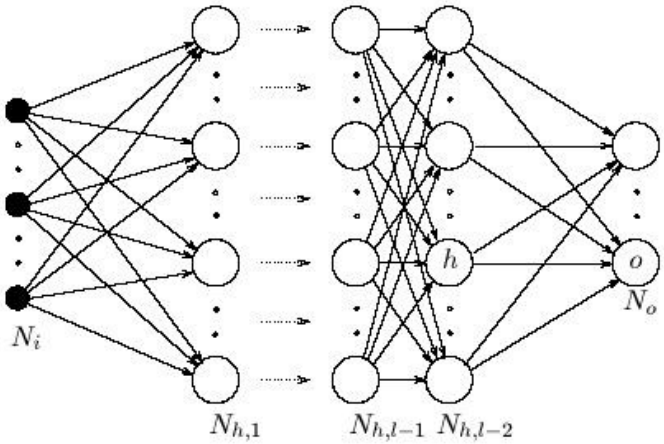


Fig. 2. Multilayer feedforward neuralnetwork architecture

Although backpropagation can be applied to networks with any number of layers, just as for networks with binary units it has been shown that only one layer of hidden units suffices to approximate any function with finitely many discontinuities to arbitrary precision, provided the activation functions of the hidden units are non-linear [5]. In most applications a feed-forward network with a single layer of hidden units is used with a sigmoid activation function for the units.

A two layer feed-forward backpropagation neural network is created with 40 representative elements corresponding to input vector and 10 elements of target vector. The size of first and second layer is set to 10. Each layer has a differentiable transfer function.

3 Results

The proposed algorithm is implemented using Matlab release 2010 on PC with Intel Core I5 processor. Figure 3 shows the handwritten numerals in five rows.

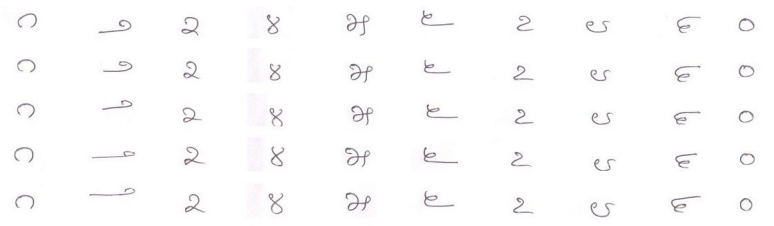


Fig. 3. Handwritten numerals used for training and testing the network

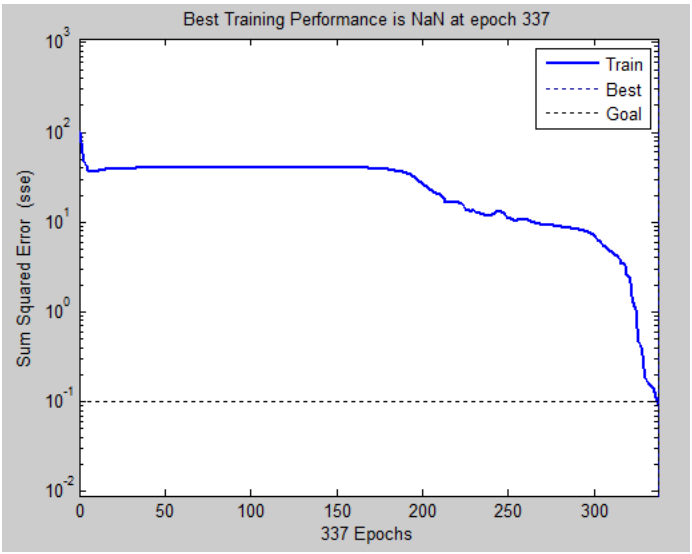


Fig. 4. Performance of the network

The first four rows of numerals are used for the training of the network and the testing has been done on the last row. The proposed algorithm could correctly detect the last row of numerals. The proposed algorithm had initial sum of squared error in the range of 100s. The network has converged to the least value of sum of squared error less than 0.1 in 337 iterations, which is shown in Figure 4. The proposed algorithm has been tested on the printed Kannada numerals and it could successfully classify all the numerals.

4 Conclusions

In this paper a two layer feed-forward neural network is used for the classification of both printed and handwritten numerals. The Proposed algorithm could successfully detect the Kannada numerals when the system is fed with both printed and handwritten numerals. The number of iterations required for printed numerals were lesser that the handwritten numerals. The algorithm is tested for few sets of printed and handwritten numerals. Especially the performance has to be validated for different printed Kannada numerals generated through different Kannada font generation software tools namely Nudi and Baraha.

References

1. Ashwin, T.V., Sastry, P.S.: A font and size independent OCR system for printed Kannada documents using support vector machines. Sadhana 27(1), 35–58 (2002)
2. Bryson, A.E., Ho, Y.C.: Applied optimal control: optimization, estimation, and control, p. 481. Blaisdell Publishing Company or Xerox College Publishing (1969)

3. Campbell, G.L.: Handbook of scripts and alphabets, p. 8485. Routledge, New York (1997)
4. O'Gorman, L., Kasturi, R.: Document image analysis. IEEE Comp. Soc. Press (1995)
5. Hartman, E., Keeler, J.D., Kowalski, J.M.: Layered neural networks with Gaussian hidden units as universal approximations. Journal of Neural Computation 2(2), 2010–2215 (1990)
6. Hegadi, R.S.: Template Matching Approach for Printed Kannada Numeral Recognition. In: Int. Conf. Comp. Int. Info. Tech. (CIIT), Pune, India, pp. 480–483 (2011)
7. Kunte, R.S., Samuel, R.D.S.: An OCR system for printed Kannada text using two-stage Multi-network classification approach employing Wavelet features. In: IEEE Com. Soc. Int. Conf. Comp. Intell. Mult. App. India, pp. 349–355 (2007)
8. Sheshadri, K., Ambekar, P.T., Prasad, D.P., Kumar, R.P.: An OCR system for Printed Kannada using k-means clustering. In: IEEE Int. Conf. Ind. Tech. (ICIT), Chile, pp. 183–187 (2010)

Content Based Image Retrieval by Combining Median Filtering, BEMD and Color Technique

Purohit Shrinivasacharya[1] and M.V. Sudhamani[2]

[1] Siddaganga Institute of Technology, Tumkur - 03
purohitsn@gmail.com
[2] RNS Institute of Technology, Bengaluru - 61
mvsudha_raj@hotmail.com

Abstract. A Content Based Image Retrieval (CBIR) system provides an efficient way of retrieving most similar images from image collections. In this paper we present a novel approach which combines color and edge features to extract similar images. We apply median filtering technique to original image to get the smoothed image. The Bi-directional Empirical Mode Decomposition (BEMD) technique is applied to extract edge information from the image. Then we replace only the values of edge position of smoothed image with the detected edge image values by BEMD and extracted 64 bins gray features. Later we apply one dimensional color histogram technique to obtain histogram vector by using RGB color space and is converted into 32 bins color features. Finally, we combine both the features to extract the most similar images from the database. The experiment is conducted on 1000 images of different categories stored in groundtruth database and the effectiveness of this technique is demonstrated. The results have been tabulated and compared with the conventional median and edge technique. We can observe that performance our proposed method is good.

Keywords: CBIR, BEMD, Indexing, Image database, Histogram, Color.

1 Introduction

Application of World Wide Web and the internet is increasing exponentially, the need for finding an image in internet is also increasing rapidly. A huge amount of image databases are added every minute and so is the need for effective and efficient image retrieval system. Retrieving an image having some characteristics in a big database is a crucial task. Searching for an image among a collection of images can be done by different approaches.

Currently is a growing interest in CBIR technique because of the limitations inherent in text based systems, as well as the large range of possible uses for efficient image retrieval. The present technology is adequate to search images using textual information. But it requires humans to personally describe every image information in the database. This is very impractical for large amount of image databases. It is possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like "cat" as a subclass of "animal" avoid this problem but still face the same scaling issues.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 969–975.
springerlink.com

A CBIR is an alternative or complementary for the textual indexing search. CBIR is the process of retrieving images from a database on the basis of features that are extracted from the images themselves. We present a CBIR system which accepts a query image as input and relevant images are retrieved based on the similarity of the features of the query image and features of the individual images stored in database. The proposed method uses the BEMD, color and median filtering [2] histogram techniques to build new system. The detailed description will be covered in section 2 and 3 of this paper. The proposed system uses the different category images. There are 10 categories have been chosen are as shown in the following Fig.1.

Fig. 1. Sample Database of CBIR System

2 CBIR System

Content based image retrieval system mainly consists of two phases.

Offline Phase: In this phase feature vectors are extracted for the collection of images and stored those features as well as their index (image name) in the database. This phase is as shown in the Fig.2.

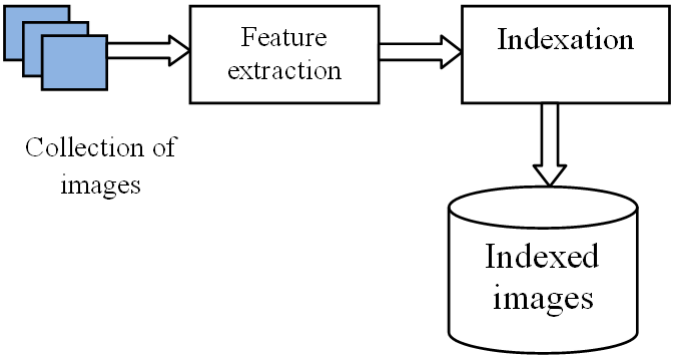


Fig. 2. Offline Phase of the CBIR System

Online Phase: In this phase a query image is accepted from the user for extracting the similar images from database. The feature vector of the query image is extracted in the same manner and compared with feature vectors in the database for finding similar images and intern to display on the Graphical User Interface (GUI). This is as shown in Fig.3.

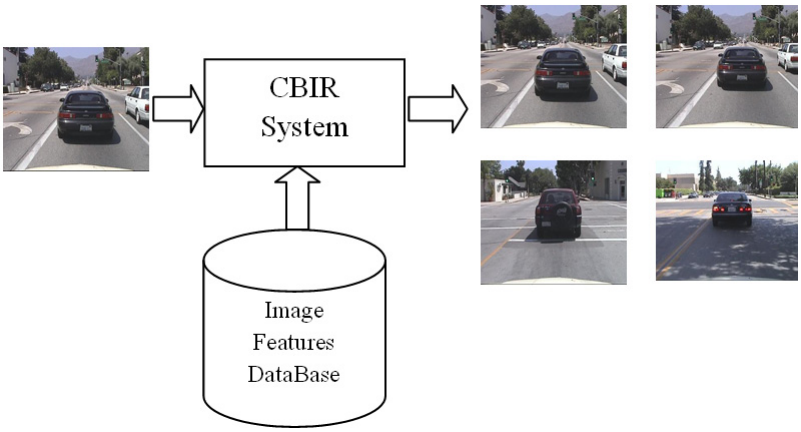


Fig. 3. Online Phase of the CBIR System

3 Feature Extraction

3.1 Color

The color feature is one of the most widely used visual features in image retrieval. The color of an image is represented through some color model. We have chosen RGB as colorimetric space. Typical characterization of color composition is done by color histograms.

The RGB image is resized into 512 × 512 and the one dimensional color histogram vector is obtained. Then the obtained color histogram vector is converted into 32 bins as a color features. The 32 bins are stored for comparison in database for similarity assessment. A histogram method is one of the technique is used for similarity [8] measure. The color histogram vector of query image and the stored images histogram of the database are compared using Bhattacharyya coefficient [9] in online phase.

3.2 Edge Detection Using BEMD

Edge detection is one of the primary means to process an image. Edges in images are areas with strong intensity contrasts, with a jump in intensity from one pixel to the next.

In this paper edge is detected based on the empirical mode decomposition algorithm. When the original image is decomposed by BEMD, first Internsic Mode Frequency (IMF) image has a very good edge characterization. After handling the first

IMF image by the suitable threshold, we obtain clear edge image. The process of shifting is extracting the edge from the image is as follows.

Assume that $X(t)$ is the original signal and let $R(t) = X(t)$, $k = 0$ and $i = 0$

1. Find the local minima and maxima of $R(t)$
2. Find the upper envelop $E_{max}(t)$ by interpolating between maxima and find the lower envelop $E_{min}(t)$ with minima.
3. Calculate the mean envelop as an approximation to the local average

$$M(t)_1 = \frac{E_{\max}(t) + E_{\min}(t)}{2} \tag{1}$$

4. Let $i = i + 1$ and define the intermediate-mode function as

$$P_i(t) = T(t) - M(t) \tag{2}$$

5. Repeat steps 1 to 4 on $P_i(t)$ until it is an IMF, then record the IMF

$$C_1 = P_i(t) \tag{3}$$

6. Let $R(t) = R(t) - C_k(t)$, if stopping criteria is reached then stop the shifting process otherwise $k = k + 1$, $i = 0$,, and goto step 1.

After IMFs are extracted through the sifting process, the original signal $X(t)$ can be represented like this:

$$X(t) = \sum_{j=i}^{n} (C_n(t) + R_n) \tag{4}$$

where $C_n(t)$ is the n^{th} IMF and $R(t)$ is the residue.

The above process is used for single dimension as a signal, for the two dimensions we have used the following standard division criteria for stopping the sifting process [7]

$$SD_k = \sum_{m=0}^{N} \frac{|P_{i-1}(m) - P_i(m)|^2}{(P_{i-1}^2(m))} \tag{5}$$

3.3 Feature Extraction Using Edge and Median Filtering

The following steps are carried out for generating feature vector for the images.

1. The image is converted to gray scale image.
2. Histogram Equalization is applied for gray scale image.
3. Edge is detected using empirical mode decomposition.
4. Median filtering is applied to the histogram equalized gray scale image block of 3 × 3.
5. Replace the values of edge position of median filtering image with detected edge values by BEMD.
6. Feature vector is stored in the database.

3.4 Similarity Measure

The Bhattacharyya coefficient technique is most popular method to correlate the color histogram images. This method returns a value in the range 0 to 1 and stored in the variable SC where,

0 = very low similarity, 0.9 = good similarity, 1 = perfect similarity

Let A be an image in database and Q be a query image and A (n) and Q(n) be the average value of pixels of each bin, where n=1, 2,... ...,64. Difference between the value of each bin is calculated as diff (n) =abs (A (n) - Q (n)) n=1, 2,....,64. The average value of diff (n) is stored in variable SE.

After getting the color and edge feature values in *SC* and *SE* respectively, we calculate the distance between the query image and images in database. The similarity measure used is a sum of weighted color and edge features. This similarity measure [5] is given by:

$$S = \alpha.SC + \beta.SE \qquad with \qquad (\alpha + \beta) = 1$$

Where *SC* is color similarity, *SE* is edge similarity, lower the value of S, higher will be the similarity and vice-versa. The choice of these parameters depends on the query image. They can also be set automatically based on general description of the image. We have conducted experiments on different values of α and β and finally set the values at α = 0.65 and β = 0.35 because color component is more dominate compared to the edge component for getting a better results.

4 Experimental Results

The evaluation of the CBIR system is done by submitting query image to retrieve similar images from various categories of database images. We conducted experiments on the ground truth database provided by James S. Wang et al. [3, 4]. There are 1000 images of 10 different categories and each category has 100 similar images. The different query images and its corresponding retrievals are shown in Fig. 4, Fig. 5 and Fig. 6. Here, only one page displays of results are shown.

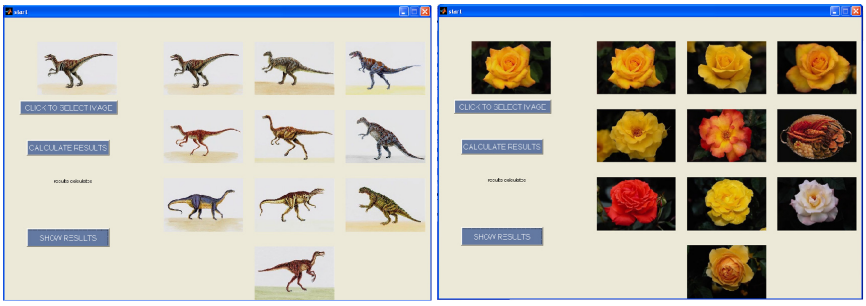


Fig. 4. Query image and results for dinosaur's and rose's category using proposed algorithm

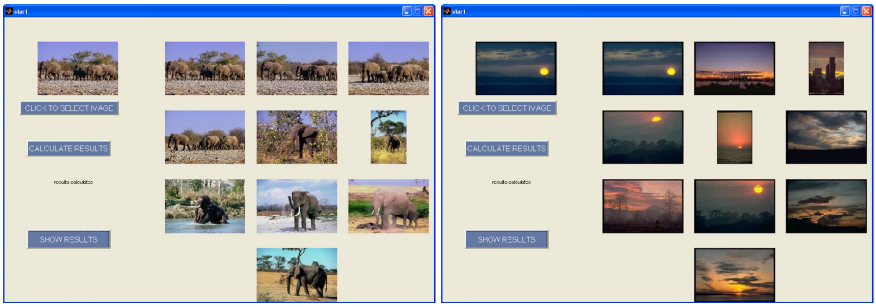


Fig. 5. Query image and results for elephant's and sunset's category using proposed algorithm

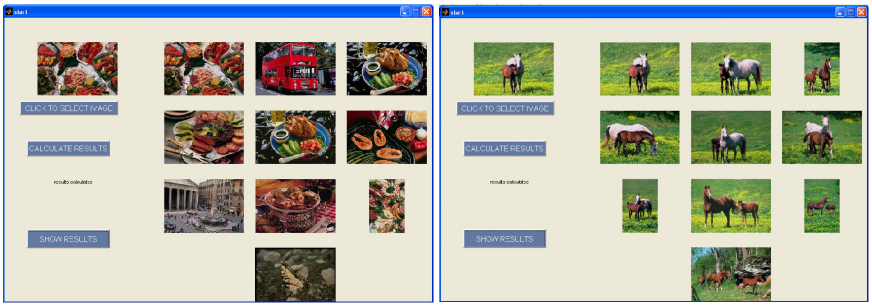


Fig. 6. Query image and results for food's and horse's category using proposed algorithm

Precision and recall are two widely used metrics for evaluating the correctness of CBIR system. The precision and recall values of the median filtering and BEMD edge histogram method [6] and our proposed system are shown in the Table 1. We could observe here that a substantial improvement in the average value of precision 65.10 % and recall value of 38.07% compared to existing system [6] values of 63.00% and 37.00% respectively.

Table 1. Precision and Recall for Existing and Proposed Method

Sl. No.	Category	Existing Method		Proposed Method	
		Precision	Recall	Precision	Recall
1	Buildings	47.25	20.00	53.00	22.30
2	Africans	73.33	42.00	75.00	40.40
3	Buses	48.33	33.00	57.00	39.40
4	Dinosaurs	99.00	98.00	99.00	89.00
5	Elephants	56.00	35.00	58.00	29.90
6	Food	57.00	33.00	55.00	30.50
7	Flowers	74.82	52.00	69.00	37.50
8	Horses	89.40	40.00	65.00	33.90
9	Mountains	28.00	19.00	63.00	29.90
10	Sunset	48.25	30.00	57.00	27.80
	Average	**63.00**	**37.00**	**65.10**	**38.07**

5 Conclusions

We have presented a novel approach for image retrieval by combining edge and color features. Around 300 queries on image database of 1000 images with 10 different categories have been considered for experimental purpose. Experimental results shown that there is a substantial improvement in the performance of image retrieval system in respect of precision and Recall with combination of color and edge features. However, futher enhancement in system performance can be achieved by exploring different features which is our curent research focus.

References

1. Liang, L.F., Ping, Z.L.: An Edge Detection Algorithm of Image Based on Empirical Mode Decomposition. In: Proc. of IEEE Second International Symposium on Intelligent Information Technology Application, vol. 1, pp. 128–132 (2008)
2. Zhao, H., Kim, P., Park, J.: Feature Analysis Based on Edge Extraction and Median Filtering for CBIR. In: 11th International Conference on Computer Modelling and Simulation, vol. 48, pp. 245–249 (2009)
3. Li, J., Wan, J.Z.: Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 25, 1075–1088 (2003)
4. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. In: Laurini, R. (ed.) VISUAL 2000. LNCS, vol. 1929, pp. 360–371. Springer, Heidelberg (2000)
5. Ait-Aoudia, S., Mahiou, R., Benzaid, B.: YACBIR-Yet Another Content Based Image Retrieval System. In: 14th International Conference Information Visualisation, pp. 570–575 (2010)
6. Shrinivasacharya, P., Kavitha, H., Sudhamani, M.V.: Content Based Image Retrieval by Combining Median Filtering and BEMD Technique. In: International Conference on Data Engineering and Communication Systems (ICDECS 2011), vol. 2, pp. 231–236 (December 2011)
7. Nunes, J.C., et al.: Texture Analysis Based on the Bidimensional Empirical Mode Decomposition. Machine Vision and Applications 16(3), 177–188 (2005)
8. Sizintsev, M., Derpanis, K.G., Hogue, A.: Histogram-based Search: A comparative study. In: Proc. IEEE CVPR, pp. 1–8 (2008)
9. Dubuisson, S.: The computation of the Bhattacharyya Distance between Histograms without Histograms. In: IPTA 2010, pp. 373–378 (July 2010)

Fuzzy Geometric Face Model for Face Detection Based on Skin Color Fusion Model

P.S. Hiremath and Manjunath Hiremath

Department of Computer Science,
Gulbarga University,
Gulbarga-585106, Karnataka, India
hiremathps53@yahoo.com,
manju.gmtl@gmail.com

Abstract. Face detection is an important problem considered in many applications. To analyze the information included in face images, a robust and efficient face detection algorithm is required. The face detection in a complex background is still more difficult. In the present paper, our objective is to propose a novel fuzzy geometric face model for single as well as multiple face detection using a skin color fusion model. We combine the skin region extraction using different color spaces, namely, RGB, YCbCr and HSI, and face detection using fuzzy based geometric face model into a robust face detection system. The skin color fusion model is used to segment the skin color region in a face image. Then, in each of the skin regions, the facial features, namely, eyes and mouth, are extracted by using fuzzy geometric face model. The experimentation has been done using several publicly available standard face databases. The experimental results show that the proposed algorithm performs satisfactorily with an average accuracy of 97.40% and is efficient in terms of accuracy and detection time in comparison with the state-of-the-art methods in the literature.

Keywords: Face detection, skin color fusion model, fuzzy face model.

1 Introduction

Automatic human face detection is a computer technology that determines the presence of human faces in digital images. It can be regarded as a specific case of object-class detection and is a more general case of face localization. In automatic human face localization, the task is to find the locations and sizes of a known number of faces.

Human face detection and segmentation is an active research area. This field of research plays an important role in many applications such as face identification system, face tracking system, video surveillance and security control system, and human computer interface. These applications often require segmented human face which is ready to be processed. There are many factors that influence the success of human face detection and segmentation. These factors include complex color background, condition of illumination, change of position and expression, rotation of head, and distance between camera and object.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 977–982.
springerlink.com

A survey of literature on the research work focusing on various potential problems and challenges in the face detection and recognition can be found in [1-5]. Many methods have been proposed and developed for human face segmentation. In [6], a detailed experimental study of feature-based face detection against skin-color like backgrounds with varying illumination are presented. The neural network based [10] and view based [11] approaches require a large number of face and non-face training examples. Skin tone color provides a useful cue for face detection. Gupta et al. [7] proposed a combined pre-processing method for facial image data for better processing of raw data for training and also detection of face is presented by using integer wavelet packet transform and SVM classifier. By Tripathi et al. [8] presented a new face detection method which combines the skin color detector and the template matching method, in which the skin color detector to find the faces is based on YCbCr model. The YCbCr model can be used to easily detect the skin color or non-skin color in the images but it is verified that, it may fail in some criteria such as illumination variations. Singh et al. [9] have carried out a detailed experimental study of face detection algorithms based on skin color. Three color spaces, RGB, YCbCr and HSI are of main concern. Hiremath and Danti [14] have proposed a method for detection of multiple faces in an image using skin color information and lines-of-seperability face model. This method is improved by Hiremath and Manjunath [15] by considering only the prominent facial features, namely, eyes and mouth, and thus speeding up the algorithm for face detection.

In this paper, our objective is to propose a fuzzy geometric face model for face detection using skin color fusion model that is able to handle a wide variety of variations in color images based on RGB, YCbCr and HSI color spaces.

2 Skin Color Fusion Model

The range of colors that human facial skin takes on is clearly a subspace of the total color space, assuming that a person framed is not having face with any unnatural color. The proposed method uses three color spaces namely, RGB, YCbCr and HSI. Assuming that the face image is not too dark or not too bright, the bounding ranges for the color component values of these color spaces are obtained from the training images and are given in the Table 1. For each color component of a color model, the binary segmented image showing the skin regions is obtained and then the three segmented binary images corresponding to the three color components are fused to obtain a single binary segmented image showing the skin regions of the input image. Such binary segmented images obtained for each of the three color models, namely, RGB, YCbCr and HSI, are fused to obtain a binary segmented image.

Table 1. The knowledge base of skin color components' range of values obtained from training images for color models, namely, RGB, YCbCr and HSI

Color model	RGB			YCbCr			HSI		
Color Components	R	G	B	Y	Cb	Cr	H	S	I
Minimum Value	120	79	60	100	115	140	0.02	0.28	0.50
Maximum Value	190	134	126	150	125	155	0.05	0.45	0.75

3 Proposed Method

In the proposed method, the input face image is segmented by the skin region extraction method described in the section 2 and obtain the binary segmented image showing the skin regions. Then, in each of the skin region, the prominent facial features namely, eyes and mouth, are searched to construct geometric fuzzy face model. The procedures for the detection of eye features and the construction of fuzzy geometric face model are given in the [15]. The block diagram of the proposed approach is shown in the Fig. 1.

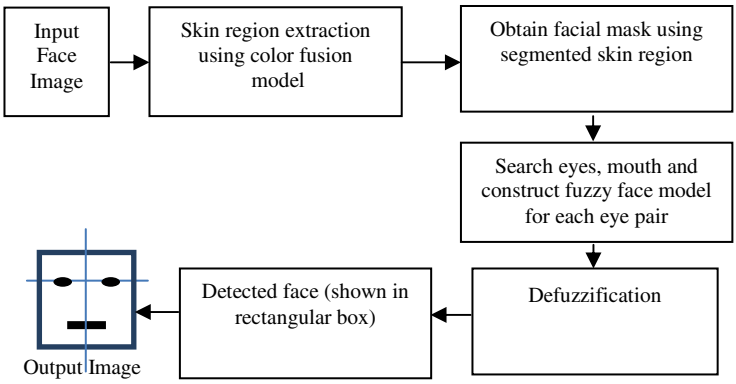


Fig. 1. Proposed method for face detection based on skin color fusion model and fuzzy geometric face model

The proposed algorithm for the detection of face(s) in the input color image, is as given below:

ALGORITHM: Single or multiple face detection algorithm

Step 1: Input RGB color image which contains human face(s).
Step 2: Transform input RGB color image to YCbCr and HSI color spaces.
Step 3: Perform skin region extraction using skin color fusion model described in the section 2 and then obtain facial mask using segmented skin region.
Step 4: For a skin region, construct fuzzy face model for the input image using the fuzzy geometric face model as described in the[15].
Step 5: Perform the defuzzification process.
Step 6: Repeat Step 4 and 5 for each skin region.

Output the detected face(s) image by showing the face(s) in rectangular box(es).

4 Experimental Results

The experimentation of the proposed method is carried out using 730 images chosen randomly from the standard databases, namely, Color FERET face database, MIT CBCL, MUCT, Georgia Tech, Indian Face Database-IIT, and CVL Color Face

Database. The implementation is done on, Intel Core 2 Quad System @2.66 Ghz machine using MATLAB 7.9. The experimental results of the single frontal face detection by the proposed method are shown in the Fig.2 and multiple face detection are shown in the Fig. 3.

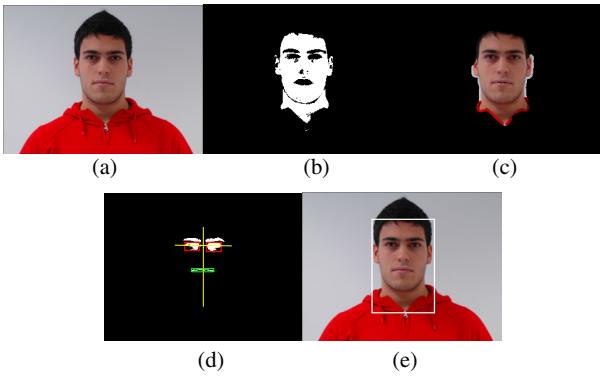


Fig. 2. (a) Original single frontal face image, (b) Segmented skin region, (c) facial mask, (d) Face features obtained by fuzzy face model and (e) Detected face

The test images in the dataset contain single frontal face as well as multiple faces with varying size, poses, expressions, head tilts, lighting conditions and background. In the dataset, there are color and gray scale images excluding too dark or too bright images. The proposed approach is able to precisely locate the facial features and the detected face is shown in a box. The performance of the proposed approach on the different databases is given in the Table 2.

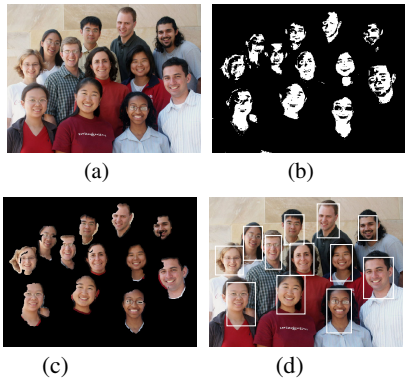


Fig. 3. (a) Original image containing multiple face, (b) Segmented skin region, (c) Facial mask and (d) Detected face(s)

Table 2. The detection performance of the proposed approach for different database set images

Face database	No. of faces	No. of faces detected correctly	Avg. detection rate (%)	Avg. detection time (in sec)
Color FERET Database	245	240	97.59	3.045
MIT CBCL Face Database	10	10	100.00	3.166
MUCT database	250	244	97.60	3.967
Georgia tech Face Database	50	48	96.00	4.011
Indian Face Database IIT	61	59	96.72	3.909
CVL Color Face Database	114	110	96.49	3.711

The comparison of the detection results obtained by the proposed method based on skin color fusion model and that by the method in [15] based on single color space RGB are given in the Table 3. It is observed that the proposed method yields better detection results.

Table 3. The comparison of average detection rate obtained by proposed method and the method in [15].

Face database	Avg. detection rate (%)	
	Proposed method	Method in [15]
Color FERET Database	97.59	96.25
MIT CBCL Face Database	100.00	97.50
MUCT database	97.60	* * *
Georgia tech Face Database	96.00	94.64
Indian Face Database IIT	96.72	91.00
CVL Color Face Database	96.49	* * *

5 Conclusion

In this paper, a novel approach for the human face detection in a digital image based on skin color fusion model and fuzzy face model is presented. The geometrical configuration of only the prominent facial features of the face, namely, eyes and mouth, is used to construct fuzzy face model. The skin color fusion model is based on the color spaces, namely, RGB, YCbCr and HSI, which has yielded better segmentation results and, hence, the improved face detection results, in comparison with single color space RGB. The average detection rate of the proposed approach is 97.40% approximately. The multiple human face(s) in an image with different face sizes, head tilts, lighting conditions, expressions and complex background are detected successfully.

References

[1] Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson Education Asia (2002)
[2] Lukac, R., Plataniots, K.N.: Color Image Processing; Methods and Applications. Taylor and Francis group CRC Press (2007)
[3] Yang, M.H., Kriegman, D., Ahuja, N.: Detecting Faces in Images: A Survey. IEEE Transaction Pattern Analysis and Machine Intelligence 24(1), 34–58 (2002)
[4] Hjelm, E., Low, B.K.: Face Detection: A Survey. Computer Vision and Image Understanding 83, 110.3, 236–247 (2001)
[5] Zhang, C., Zhang, Z.: A Survey of Recent Advances in Face Detection. Tech. Rep., Microsoft Research (2010)
[6] Hu, W.-C., Yang, C.-Y., Huang, D.-Y., Huang, C.-H.: Feature-based Face Detection Against Skin-color Like Backgrounds with Varying Illumination. Journal of Information Hiding and Multimedia Signal Processing 2(2) (2011)
[7] Gupta, M., Gupta, N.: A novel approach for pre-processing of face detection system based on HSV color space and IWPT. International Journal of Advanced Computer Science and Applications, IJACSA 2(11), 144–147 (2011)
[8] Tripathi, S., Sharma, V., Sharma, S.: Face Detection using Combined Skin Color Detector and Template Matching Method. International Journal of Computer Applications (0975 – 8887) 26(7) (July 2011)
[9] Singh, S.K., Chauhan, D.S., Vatsa, M., Singh, R.: A Robust Skin Color Based Face Detection Algorithm. Tamkang Journal of Science and Engineering 6(4), 227–234 (2003)
[10] Rowley, H.A., Baluja, S., Kanade, T.: Neural Network-Based Face Detection. IEEE Trans. Pattern Analysis and Machine Intelligence 20, 23–38 (1998)
[11] Sung, K.K., Poggio, T.: Example-based Learning for View-based Human Face Detection. IEEE Trans. Pattern Recognition and Machine Intelligence 20, 39–51 (1998)
[12] Hiremath, P.S., Danti, A., Prabhakar, C.J.: Modeling uncertainty in representation of facial features for face recognition. In: Delac, K., Grgic, M. (eds.) Face Recognition, p. 183. ITech, Vienna (2007)
[13] Liang, Y., Ma, L., Zhang, L., Miao, Q.: Face Localization Based On Edge Information Of Skin Color And Eye. Energy Procedia 13, 3678–3683 (2011)
[14] Hiremath, P.S., Danti, A.: Detection of multiple faces in an image using skin color information and lines-of-separability face model. International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI) 20(1), 39–61 (2006)

A Novel Approach for Prefetching of Web Pages through Clustering of Web Users to Reduce the Web Latency

G.T. Raju[1] and M.V. Sudhamani[2]

[1] Dept. of Computer Science and Engineering
RNS Institute of Technology, Bangalore-98
gtraju1990@yahoo.com
[2] Dept. of Information Science and Engineering
RNS Institute of Technology, Bangalore-98
mvsudha_raj@hotmail.com

Abstract. Web users are experiencing a long latency while retrieving the Web pages due to the amount of network traffic increased with the WWW expansion. Potential sources of latency are the Web servers' heavy load, network congestion, low bandwidth, bandwidth underutilization, and propagation delay. To solve the latency problem, prefetching technique that predicts the destination pages for user community has become critical to save the communication overhead. Prefetching means fetching of Web pages before the users request them so as to reduce the user perceived latency. A novel Cluster and Prefetch (CPF) approach is proposed in this paper. Experimental results shows that the CPF approach effectively reduces the user perceived latency without wasting the network resources with high prediction accuracy.

1 Introduction

Prefetching technique is motivated by the fact that, in general, once a user goes to a Web site; he/she generally browses around for several pages before leaving for another site. Since the user follows hyperlinks upon his/her interests, it is likely that links are not followed uniformly. It is possible to either predict each user's interest using cookies or mine a consensus of interests with some confidence from access log files recorded by the Web server. This information not only is valuable for the Web administrator to eliminate uninterested pages, or balance load among the servers, but also can help to improve Web-browsing time. Most prefetching techniques predict the Web page requests for individual user. These techniques can easily overload the network when there are large numbers of users. To overcome this, Cluster and Prefetch (CPF) approach is proposed in this paper. This approach uses the ART1 NN clustering algorithm for clustering the Web users. The prototype vector of each cluster gives a generalized representation of the Web pages that are most frequently requested by all the members of that cluster. Whenever a host connects to the server or a proxy, the proposed prefetching strategy returns the Web pages for the cluster to which the host belongs to. Advantage of the CPF approach is that the better network resource utilization by prefetching the Web pages for a user community rather than a single user, thereby improving the Web browsing time of user.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 983–989.
springerlink.com

2 Analytical Model of Web Prefetching

The objective of analytical model for web prefetching is to study the perceived latency in retrieving a web page by a browser/user with the given web traffic parameters such as – *number of users* and the *bandwidth of access link* etc.,

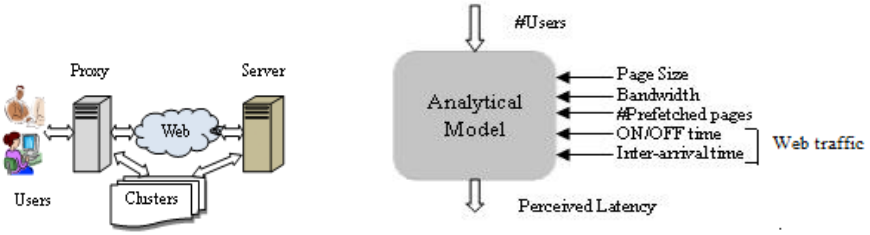


Fig. 1. (a) Perspective of Web Prefetching (b) Analytical Model of Web Prefetching

Perspective of Web prefetching and its analytical model are as shown in Fig. 1. When a user request for a page at particular time instance, the proxy identifies the web user and the cluster to which user belongs to. If the page is already prefetched by the prefetcher and is consistent with the original page on the remote server, the proxy sends the page to the user (Hit). If the proxy does not have a copy of the requested page, then the proxy prefetches the pages represented by the prototype vector of the cluster to which the user belongs to, sends the requested page to user (Miss) and keeps a copy in the cache. What is interested here is the page delivering latency or the response time which is defined as the time interval from the browser clicking an object to the requested object being displayed on the monitor. The response time depends on various parameters such as: *Web traffic (ON time, OFF time, Inter-arrival time), Page size , Number of prefetched objects, Number of users, and the Bandwidth of access link.*

Web traffic is modeled as *ON-OFF* process which is shown in Fig. 2, with *ON state* corresponding to the request and downloading time of the objects and the *OFF state* corresponding to the inactive time (viewing time).

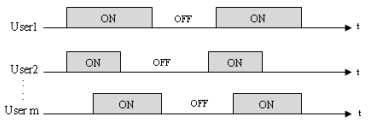


Fig. 2. ON-OFF Web traffic model

ON state is initiated by user click on the hyper link and the page is downloaded during this state. It is possible that the *ON state* lasts over multiple web request periods when the downloading of the last embedded object and the next HTML object overlaps. For example this can happen when the user requests a new object in the middle of the object download. *ON state* is found to follow a Weibull distribution whose probability density function is given by

$$f(x) = \frac{\beta}{\delta}\left(\frac{x}{\delta}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\delta}\right)^{\beta}\right] \quad for\ x > 0 \tag{4}$$

Where β is the shape parameter and δ is the scale parameter. Let x_2 be the continuous random variable represents the ON time with β=0.77 and δ = $e^{4.4}$. Intervals between adjacent requests to follow another Weibull distribution with β=0.5 and δ=1.5. Let x_2 be the continuous random variable represents inter-arrival time.

OFF state is the user viewing time or time that the user is away from the system. OFF state is found to follow Pareto distribution whose probability density function is given by

$$f(x) = \alpha k^{\alpha} x^{-\alpha-1} \tag{5}$$

Where α is the shape parameter and *k* is the scale parameter. Let x_3 be the continuous random variable represents the OFF time with α=0.58 and *k*=60.

Page size to follow another Pareto distribution with α= 1.3 and k=300. Let x_4 be the continuous random variable represents *page size*. Number of *prefetched objects* to follow Weibull distribution with β=0.5 and δ=1.5 represented by continuous random variable x_5 and number of *non-prefetched objects* to follow another Weibull distribution with β=0.9 and δ =4 represented by another continuous random variable x_6.

Let *m* be the *number of users* (250, 500, and 1000) and *B* be the *bandwidth* of the access link (512kbps, 2048kbps, and 4056kbps). Let *c* be a constant whose value is given by $c = B/m$.

The joint probability density function of the *response time (latency)* with prefetching is modeled as

$$f_{X_1X_2X_3X_4X_5}(x_1,x_2,x_3,x_4,x_5) = c[6.986\times10^{-4}(net)x_4^{2.1}x_5^{-0.5}e(-0.816x_5^{0.5})] \tag{6}$$

Where $$net = [2.59\times10^{-2}x_1^{-0.23}e(-0.03x_1^{0.77}) + 0.4078x_2^{-0.5}e(-0.816x_2^{0.5}) + 6.99x_3^{-1.58}] \tag{7}$$

The joint probability density function $f_{X_1X_2X_3X_4X_5}(x_1,x_2,x_3,x_4,x_5)$ satisfies the following properties:

(1) $f_{X_1X_2X_3X_4X_5}(x_1,x_2,x_3,x_4,x_5) \geq 0$ for all x_1, x_2, x_3, x_4, x_5

(2) $$\int_0^{86400}\int_{x_1}^{86400}\int_0^{86400}\int_0^{1048}\int_0^{1000} f_{X_1X_2X_3X_4X_5}(x_1,x_2,x_3,x_4,x_5)\,dx_1\,dx_2\,dx_3\,dx_4\,dx_5 = 1$$

(3) For any region *R* of p-dimensional space

$$P([X_1,X_2,X_3,X_4,X_5] \in R) = \iint_R \int\int\int f_{X_1X_2X_3X_4X_5}(x_1,x_2,x_3,x_4,x_5)dx_1dx_2dx_3dx_4dx_5$$

Similarly, the joint probability density function of the *response time (latency)* without prefetching is modeled as

$$f_{X_1X_2X_3X_4X_6}(x_1,x_2,x_3,x_4,x_6) = c[4.4246\times10^{-4}(net)x_4^{2.1}x_6^{-0.1}e(-0.287x_6^{0.9})] \tag{8}$$

The joint probability density function $f_{X_1X_2X_3X_4X_6}(x_1,x_2,x_3,x_4,x_6)$ satisfies the following properties:

(1) $f_{X_1X_2X_3X_4X_6}(x_1,x_2,x_3,x_4,x_6) \geq 0$ for all x_1,x_2,x_3,x_4,x_6

(2) $\int_0^{86400}\int_{x_1}^{86400}\int_0^{86400}\int_0^{1048}\int_0^{1000} f_{X_1X_2X_3X_4X_6}(x_1,x_2,x_3,x_4,x_6)dx_1dx_2dx_3dx_4dx_6 = 1$

(3) For any region *R* of p-dimensional space

$$P([X_1,X_2,X_3,X_4,X_6] \in R) = \iiint_R \int\int\int f_{X_1X_2X_3X_4X_6}(x_1,x_2,x_3,x_4,x_6)dx_1dx_2dx_3dx_4dx_6$$

Experiments have been conducted to show the comparative analysis on the user perceived latency from the analytical model and the actual trace for varying number of users and given bandwidth. Results presented in section 4.2 shows that the response time from the model and the trace is fairly well matched.

3 The CPF Model

The architecture of the CPF model is as shown in Fig. 3. The feature extractor module extracts each client's feature vector. ART1 clustering module identifies the group to which the client belongs to and returns that group's prototype vector. The prefetching module prefetches the URLs that are most frequently accessed by all the members (hosts) of that cluster represented by a prototype vector. The proxy server responds to the client with prefetched URLs. The prefetching accuracy is measured by predicting the URLs for each member of the cluster and then the prediction is verified with access logs recorded for the next *t* days (prediction period). The pseudo code for CPF approach is given in Fig. 4.

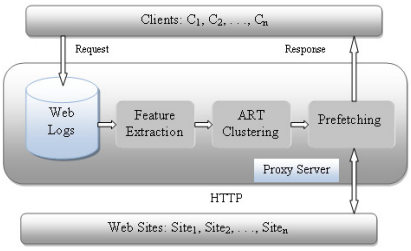


Fig. 3. Architecture of CPF Approach

```
CPF (Host_Id)
 {
        //Takes input: Host_Id of the host that request a URL
        //Cluster the hosts using ART Neural Network Clustering Algorithm;
  ART1_Clustering(P,ρ);
        /* P is the Array of Pattern vectors and ρ is the Vigilance Parameter. Let 'n' is the
number of clusters and C1, C2,….,Cn are the clusters represented by the prototype vectors.
The prototype vector for the kth cluster is of the form Tk = ( tk1, tk2, …, tkm) where tkj for
j=1,2,..m are the top-down weights corresponding to node k in layer F2 of the ART neural
network. */
   Initialize Count=0;
   Repeat for each cluster Ck of the n clusters
   If ( Host_ID is a member of cluster Ck)
   {
    Repeat for j = 1,2,….m
    If ( tkj = 1) {
     Prefetched_URLs[Count] = URLj
     Count++; }
   }
   Return Prefetched_URLs[] ;
 } // End of CPF prefetching scheme
```

Fig. 4. Pseudo code for CPF approach

4 Experimental Results

4.1 Performance of CPF Approach

Let *n* be the number of URLs prefetched, *k* be the number of URLs requested from the prefetched URLs, and *m* be the number of URLs requested by the user. Two parameters are used to assess the performance of CPF prefetching scheme:

1. *Hits*: The number of URLs requested from the prefetched URLs
2. *Accuracy*: The ratio of *Hits* to the number of prefetched URLs

Prediction accuracy is computed as

$$\frac{\sum_{r=0}^{k} URL_r}{\sum_{j=1}^{n} prefetched\_URL_j} \tag{9}$$

To verify the accuracy, URLs for each host are prefetched and compared with the predicted URLs over the next *t* days where *t* is the prediction period. Table 1 presents the results obtained by executing the CPF pseudo code on NASA Web log files for 6 days (1/Aug/1995 to 6/Aug/1995).

Table 1. Results of CPF approach

Cluster Id	Users in Clusters	User Id	No. of Requests	Number of URLs Prefetched	Hits	Prediction Accuracy
C1	U1,U2,U3, U4, U5	1	162	36	35	97.22
		2	184		33	91.66
		3	132		34	94.44
		4	168		35	97.22
		5	190		34	94.44
C2	U6,U7,U8, U9	6	200	62	60	96.77
		7	135		56	90.32
		8	146		58	93.54
		9	202		61	98.38
C3	U10,U11, U12	10	181	28	27	96.42
		11	126		26	92.85
		12	0		-	100
C4	U13,U14, U15, U16	13	186	24	22	91.67
		14	54		20	83.3
		15	147		22	91.67
		16	85		21	87.5

Fig. 5 shows the Web traffic (the number of URLs requested by each host/users). It is observed from the Fig. 6 that, the prediction accuracy ranges from 83.33 to 98.38%. The average prediction accuracy is 93.16% excluding the deviated one**.** Experimental results show that, the proposed CPF approach has very high prediction accuracy compared to other approaches [1,2,3,4].

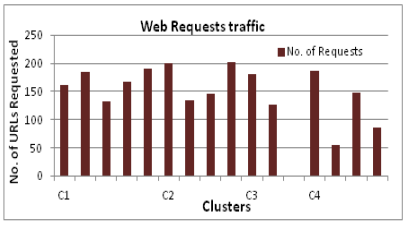


Fig. 5. Web Requests traffic

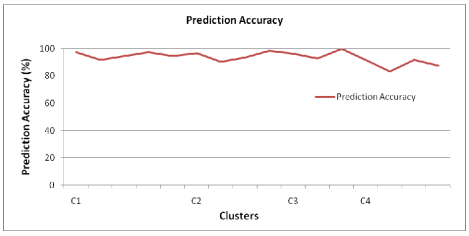


Fig. 6. Prediction Accuracy

4.2 Comparative Analysis

Comparative analysis has been made on the response time (latency) obtained from the analytical model and the trace collected from NASA Web site. The model parameters are derived from the trace data. To validate the model, the two metrics related to *user perceived latency* and *the traffic increase* are used. *Latency per page ratio* is the ratio of the latency that prefetching achieves to the latency with no prefetching. Lower the *latency ratio* value better the *performance*. *Traffic increase* denotes the bytes transferred through the network when prefetching is employed divided by the bytes transferred when prefetching is not employed. Lower the values of *traffic increase* better the *performance*. Cumulative Distribution Function (CDF) comparison of latency with and without prefetching for the trace and model with the given bandwidth is shown in Fig. 7. Average Latency Ratio v/s Traffic Increase is shown in Fig. 8. A summary of prefetching techniques is provided in Table 2.

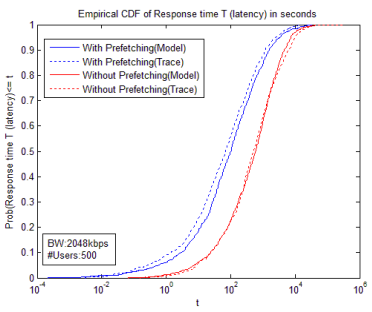


Fig. 7. CDF comparison of Response time

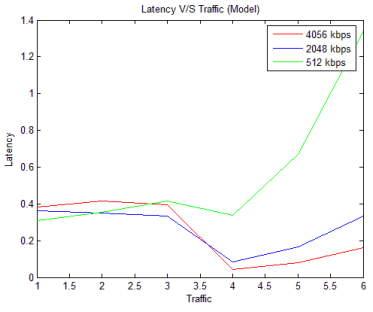


Fig. 8. Average Latency Ratio v/s Traffic Increase

Table 2. Summary of prefetching techniques

Reference	Method	Single User	Multiple Users	Prediction Accuracy
[1]	User profiles, weighted directed graph	Yes	-	50-75%
[2]	ANN, Keywords in Anchor Text	Yes	-	60-70%
[3]	PPM algorithm	Yes	-	40-73%
[4]	Top-10 prefetching approach	Yes	-	60%
[5]	Directed Graph	Yes	-	70-75%
[6]	Intelligent adaptive NN predictor	Yes		80%
Proposed method	***CPF method (thru clustering)***	***Yes***	***Yes***	***83-98%***

5 Conclusions

CPF approach that showed its usefulness in reasonable utilization of network resources through prefetching of Web pages for a community of users instead of a single user with an average prediction accuracy of 93.16% has been presented. Though the CPF approach results in substantial increase of network traffic, it effectively reduces the user perceived latency. Future research directions in this regard concern with the development of adaptive predictive systems that use hybrid approach such as use of statistical, neural, and Bayesian learning algorithms.

References

1. Loon, T.S., Bhargavan, V.: Alleviating the latency and bandwidth problem in WWW browsing. In: Proc. of the USENIX Symposium on Internet Technologies and Systems, USITS 1997 (1997)
2. Ibrahim, T., Xu, C.Z.: Neural Nets based Predictive prefetching to tolerate WWW latency. In: Proc. of the 20th International Conference on Distributed Computing Systems. IEEE, Taipei (2000)
3. Fan, L., Cao, P., Jacobson, Q.: Web prefetching between low-bandwidth clients and proxies: Potential and performance. In: Proc. of the Joint International Conference on Measurement and Modeling of Computer Systems, SIGMETRICS 1999, Atlanta, GA (1999)
4. Markatos, E.P., Chronaki, C.E.: A Top-10 approach to prefetching on the Web. In: Proc. of the 8th Annual Conference of the Internet Society, INET 1998, Geneva, Switzerland (1998)
5. Padmanabhan, V.N., Mogul, J.C.: Using predictive prefetching to improve WWW latency. Proc. of ACM Computer Communication Review 26(3), 23–36 (1996)
6. Tian, W., Choi, B., Phoha, V.V.: An Adaptive Web cache access predictor using Neural Network. In: Proc. of the 15th International Conference on Industrial and Engineering Applications of AI and Expert Systems, Cairns, Australia, pp. 450–459 (2002)

A Neuro-Fuzzy Based Intelligent Agent for Text Based Emotion Recognition

G. Sharada[1] and O.B.V. Ramanaiah[2]

[1] VNR Vignana Jyothi Inst. of Engg. & Technology,
HYD, India
sharada_g@vnrvjiet.in
[2] Jawaharlal Nehru Technological Univ. College of Engg.,
HYD, India
obvramanaiah@gmail.com

Abstract. Emotion recognition is an important aspect of Affective Computing. This paper deals with the development of an intelligent agent for automatic recognition of emotion from text based events. The approach chosen is Soft Computing and the architecture used is a Neuro-Fuzzy system. The input (event) string is divided into tokens which are then compared to a standard corpus (i.e.,Wordnet-Affect) of emotional keywords. The computed values of emotional weight and polarity are then processed by a Neuro-Fuzzy Controller, which generates the emotion underlying the event. The system considers Ekman's six basic emotions {Happiness, Despair, Disgust, Fear, Anger, Surprise}. The controller is trained to generate correct output through the backpropagation algorithm. The system is implemented using Java, and, Matlab is used for mathematical analysis. The performance of the system is graded based on the measures of Precision and Accuracy.

1 Introduction

Human emotion can be expressed through different kinds of medium like speech, images, facial expression, text etc. Human-machine communication will benefit from the ability to recognize emotions. If our aim in AI is to build systems that behave like human beings then it is necessary that we incorporate elements of both rational and emotional intelligence into the system. An emotional agent can be implemented by developing a computational approach.

Soft Computing is a synergistic integration of three computing paradigms: neural networks, fuzzy logic and probabilistic reasoning which provides a flexible framework to construct computationally intelligent systems. The role model for soft computing is the human mind.

Emotion recognition in text has a number of applications in various fields like Business (CRM), Education (Intelligent Tutoring Systems), Psychology, Text based communication environments (blogs, mails), Computational Linguistics.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 991–996.
springerlink.com

2 Related Work

Traditionally, research on the recognition of emotion from text was focused on the discovery and utilization of emotional keywords. Subasic & Huettner [9] classified a group of emotional words by manually scoring the emotion level for each word. Research in text based emotion processing has focused on emotion detection and classification for a sentence or document [2].

Aman and Szpakowicz, [3] studied how to identify emotion categories and intensity of emotions. A semantic network based emotion recognition mechanism [10] was proposed using emotional keywords, semantic/syntactic info., and emotional history in order to recognize the emotional state of a speaker. Emotion theories of Ortony, Clore & Collins [7] have been used widely in order to detect emotion within interaction systems.

A distributed architecture for a system simulating the emotional state of an agent acting in a virtual environment was presented by Aard-Jan Van Kesteren, Rieks Op Den Akker, Mannes Poel, Anton Nijholt [1]. Devillers et al.[5] found the most appropriate emotional state by calculating the conditional probability between the emotional keywords and emotional states.

3 System Architecture

The system comprises of the following units:

Input Unit
Event Processor

Emotional State Calculator (Neuro-Fuzzy System) with following five layers:

Input Layer
Input Fuzzy Layer
Conjunction Layer
Output Fuzzy Layer
Output Layer
Output Unit

Input Unit: The input is a textual string representing a real world event which is parsed and tokenized by this unit.

Event Processor: The Event Processor processes the tokens of the string and compares them to a corpus of emotional keywords and a lexicon of the English language . The emotional weights and the polarity of the matched tokens are converted into corresponding linguistic variables and fed to the emotional state calculator which is a neuro-fuzzy system.

Emotional State Calculator**:** In the emotional state calculator the input value is passed onto fuzzy set units, which then translate the value into a degree of membership as the activation level of a fuzzy set unit. The conjunction unit will take the minimum of the

inputs (degrees of membership) it receives from the input fuzzy set units. An output fuzzy unit takes the maximum of its inputs.

The activation function used in each layer is as follows:

The input layer: no activation function.

The input fuzzy layer: the trapezoidal function.

The conjunction layer: the min. function.

The output fuzzy layer: the mean-of-maxima function.

The output layer: no activation function.

Output Unit: The Output Unit generates the result of the system which is an emotion indication bar with the emotion and corresponding intensity, in response to the input external event.

4 Methodology

ALGORITHM: RECOGNITION TASK
Input the event string.
Divide the string into tokens using the tokenizer.
Compare the tokens to the database of emotional keywords.
If the tokens do not match any keyword then generate output as NEUTRAL.
For the tokens that match do the following operations:
Compute the emotional weight of the keyword using K-Window1 algorithm.
Compute the polarity of the keyword using K-Window2 algorithm.
Input the category to the Neural Network Selector.
Input the weight and polarity to the specific network selected.
Evaluate using the inference rules of Neuro-Fuzzy System.
Generate output in the form of "Emotion Indication Bar" or Emotion Chart.

Similarly an algorithm is used for the training the network using the Backpropagation concept using the feedback of Annotators as the standard.

The Neuro Fuzzy system (Fuzzy Controller) has four modules:

Fuzzification Module: converts crisp inputs into appropriate fuzzy sets.

Fuzzy Inference Engine: evaluates the fuzzy rules stored in the fuzzy rule base.

Fuzzy Rule Base: consists of fuzzy inference rules which formulate the knowledge of the problem.

Defuzzification Module: converts fuzzy set into a single crisp value.

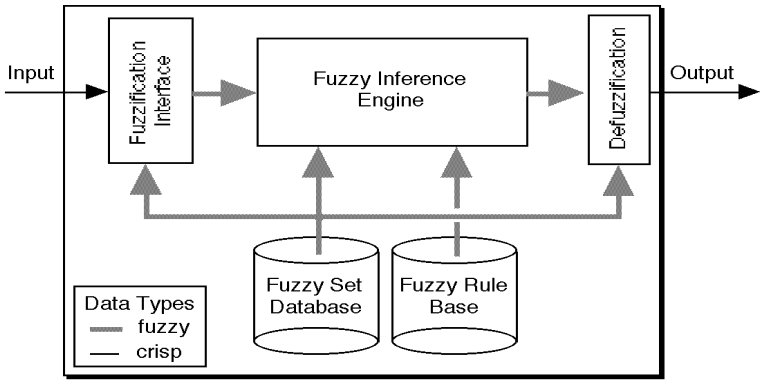


Fig. 1. Fuzzy Controller

Input Variable : Weight has three linguistic states {low, medium, high}
Input Variable : Polarity has two linguistic states {positive, negative}
Output Variable : Emotion has six linguistic states
{PositiveLow, PositiveMedium, PositiveHigh,
NegativeLow, NegativeMedium, NegativeHigh }

Fuzzification function : $f_w : [x, y] \rightarrow R$
$f_w : [0, 20] \rightarrow R$
$f_p : [a, b] \rightarrow R$
$f_p : [-1, 1] \rightarrow R$

Reasoning Schema :
Rule 1 : If < w, p > is A1 x B1 then e is C1.
:
Rule 6 : If < w, p > is A3 x B2 then e is C6.
Fact : < w, p >. is f_w (x0) x f_p(y0)
Conclusion : e is C.

The symbols Aj, Bj, Cj (j = 1, 2…. n) denote fuzzy sets of linguistic states.

The activation function is the trapezoidal function in each neuron given by :

$$\text{trapezoid}(x; a, b, c, d) = \max\left(\min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right)$$

The defuzzification function used is the Mean-Of-Maxima method and the defuzzified value, d_{MM} (C) is the average of all values in the crisp set

M = { z_k / C(z_k) = h (C) } and is denoted as d_{MM} (C) = $\sum_{z_k \in M}$ (z_k / | M |)

5 Results and Discussion

A sample of events involving the emotions was processed accordingly by NFS.

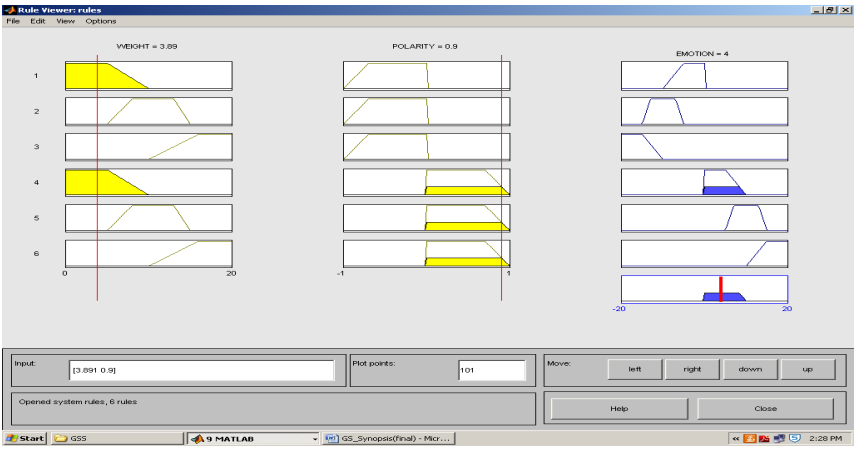


Fig. 2. Inference Rule Viewer for the Neuro-Fuzzy System

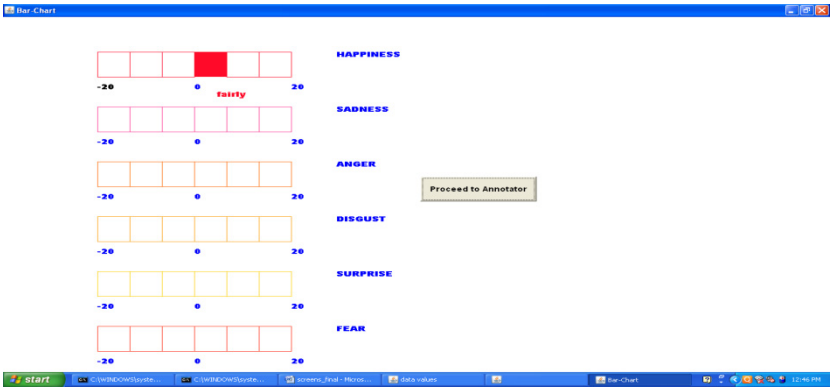


Fig. 3. Emotion Indication Bar for an event with LOW intensity of HAPPINESS

Performance of the system was measured by comparing the outputs of the system with those collected from the annotators and is as given below:

True positive, t_p represents correct result.
False positive, f_p represents unexpected result.
False negative, f_n represents missing result.
True negative, t_n represents correct absence of result.
For a sample of 39 events, the values are : $t_p = 32$, $f_p = 5$, $f_n = 2$, $t_n = 0$

$$\text{Precision} = t_p / (t_p + f_p) = 32 / (32 + 37) = 0.86$$

$$\text{Accuracy} = (t_p + t_n) / (t_p + t_n + f_p + f_n) = (32 + 0)/(32+0+5+2) = 0.82$$

5 Conclusion

In this paper we have put forth an approach to enable a computer to analyze events and detect the emotion associated with the event, if any. The system was implemented as a neuro-fuzzy system. The emotional values were computed through JAVA programs and fed to the simulated Neuro-Fuzzy System in MATLAB. The output was generated based on the membership functions and inference rules constructed. Training of the NFS is also provided using the backpropagation algorithm. If a computer system has elements of the rational mind and also the emotional mind then it will eventually take the processing power of computers to new heights.

References

1. van Kesteren, A.-J., op den Akker, R., Poel, M., Nijholt, A.: Simulation of emotions of agents in virtual environments using neural networks. Department of Computer Science, University of Twente, Enschede, Netherlands
2. Alm, Roth, Sproat: Emotions from text - machine learning for text based emotion prediction. In: Proceedings of HLTC, Vancouver, Canada (2005)
3. Aman, S., Szpakowicz, S.: Identifying expressions of emotion in Text, Speech and Dialogue, Master's thesis, Univ. of Ottawa, Canada (2007)
4. Strapparava, C., Mihalcea, R.: Learning to identify emotions in text. In: SAC (2008)
5. Devillers, L., Luniel, L.: Emotion detection in task-oriented spoken dialogues. In: Proceedings of the International Conference on Multimedia and Expo (2003)
6. Ghazi, D., Inkpen, D., Szpakowicz, S.: Hierarchical versus Flat Classification of Emotions in Text. In: Proc. of NAACL Workshop on CAAGET, California (2010)
7. Ortony, A., Clone, G.L., Collins, A.: The cognitive Structure of Emotions. Cambridge University Press, New York (1988)
8. Sivanandam, S.N., Deepa, S.N.: Principles of Soft Computing. Wiley India (P) Ltd. (2008)
9. Subasic, P., Huettner, A.: Affect analysis of text using fuzzy semantic typing. IEEE Trans. Fuzzy Systems (2001)
10. Wu, Chuang, Lin: Emotion recognition from text using semantic labels and separable mixture models. ACM Transactions (2006)

Feature Selection for Decoding of Cognitive States in Multiple-Subject Functional Magnetic Resonance Imaging Data

Accamma I.V. and H.N. Suma

BMS Research Center,
B M S C E, Bangalore
accamma@gmail.com, hnsuma@yahoo.co.in

Abstract. The last two decades have seen a surge in the interest in research based on Functional Magnetic Resonance Imaging (fMRI) data. Decoding of cognitive states based on fMRI activation profiles has become a very active topic in this area. fMRI data is very high dimensional and noisy. However, there is a dearth of datasets to work on. Sharing of learning by analyzing datasets drawn across multiple subjects in an experiment can help in increasing the amount of data we have for analysis. Decoding of cognitive states using classifiers trained across multiple subjects is a challenging task because of differences in anatomy and cognition. Selecting features to analyze from the dataset is a key step in the analysis of fMRI data. In this paper we apply PCA, ICA and five non-linear dimensionality reduction techniques to the fMRI data. The aim of this work is to analyze which technique can provide the best feature selection to capture the commonality across multiple subjects. The reduced datasets are then used to train classifiers to solve a multiple-subject decoding problem.

1 Introduction

fMRI is a technique used to capture the images of the activity in the brain. It captures MR images measuring the changes in blood oxygenation level. Since blood oxygenation varies according to the levels of neural activity, these differences can be used to detect brain activity. In the last few years, fMRI has emerged to become the preferred method to map the cognitive states of a human subject to the specific functional areas of the brain.

Most of the literature in fMRI analysis is dedicated to discovering the neural correlates of various activities. More recently there is increasing interest in 'reverse inference' or a predictive method of fMRI analysis. Here motor, sensory or cognitive states of a subject are predicted based on the corresponding fMRI data. Most of the work done so far focuses on reverse inferring cognitive states of a subject based on analysis of data pertaining to the same subject. A more challenging problem is to be able to infer the cognitive states of a particular subject based on the data drawn from other subjects. This is challenging because different brains have different sizes and shapes, and because different people may generate different brain activations for the same cognitive state.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 997–1004.
springerlink.com

Despite differences in anatomical structure and degree of cognitive response, it has been found that there are certain commonalities in the mapping of neural response to stimuli. This paper aims to explore feature selection as a means of identifying the commonality in neural response in data drawn from multiple subjects. Feature selection, or the selection of dimensions to analyze, is a critical step in reverse inference. We reduce the dimensions of the fMRI data using various dimensionality reduction techniques and analyze which technique can best abstract the brain activations to use in the information drawn across subjects.

The rest of the paper is organized as follows - Section 2 contains survey about related work done. Section 3 discusses briefly various feature selection used in our analysis. Section 4 provides details about the dataset and the methodology used for comparative analysis. Section 5 details the steps in the evaluation technique used. Section 6 contains the analysis of results followed by the conclusion in Section 7.

2 Related Work

A great literature of research has been generated in the area of fMRI data analysis. We focus mainly on the research related to reverse inference. Reverse inference is a type of problem that is addressed very well in a machine learning paradigm. Starting with fMRI activation profiles, machine learning classifiers have been designed to map and distinguish neural responses to spatial characteristics such as the direction of motion [10] or the orientation of stimulus [4][9], lying during a card game [2] decision making [20] mental arithmetic [12] and semantic categories[13][22]. Apart from two category distinctions, researchers have successfully used machine learning classifiers to distinguish between multiple categories of objects by reverse inference [1]. Many studies have mapped neural responses to the building blocks of visual perception, There have been attempts made toward image reconstruction[11][15][24] by measurement of brain activity and also an effort to reconstruct the visual experience of movies[16].

One of the very impressive contributions was made by Mitchell et al., where they used a combination of machine learning techniques and neuro-imaging to decode semantic content [14]. The researchers were able to extrapolate from the brain imaging data collected, using machine learning, the brain activation for words the participants had not viewed during the experiment. The predictions were generated based on the semantic features of words as drawn from an online trillion-token Google text corpus. This work was extended by Chang et al to classify not just words, but word pairs containing adjective-noun phrases, and obtain a prediction with accuracy above chance levels [8]. Classification analysis shows that the data from distributed pattern neural activity contains sufficient signals to decode differences among phrases. These techniques, methodologies and conclusions, as well as ethical issues that may arise in fMRI data analysis have been discussed in detail by many reviewers[18][3][17] .

Most of the predictive approaches in fMRI data analysis have been limited in the sense that they can be applied only individually to a particular subject's data from a particular fMRI study. There have been very few efforts made in the area of multiple subject analysis. Wang et al. explored two techniques of applying multiple-subject

feature selection to fMRI data. Region of Interest (ROI) mapping and transformation to Talairach coordinates were used to classify a two category dataset [26]. Another study by Just et al., using factor analysis revealed significant cross participant commonality in neural representations of word meaning[7]. They were able to predict the activation patterns for previously unseen words based on factored abstractions of the content of the representation. Factor analysis could determine the common semantic dimensions underlying the activations across participants. Rustandi et al., used Canonical Correlation analysis(CCA) to find the common dimensions among different data sets[6] They created a model called CCA-mult and proved that the greater the number of subjects and studies involved in an analysis, the more accurate the predictions will be.

As mentioned above the reverse inference problem fits very well into a machine learning paradigm. Given that fMRI data is high dimensional and noisy, feature selection forms a crucial step in the analysis of fMRI data. ROI, activity based selections, Principal Component Analysis (PCA), factor analysis and CCA are some of the methods that have been used so far. Non-linear feature selection techniques have not been explored much in the context of fMRI data.

3 Feature Selection

Every input image in fMRI may contain thousands of voxels. It is important to choose the right subset of these voxels to use in further analysis. We explored a variety of approaches to reducing the dimension of the input data, concentrating mainly on non-linear techniques. We compare all the feature selection methods to the results obtained by using the 500 most active voxels as the feature vector.

The first approach - PCA - is one of the most popular techniques for dimensionality reduction. It represents the variance in the data as new dimensions. This is done by creating a linear mapping of the principal eigenvectors of the covariance matrix of the data. PCA has been applied to a wide variety of problems. However, its effectiveness is limited by its global linearity.

Independent component analysis (ICA) is a blind signal separation technique to identify the statistically independent components in the data. For the purposes of this study, we have considered fast ICA implementation [5] provided by Prof. Aapo Hyvarninen of the Helsinki University of Technology. In the Isomap technique, a graph is constructed for every data point connecting it, to its k nearest neighbours. Using the graph, a distance matrix is constructed by computing the shortest path between the data points. By applying multidimensional scaling to the distance matrix, lower dimensions are obtained. We use the implementation suggested by Tenenbaum [23].

Local Linear Embedding (LLE) is a non linear dimensionality reduction technique like Isomap. The difference is that it tries to preserve the local properties of a data point. It does this by representing the data point as a linear combination with its k nearest neighbors. We use the implementation by Roweis [21]. In the Diffusion map method, a graph of the data points is constructed first. The proximity of the data points is measured by a Markov random walk defined on the graph for a number of steps.

The Kernel PCA works like the traditional linear PCA except that it tries to find the principal Eigen vectors of the Kernel matrix constructed using a Kernel function. Multilayer Autoencoders are neural networks with odd number of hidden layers. They try to learn a non-linear mapping between the high dimensional space and the low dimensional space. The training is done using Restricted Boltzmann Machines. All of the above three methods are implemented using the Dimensionality reduction toolbox for Matlab [25].

4 Dataset and Methodology

We evaluated our feature selection techniques using the previously analyzed dataset by a fMRI study by Mitchell et al. In this study nine healthy, college-age participants were presented with stimuli corresponding to 60 different concrete noun words which can be grouped to 12 semantic categories (animals, body parts, buildings, building parts, clothing, furniture, insects, kitchen items, tools, vegetables, vehicles, and other man-made items). The data that has been made available has been processed and a single fMRI mean image has been created for each of the 360 trials (60 words × 6 runs) This has been achieved by taking the mean of the images collected at 1 second intervals starting at 4 seconds after stimulus onset up to 7 seconds. We created a representative fMRI image for each stimulus by computing the mean fMRI response over its six presentations, and the mean of all 60 of these representative images was then subtracted from each.

It is assumed that there are intermediate semantic features that represent the meaning of each word underlying the brain activations associated with the thinking about that word. The semantic features used here are Intel 218 features. Intel 218 is the set of answers to 218 yes/no questions about various objects gathered by Dean Pomerleau at Intel Research Pittsburgh. They characterize the semantic information present in words, as answers to questions about properties of the objects [19]. To obtain the common abstractions of all the subjects, all the above mentioned dimensionality reduction methods are applied to the dataset. This reduces the dimensionality maintaining only the essential structures in the fMRI data.

5 Evaluation

The experiment consists of data from nine subjects. We try to predict the neural activations of one subject based on the data from other subjects. The model was trained using leave-two-out cross validation. This provides us with 1770 unique pairs per subject.

Each trained model was first tested by letting it predict the two held-out words from a particular subject. The match between the two predicted and the two held-out words was determined by which match had higher cosine similarity. For training we use the data from the other eight subjects. We use Mitchell's relation that the predicted activation at voxel v is given by the product of the semantic features and a learned weight.

$$Y_v = \sum_{i=1}^{n} W_{vi} \, f_i(word)$$

Where n is the number of words, W is the learned Weight, f(word) is the feature vector with respect to the word.

We perform a Multiple Regression on each of the subject's reduced datasets to obtain the weights. We consider one unique pair from one held-out subject. We consider the average of the learned weights from other eight subjects and the activation profile corresponding to one word from the unique pair. Using these we predict the semantic features corresponding to the word. We compare the predicted features with the true feature vectors using the cosine similarity metric. Between the two words, if the distance between the true and predicted features is large enough to distinguish between them, then we consider it a correct prediction. Aggregate of the scores have been computed for each of the subjects.

6 Results

This section describes two types of analyses and results.

6.1 Decoding Based on all Active Voxels from Other Subjects

To predict the neural activations to words of a particular subject, we train the classifier using all responses from other subjects including the held-out words. Therefore there are 60 words from each of the subjects used for training. These results (Table-1) indicate how well a feature selection technique can encapsulate the fMRI activation of a particular word so that it is common to all subjects. The response of each of the nine subjects is predicted based on the other eight subjects. From earlier studies, the accuracy of the prediction for subjects 1 through 9 for single subject decoding is 0.83, 0.76, 0.78, 0.72, 0.78, 0.85, 0.73, 0.68 and 0.82. The accuracies of our multiple subject predictions do not match up to single subject accuracies.

Table 1. Accuracy of decoding multiple subject fMRI data (in percentage)

Method	Subj1	Subj2	Subj3	Subj4	Subj5	Subj6	Subj7	Subj8	Subj9
Most active 500	55	64	50	51	53	48	53	60	53
PCA	54	51	54	58	50	46	52	46	56
ICA	56	56	53	56	49	41	57	58	50
Isomap	55	60	53	49	62	47	51	59	50
Diffusion Maps	48	47	55	53	50	44	46	58	48
LLE	44	43	48	49	49	38	44	48	45
Kernal PCA	54	45	53	53	51	51	47	53	55
AutoEncoder RBM	61	52	45	45	42	42	57	51	58

This is because of the differences in anatomy and degree of cognitive response across subjects. Subject-6 performs well in single subject studies but worse than other subjects in multi subject studies. This could be because of increased variability

between this subject and other subjects. Ignoring Subject-6, if the averages of the accuracies are considered we see that LLE and Diffusion maps do not perform very well. ICA and Isomap perform better than other techniques, but not significantly better than choosing the 500 most active voxels. Therefore, there is a need to refine these methods.

6.2 Decoding Based on Feature Extraction from Each of the Active ROIs

We consider the active ROIs from the dataset. We apply feature selection on voxels from each ROI. The features extracted from each of the ROIs are concatenated together for each of the subjects. Then the analysis proceeds as above with the transformed data. The neural responses of each of the nine subjects are predicted using this technique and the accuracies are calculated (Table-2).

Table 2. Decoding accuracy based on active ROI (in percentage)

Method	Subj1	Subj2	Subj3	Subj4	Subj5	Subj6	Subj7	Subj8	Subj9
PCA	51	41	52	58	57	49	42	56	50
ICA	56	51	49	56	55	53	52	48	49
Isomap	54	52	57	54	54	57	58	48	52
Diffusion Maps	46	50	54	55	61	52	50	52	52
LLE	55	49	47	54	47	57	50	43	51
Kernal PCA	45	48	41	47	47	48	58	58	49
AutoEncoder RBM	51	46	50	46	48	46	45	54	48

We see that applying feature selection to the ROIs does not offer a significant advantage. One noteworthy improvement is that Subject-6 performs as well as others. Therefore, this technique is robust to variations across subjects. As opposed to previous analysis, using Diffusion Maps best extracts features from ROIs compared with other techniques.

7 Conclusion and Future Work

This paper aimed to explore some feature selection methods for capturing the commonality across multiple-subject fMRI data. In our analysis above, whole brain feature extraction using ICA, Isomap and ROI feature extraction using Diffusion maps predict with accuracies above 50%. For application in analysis of patients exhibiting cognitive disorders and brain impairment due to injury, higher accuracy levels are required. Therefore, there is a need to work on refined feature selection methods. The classifier we used was based on simple multiple regression. Support Vector Machines and Gaussian Naive Bayes classifiers have been proven to perform well on fMRI data. We intend to pursue further research in this direction.

References

1. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fMRI) 'brain reading': detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270 (2003)
2. Davatzikos, C., et al.: Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. Neuroimage 28, 663–668 (2005)
3. Haynes, J., Rees, G.: Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7, 523–534 (2006)
4. Haynes, J.D., Rees, G.: Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience 8, 686–691 (2005)
5. Hyvarinen, A.: Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. IEEE Transactions on Neural Networks 10(3), 626–634 (1999), http://research.ics.tkk.fi/ica/fastica/ (accessed March 15, 2012)
6. Rustandi, I., Just, M.A., Mitchell, T.M.: Integrating multiple- study multiple-subject fMRI datasets using canonical correlation analysis. In: Proceedings of the MICCAI Workshop: Statistical Modeling and Detection Issues in Intra- and Inter-Subject Functional MRI Data Analysis (2009)
7. Just, M.A., Cherkassky, V.L., Aryal, S., Mitchell, T.M.: A Neurosemantic Theory of Concrete Noun Representation Based on the Underlying Brain Codes. PLoS ONE 5(1), e8622 (2010)
8. Chang, K.K., et al.: Quantitative modeling of the neural representation of adjective-noun phrases to account for fMRI activation. In: Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp. 638–646 (2009)
9. Kamitani, Y., Tong, F.: Decoding the visual and subjective contents of the human brain. Nature Neuroscience 8, 679–685 (2005)
10. Kamitani, Y., Tong, F.: Decoding seen and attended motion directions from activity in the human visual cortex. Current Biology 16, 1096–1102 (2006)
11. Kay, K.N., et al.: Identifying natural images from human brain activity. Nature 452, 352 (2008)
12. Knops, A., et al.: Recruitment of an area involved in eye movements during mental arithmetic. Science 324, 1583–1585 (2009)
13. Mitchell, T.M., et al.: Learning to decode cognitive states from brain images. Machine Learning 5, 145–175 (2004)
14. Mitchell, T.M., et al.: Predicting human brain activity associated with the meanings of nouns. Science 320, 1191 (2008)
15. Miyawaki, Y., et al.: Visual image reconstruction from human brain activity using a combination of multiscale local image decoders. Neuron 60, 915–929 (2008), doi:10.1016/j.neuron.2008.11.004
16. Nishimoto, et al.: Reconstructing Visual Experiences from Brain Activity Evoked by Natural Movies. Current Biology (2011), doi:10.1016/j.cub.2011.08.031
17. Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V.: Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends in Cognitive. Science 10, 424–430 (2006)
18. Otoole, A.J., et al.: Theoretical, statistical, and practical perspectives on pattern-based classification approaches to functional neuroimaging analysis. Journal of Cognitive Neuroscience 19, 1735–1752 (2007)
19. Palatucci, M., Pomerleau, D., Hinton, G., Mitchell, T.: Zero-shot learning with semantic output codes. In: NIPS 2009 (2009)

20. Pessoa, L., Padmala, S.: Quantitative prediction of perceptual decisions during near-threshold fear detection. Proceedings of the National Academy of Sciences, USA 102, 5612–5617 (2005)
21. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500), 2323–2326 (2000), http://cs.nyu.edu/~roweis/lle/ (accessed March 8, 2012)
22. Shinkareva, S.V., et al.: Using fMRI Brain Activation to Identify Cognitive States Associated with Perception of Tools and Dwellings. PLoS ONE 3(1), e1394 (2008), doi:10.1371/journal.pone.0001394
23. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290(5500), 2319–2323 (2000), http://isomap.stanford.edu/ (accessed March 10, 2012)
24. Thirion, B., et al.: Inverse retinotopy: inferring the visual content of images from brain activation patterns. Neuroimage 33, 1104–1116 (2006)
25. Van der Maaten, L.J.P., Postma, E.O., Van den Herik, H.J.: Dimensionality reduction: A comparative review, Technical Report TiCC TR 2009-005
26. Wang, X., Hutchinson, R., Mitchel, T.M.: Training fMRI Classifiers to Detect Cognitive States across Multiple Human Subjects. In: NIPS 2003 (2003)

A New Approach to Partial Image Encryption

Parameshachari B.D.[1] and K.M.S. Soyjaudah[2]

[1] Dept. of Electronics and Communication Engineering,
JSS Academy of Technical Education, Mauritius,
Avenue Droopanath Ramphul, Bonne Terre, Vacoas, Mauritius
parameshbkit@gmail.com
[2] Dept. of Electrical and Electronic Engineering,
University of Mauritius, Reduit, Mauritius

Abstract. The traffic of digital images and video has grown rapidly in the internet. Security becomes important for several applications like military image database, confidential video conferencing, medical images, etc. Several techniques have been developed for textual data but are not appropriate for images and video with huge amount of file size. In traditional image and video content protection schemes, called fully layered, the whole content is first compressed. Then, the compressed bitstream is entirely encrypted using a standard cipher. The specific characteristics of this kind of data make standard encryption algorithms inadequate. Partial encryption is a recent approach to reduce the encryption time of images in distributed network. Partial image encryption is used to reduce the amount of data to encrypt while achieving a sufficient and inexpensive security. The proposed approach involves two ways, the first is by pixel value manipulation and other second is by using SCAN mapping method. The PSNR comparison of proposed technique with the existing partial image encryption techniques shows that proposed technique gives better security than the existing techniques.

Keywords: mapping, partial image encryption, PSNR, security, SCAN.

1 Introduction

The increased popularity of multimedia applications has demanded a certain level of security. in some applications, it is relevant to hide the content of a message when it enters an insecure channel. the initial message prepared by the sender is then converted into cipher text prior to transmission. the process of converting plain text into cipher text is called encryption. the encryption process requires an encryption algorithm and a key. the process of recovering plain text from cipher text is called decryption. because common encryption methods generally manipulate an entire data set, most encryption algorithms tend to make transfer of information more costly in terms of time and sometimes bandwidth. traditionally, an appropriate compression algorithm is applied to the multimedia data and its output is encrypted by an independent encryption algorithm. this process must be reversed by the receiver.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1005–1010.
springerlink.com

Unfortunately, The Processing Time For Encryption And Decryption Is A Major Factor In Real-Time Image Communication. In Addition, The Processing Time Required For Compression And Decompression Of An Associated Image Data Is Important. Encryption And Decryption Algorithms Are Too Slow To Handle The Tremendous Amount Of Data Transmitted. One Difference Between Text Data And Image Data Is That The Size Of Image Data Is Much Larger Than The Text Data. The Time Is A Very Important Factor For The Image Encryption. We Find It At Two Levels, One Is The Time To Encrypt And Other Is The Time To Transfer Images. To Minimize The Time, The First Step Is To Choose A Robust, Rapid And Easy Method To Implement Cryptosystem. The Other Important Criteria Concerns The Method Of Compression Is That To Decrease The Size Of Images Without Loss Of Image Quality [1]. One Possible Solution Is A System Of Partial Encryption, Encrypting Only The Smallest Portion Of The Data That Makes The Entire Data Set Unusable. Partial Encryption Is A Recent Approach To Reduce The Computational Requirements For Huge Volumes Of Multimedia Data. Partial Encryption Is Currently An Important Research Area.

The Rest Of This Paper Is Organized As Follows: Section 2 Explains The Related Work. The Results Are Described In Section 3.This Paper Is Concluded By Providing The Summary Of The Present Work In Section 4.

2 Related Work

2.1 Partial Image Encryption Techniques

The encryption algorithms, which have been originally developed for text data, are not suitable for securing many real time algorithms, which have been originally developed for text data, are not suitable for securing many real time multimedia applications because of large data sizes. Software implementations of ciphers are usually too slow to process image and video data in commercial systems. Hardware implementations on the other hand, add more cost to service providers and consumer electronics device manufacturers. Recent trend is to minimize the computational requirements for secure multimedia distribution by “partial encryption” where only parts of the data are encrypted.

- **Cheng and Li, 2000**

Cheng and Li [2] proposed partial encryption methods that are suitable for images compressed with two specific classes of compression algorithms: Quadtree compression algorithms and wavelet compression algorithms based on zerotrees.

- **Droogenbroeck and Benedett, 2002**

In 2002, Droogenbroeck and Benedett proposed the selective encryption methods for uncompressed (raster) images and compressed (JPEG) images [3]. According to Droogenbroeck and Benedett, at least 4-5 least significant bitplanes should be encrypted to achieve the satisfactory visual degradation of the image. In this scheme encryption ratio vary from 50 to 60%. It is fast as XOR operation takes less time. It is not robust against cryptanalysis attack. So, security level is low.

- **Pommer and Uhl, 2003**

In 2003, Pommer and Uhl, proposed wavelet packet based compression instead of pyramidal compression schemes in order to provide confidentiality. Header information of a wavelet packet image coding scheme that is based on either a uniform scalar quantizer or zero trees is protected: it uses AES to encrypt only the sub band decomposition structure. In this approach the encoder uses different decomposition schemes with respect to the wavelet packet sub band structure for each image. it is based on AES encryption of the header information of wavelet packet encoding of an image, this header specifies the sub band tree structure[4].

- **Roman Pfarrhofer and Andreas Uhl,2005**

In 2005, Roman Pfarrhofer and Andreas Uhl [5], proposed selective encryption of JBIG encoded visual data exploiting the interdependencies among resolution layers in the JBIG hierarchical progressive coding mode. Contrasting to earlier ideas when selectively encrypting a subset of bitplanes, they able to show attack resistance even in case of restricting the amount of encryption to 1% – 2% of the data only. The extremely low amount of data required to be protected in their technique also allows the use of public-key cryptography thereby simplifying key management issues.

- **Y.V. Subba Rao, Abhijit Mitra, and S.R. Mahadeva Prasanna,2006**

In 2006, Y.V. Subba Rao, Abhijit Mitra, and S.R. Mahadeva Prasanna[6], proposed partial encryption of image using pseudo random sequences with simple hardware. According to Y.V. Subba Rao, Abhijit Mitra, and S.R. Mahadeva Prasanna partial encryption method achieves the same security with the improvement in processing speed. The performance of the method mainly depends on the differentiation of correlated and uncorrelated information in the image.

- **Yang Ou, Chul Sur, and Kyung Hyune Rhee,2007**

In 2007, Yang Ou, Chul Sur and Kyung Hyune Rhee [7] proposed two types of region-based selective encryption schemes to achieve secure access for medical images. The first scheme randomly flips a subset of the bits belonging to the coefficients in a Region of Interest inside of several wavelet sub-bands, The second scheme employs AES to encrypt a certain region's data in the code-stream. Both of two schemes support backward compatibility so that an encryption-unaware format-compliant player can play the encrypted bit-stream directly without any crash.

- **Hammed A younis,Turki Y Abdalla and Abdulkareem Y Abdalla,2009**

In 2009, Hammed A younis,Turki Y Abdalla and Abdulkareem Y Abdalla [8],proposed only 6.25%-25% of the original data is encrypted for four different images, resulting in a significant reduction in encryption and decryption time. They able to show the low computation complexity and keep an unchanged compression ratio, it could be a nice solution for real time image encryption.

- **Ju-Young Oh, Dong-Il Yang, Ki-Hwan Chon,2010**

Ju-Young Oh, Dong-Il Yang, Ki-Hwan Chon [9], Proposed to expand the advanced encrytion standard (AES)-Rijndael with five criteria: the first is the compression of plain data, the second is the variable size of the block, the third is the selectable round, the fourth is the optimization of software implementation and the fifth is the selective function of the whole routine.

- **Jay M. Joshi, Upena D. Dalal,2011**

Jay M. Joshi, Upena D. Dalal [10], proposed to gain a deep understanding of video data security on multimedia technologies and to provide security for real time video applications using selective encryption for H.264/AVC. The selective encryption in different levels provides encryption of intra-prediction mode, residue data, inter-prediction mode or motion vectors only.

2.2 SCAN Based Encryption Method

This method converts a 2D image into a 1D list, and employs a SCAN language [11]. SCAN is the method for image encryption together with information hiding. This algorithm is based on permutations of the image pixels and replacement of the pixel values. The encryption power of the SCAN method is based on the very large number of private keys. The SCAN language includes an alphabet consisting of primitive scanning techniques, as, letters, and a simple grammar to manipulate and combine the alphabet symbols by generating new scanning patterns (words) from simple ones. The development of SCAN provides an efficient approach to the problem of modeling and generating all accessing algorithmic patterns of an image of nxn. Therefore, the size of the image is very large, and thus it is inefficient to encrypt or decrypt the image directly.

3 Results and Discussion

In this section, a number of experiments which are used to examine our proposed techniques will be presented. The techniques were programmed in MATLAB version7.0. To evaluate the proposed techniques examined by the following aspects:

i) Security: Security in this work means confidentiality and robustness against attacks to break the images. It is obvious that the goal is not 100% security, but the proposed techniques that make them difficult to cryptanalysis.

ii) Speed: Less data (important part) to encrypt means less CPU time required for encryption. So, partial encryption techniques are used to reduce encryption and decryption time.

iii) **PSNR:** Peak Signal to Noise Ratio (PSNR) measures are estimate of the quality of a reconstructed image compared to an original image.

Table 1. PSNR Value for the proposed encryption Techniques

Image / Techniques used	home image	step image	aeroplane image
only pixel value manipulation	20.909	18.101	19.899
only scan mapping	13.416	11.991	15.05
pixel manipulation + scan mapping	2.908	2.9831	5.001

Table 2. Processing time for Partial image encryption using pixel value manipulation and scan mapping

Image	Amount	Encryption time (second)	Decryption time (second)
home image	Full (100%)	0.212	0.171
	40%	0.139	0.101
	15%	0.072	0.072
	5%	0.025	0.043
step image	Full (100%)	0.479	0.365
	40%	0.318	0.254
	15%	0.103	0.187
	5%	0.078	0.142
aeroplane image	Full (100%)	0.683	0.517
	40%	0.475	0.385
	15%	0.283	0.179
	5%	0.176	0.091

It is observed that the proposed scrambling techniques give varying amount of transparency. This allows the degraded vision to all the users while high quality viewing of the multimedia content to authorized users only. The PSNR value obtained by these proposed partial encryption techniques are shown in table 1 and it can be observed that changing the coefficient values by pixel manipulation gives a fairly low value of peak signal to noise ratio while changing only the pixel value manipulation further lowers down the PSNR. It is observed through simulations that combining the first two techniques i.e. SCAN mapping and the pixel value manipulation of selected coefficients in their binary form will further improve the results, producing degradation in the visual results. A drastic change in the PSNR value is observed with both the techniques are combined. This is due to increased number of selected coefficients. Although a user can view the degraded version of the original image but the decrease in PSNR value makes it difficult for an intruder to retrieve the original image without having the key.

Processing time for the proposed method shown in table 2. Proposed method reduces the encryption time and decryption time in the communication process.

4 Conclusions

From the experimental results we conclude that the proposed image encryption method gives very good results. We implemented a partial image encryption technique involves two methods, the first is by pixel value manipulation and other second is by

using SCAN mapping method. The development and implementation of partial image encryption based on percentage of encryption. In partial encryption, only part of image (Important part) is encrypted whereas the remaining part (unimportant part) is transmitted without encryption. The various proposed techniques offer different levels of transparency, security and consumption of computational sources consumed. As per Human Visual System (HVS), the degraded version of original image by the proposed technique gives a sufficient level of transparency. PSNR calculation shows that the proposed technique is better than the existing partial image encryption techniques. Hence, a suitable technique can be adopted which is best suitable to the application.

References

1. Borie, J., Puech, W., Dumas, M.: Cryptography Compression System for Secure Transfer of Medical images. In: 2nd International Conference on Advances in Medical Signal & Information Processing (MEDSIP 2004) (September 2004)
2. Cheng, H., Li, X.: Partial Encryption of Compressed Images and Video. IEEE Transactions on Signal Processing 48(8), 2439–2451 (2000)
3. Van Droogenbroeck, M., Benedett, R.: Techniques for a Selective Encryption of Uncompressed and Compressed Images. In: Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS) 2002, Ghent, Belgium, September 9-11 (2002)
4. Podesser, M., Schmidt, H.P., Uhl, A.: Selective Bitplane Encryption for Secure Transmission of Image Data in Mobile Environments. In: 5th Nordic Signal Processing Symposium, on Board Hurtigruten, Norway, October 4-7 (2002)
5. Pfarrhofer, R., Uhl, A.: Selective Image Encryption Using JBIG. In: Dittmann, J., Katzenbeisser, S., Uhl, A. (eds.) CMS 2005. LNCS, vol. 3677, pp. 98–107. Springer, Heidelberg (2005)
6. Rao, Y.V.S., Mitra, A., Mahadeva Prasanna, S.R.: A Partial Image Encryption Method with Pseudo Random Sequences. In: Bagchi, A., Atluri, V. (eds.) ICISS 2006. LNCS, vol. 4332, pp. 315–325. Springer, Heidelberg (2006)
7. Ou, Y., Sur, C., Rhee, K.-H.: Region-Based Selective Encryption for Medical Imaging. In: Preparata, F.P., Fang, Q. (eds.) FAW 2007. LNCS, vol. 4613, pp. 62–73. Springer, Heidelberg (2007)
8. Younis, H.A., Abdalla, T.Y., Abdalla, A.Y.: Vector Quantization Techniques For Partial Encryption of Wavelet Compressed Digital Images. Iraq J. Electrical and Electronic Engineering (2009)
9. Oh, J.-Y., Yang, D.-I., Chon, K.-H.: A Selective Encryption Algorithm Based on AES for Medical Information. The Korean Society of Medical Informatics (2010)
10. Joshi, J.M., Dalal, U.D.: Selective Encryption using ISMA Cryp in Real Time Video Streaming of H.264/AVC for DVB-H Application, World Academy of Science, Engineering and Technology 79 (2011)
11. Parameshachari, B.D., Chaitanyakumar, M.V.: Image Security using SCAN Based Encryption Method. In: 42nd IETE Mid-term Symposium on Telecom Paradigms - Indian Scenario, Bangalore, pp. 115–118 (April 2011)

Fuzzy Number with Nonlinear Membership Functions to Provide Flexibility in a Multi Objective Travelling Salesman Problem

Atul Kumar Tiwari[1], Cherian Samuel[1], Vinay Pratap Singh[2], and Vivek Saraswati[2]

[1] Institute of Technology, Banaras Hindu University Varanasi
atultiitb@gmail.com,
csamuel.mec@itbhu.ac.in
[2] Department of Computer Science, Banaras Hindu University Varanasi
{bhu.vinay,vicky.bhu17}@gmail.com

Abstract. Travelling Salesman Problem (TSP), as extensively discussed in literature is an NP hard problem and among the most challenging problems in operations research, industrial engineering and computational mathematics, which has been deciphered and scrutinized under different headings and using different approaches e.g. Artificial Intelligence techniques, evolutionary algorithms and linear programming models under deterministic conditions. However, the information about real life processes is not always crisp but is often available as vague, uncertain and imprecise data. Fuzzy numbers finds application in handling vague terms, and therefore they can be suitably used to model real life scenarios involving vague parameters so as to obtain optimal solutions. Fuzzy multi-objective linear programming usually deals with flexible aspiration levels that are indicative of optimality when considering all objectives or goals simultaneously with possible deviation in objectives or constraints. Therefore in this study we develop a fuzzy multi-objective linear programming model with nonlinear membership functions for solving a multi objective TSP in order to simultaneously minimize the three parameters cost, distance and time. The importance of these parameters is assigned as weights to these objectives in the final model using AHP. The proposed model will give a compromised solution for best optimality and higher satisfaction level for the three parameters being considered in uncertain environment. The primary contribution of this study is a fuzzy mathematical model using nonlinear membership functions, more precisely the exponential functions to ensure an optimal solution in vague, imprecise and uncertain environment.

Keywords: Travelling Salesman Problem (TSP), nonlinear fuzzy numbers, exponential fuzzy numbers, fuzzy multi objective linear programming, vague parameters, AHP.

1 Introduction

Travelling Salesman Problem is a NP hard combinatorial problem to determine the shortest length or the least cost or minimum time to pass through a given set of cities

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1011–1019.
springerlink.com

where each city is visited exactly once in such a way that starting and ending city is the same. As TSP is considered a Multi-Objective Optimization Problem, the three objective functions for cost, distance and time must be characterized in distinct dimension. Optimal solution for a Multi-Objective TSP means to determine *k*- dimensional points in the space of viable solutions of problem such that it possesses minimum possible values in all *k* dimensions. However it may not possible to attain all the objectives entirely. In such cases permissible deviation from particular dimensions can lead to more flexible objectives which becomes achievable in real like scenarios. A conventional programming technique may not deal with such situation. A number of novel techniques and methods are being employed by researchers worldwide to deal with such situations. Fischer and Richter (1982) used Branch and Bound approach to solve TSP with two sum criteria. Sigal (1994) proposed decomposition approach to solve a TSP problem with two criteria route length and bottlenecks. He obtained both objectives for his model from the same matrix of cost as parameter. A multiple labelling scheme coupled with a branch and bound method was used by Tung (1994) to obtain possible Pareto optimal routes. A bi-objective TSP was analysed by Melamed and Sigal (1997) with the help of an *e*-constraint based algorithm. Ehrgott (2000) solved TSP with an approximation algorithm with bound on worst case performance. Borges and Hansen (2000) used weighted sums for study of global convexity of Multi-Objective TSP model. Hansen (2000) applied Tabu Search algorithm for a Multi-Objective TSP. Yan et al. (2003) proposed Evolutionary Algorithm for Multi-Objective TSP. A dynamic search algorithm and hence Dynamic Programming and Rounding technique was proposed by Angel et al. (2004) which uses Local Search method with exponential size neighbourhoods that is possible to be searched in polynomial time. Paquete et al. (2004) used Pareto Local Search technique which extended the Local Search algorithm for a single objective TSP to a Bi- objective case. Rehmat et al. (2007) solved the multi objective TSP problem using fuzzy logic approach with linear membership functions. Chaudhury and De (2011) used the Branch and Bound algorithm to approach TSP with two sum criteria. In this study we develop a fuzzy multi-objective linear programming model with nonlinear membership functions for solving a multi objective TSP in order to simultaneously minimize the three imprecise parameters cost, distance and time. The importance of these parameters is assigned as weights to these objectives in the final model using AHP. The proposed model will give a compromised solution for best optimality and higher satisfaction level for the three parameters being considered in uncertain environment.. The decision maker can introduce tolerances to accommodate uncertainty and vagueness. By adjusting these tolerances, a range of solutions with different satisfaction or aspiration level for all objectives are found. Decision maker can choose one of these solutions that best meets his requirements within the given domain. A simulation example is presented followed by conclusions.

2 Fuzzy Multi-Objective Linear Programming

The concept of decision making in Fuzzy environment involving several objectives was proposed by Jadeh (1965). It was applied to vector maximum problem by transforming Fuzzy Multi-Objective Linear Programming (FMOLP) Problem to single

objective linear program by Zimmerman (1978). For the following Multi-Objective Linear Programming model being considered,

Max Z= CX
Subject to $AX \leq b$

The adopted Fuzzy model can be given by,

Max $Z^0 \leq CX$
Subject to $AX \leq b$
In case of minimizing objective function, Linear Fuzzy Membership function is,

$$\mu_{1k}(C_k X) = \begin{cases} 0 & \text{if } C_k X \geq Z_k^0 + t_k \\ 1 - \dfrac{C_k X - Z_k^0}{t_k} & \text{if } Z_k^0 \leq C_k X \leq Z_k^0 + t_k, k = 1...n \\ 1 & \text{if } C_k X \leq Z_k^0 \end{cases}$$

According to Fuzzy Sets, membership function of the intersection of any two or more sets is the minimum Membership function of these sets. By virtue of this the objective function becomes:

Min($\mu_{11}(C_1 X)$,....., $\mu_{1k}(C_k X), \mu_{21}(a_1 X),\ldots\ldots, \mu_{2m}(a_m X)$)

A similar set of equations can be written for maximizing objective functions.
An exponential membership function for the k^{th} objective function can be formulated as below.

$$\mu_e(Z_k) = \begin{cases} 1 & \text{if } Z_k \leq Z_k^0 \\ \dfrac{e^{S(Z_k^0 - Z_k)/t_k} - e^{-S}}{1 - e^{-S}} & \text{if } Z_k^0 < Z_k < Z_k^0 + t_k \\ 0 & \text{if } Z_k \geq Z_k^0 + t_k \end{cases}$$

Where, S is a non-zero parameter prescribed by the decision maker. It denotes the risk perceived by the decision maker in an environment. As this parameter decreases, the overall aspiration or the satisfaction level of the objective functions increases. t_k is the tolerance or admissible flexibility in the parameters.

3 Fuzzy Multi-Objective Linear Programming for TSP

The most commonly considered objective for TSP is to determine an order for traveling across all cities such that total cost, total time and overall distance is minimized. Therefore the individual objective functions for each objective can be formed. Let X_{ij} denote the binary variable which equals one when city j is visited from city i otherwise it is zero i.e.

$$X_{ij} = \begin{cases} 1, City(i) \rightarrow City(j). \\ 0, Otherwise \end{cases}$$

Let C_{ij} denote the cost of travelling from city i to city j. The overall cost of a route is sum of costs on each links comprising that route. We have to minimize total traveling cost. The goal is set for total estimated cost for entire route for TSP and it is denoted by Z^0_1. For situations when estimated cost doesn't meet, we set tolerancet_1 for estimated cost. Therefore the objective function for minimization of estimated cost is given as follows:

$$Z_1 : min \sum_{i=1}^{n} \sum_{j=1}^{n} C_{ij} X_{ij} \lesssim Z_1^0 \tag{1}$$

Let d_{ij} denote the distance from city i to city j and Z^0_2 be corresponding aspiration level for this objective function. Then objective function for minimization of distance with tolerance t_2 is:

$$Z_2 : min \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} X_{ij} \lesssim Z_2^0 \tag{2}$$

Let t_{ij} denote the time taken to travel from city i to city j and Z^0_3 be its corresponding aspiration level. Then the objectives function for minimization the total time with tolerancet_3. The objective function is written as follows:

$$Z_3 : min \sum_{i=1}^{n} \sum_{j=1}^{n} t_{ij} X_{ij} \lesssim Z_3^0 \tag{3}$$

The three objective functions as formulated above are not independent of each other; rather in most of the cases they do depend on each other. However, we are not going to discuss the dependencies of parameters in this paper. The solution methodology developed here will work in all scenarios in case feasible solution is available. The membership functions for these objectives are set using a nonlinear function to check their level of acceptance in real world like framework. The constraints are as follows. Equation (1) and (2) ensures that every city is visited from exactly one neighboring city and vice versa, that is,

$$\sum_{i=1}^{n} X_{ij} = 1, \forall j \tag{4}$$

$$\sum_{j=1}^{n} X_{ij} = 1, \forall i \tag{5}$$

Further, a route has to be selected at most once, i.e.

$$X_{ij} + X_{Ji} \leq 1, \forall i, j \tag{6}$$

And non-negativity constraints:

$$X_{ij} \geq 0 \tag{7}$$

4 Simulation Example

The proposed Fuzzy Multi Objective Linear Program approach to solve a symmetric TSP is carried out in this section where a salesman starts from city 0, visits three cities one after the other exactly once and returns back to starting city 0 making a close mesh and adopting the route with optimal cost incurred, time taken and distance travelled. The input data is taken from Chaudhuri et al. (2011).Figure 1 represents map of cites to be visited and these cities are listed along with their parameters cost, time and distance in matrix form in Table 1, where the triplet (c, d, t) represents cost, distance and time respectively for given pair of cites. Let links x_{ij} be the binary decision variable for selection of link (i, j) from city i to city j. The objective functions Z_1,Z_2 and Z_3 are formulated in equation below for cost, distance and time, respectively. The aspiration levels for these objectives turn out to be 65, 16 and 11 when each objective is solved separately, subject to given set of constraints (4) to (7).The objective functions can be written as follows:

$$\text{Min}Z_1=20X_{01}+15X_{02}+11X_{03}+20X_{10}+30X_{12}+10X_{13}+15X_{20}+30X_{21}+20X_{23}+11X_{30}+10X_{31}+20X_{32}$$

$$\text{Min}Z_2=5X_{01}+5X_{02}+3X_{03}+5X_{10}+5X_{12}+3X_{13}+5X_{20}+5X_{21}+10X_{23}+3X_{30}+3X_{31}+10X_{32}$$

$$\text{Min}Z_3=4X_{01}+5X_{02}+2X_{03}+4X_{10}+3X_{12}+3X_{13}+5X_{20}+3X_{21}+2X_{23}+2X_{30}+3X_{31}+2X_{32}$$

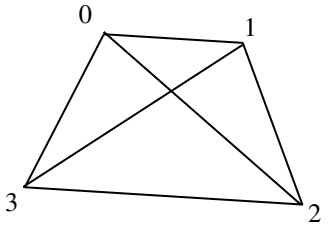


Fig. 1. Symmetric Traveling Salesman route

Table 1. The matrix for time, cost and distance for each pair of cities.(Chaudhuri et al. 2011)

City	0	1	2	3
	(c,d,t)	**(c,d,t)**	**(c,d,t)**	**(c,d,t)**
0	(0,0,0)	(20,5,4)	(15,5,5)	(11,3,2)
1	(20,5,4)	(0,0,0)	(30,5,3)	(10,3,3)
2	(15,5,5)	(30,5,3)	(0,0,0)	(20,10,2)
3	(11,3,2)	(10,3,3)	(20,10,2)	(0,0,0)

The Fuzzy Membership functions for cost, distance and time objective functions are defined below respectively

$$\mu_e(Z_1)=\begin{cases}1 & \text{If } Z_1 \le 65 \\ \dfrac{e^{S(65-Z_1)/t_1}-e^{-S}}{1-e^{-S}} & \text{If } 65 < Z_1 < 65+t_1 \\ 0 & \text{If } Z_1 \ge 65+t_1\end{cases} \tag{8}$$

$$\mu_e(Z_2)=\begin{cases}1 & \text{If } Z_2 \le 16\\ \dfrac{e^{S(16-Z_2)/t_2}-e^{-S}}{1-e^{-S}} & \text{If } 16<Z_2<16+t_2\\ 0 & \text{If } Z_2 \ge 16+t_2\end{cases} \tag{9}$$

$$\mu_e(Z_3)=\begin{cases}1 & \text{If } Z_3 \le 11\\ \dfrac{e^{S(11-Z_3)/t_3}-e^{-S}}{1-e^{-S}} & \text{If } 11<Z_3<11+t_3\\ 0 & \text{If } Z_3 \ge 11+t_3\end{cases} \tag{10}$$

Now we formulate a crisp single objective linear programming with above fuzzy numbers and least satisfaction level of each of the objectives α

Maximize $\alpha = \lambda_1 U_1 + \lambda_2 U_2 + \lambda_3 U_3$

Subjected to:

$$U_1 \le \frac{e^{S(65-Z_1)/t_1}-e^{-S}}{1-e^{-S}}$$

$$U_2 \le \frac{e^{S(16-Z_2)/t_2}-e^{-S}}{1-e^{-S}}$$

$$U_3 \le \frac{e^{S(16-Z_3)/t_3}-e^{-S}}{1-e^{-S}}$$

$\lambda_1+\lambda_2+\lambda_3=1$ and Constraints (4) – (7)

$X_{ij} \epsilon \{0,1\}, U_1, U_2, U_3 \ge 0$

λ_1, λ_2 and λ_3 are parameters to assign weights to various objectives. It should be application oriented and can be different for different industry. However without loss of generality we assume each objective to be equally important and assign equal weight i.e. 1/3 to each of these parameters for this model. Lingo 11.0 computer software is used to run this ordinary LP model on an Intel® 1.60GHz Processor with 1 GB RAM. As shown in Table.2 solution is infeasible when tolerances are 5, 2 and 1 for Z_1, Z_2 and Z_3 respectively. By relaxing tolerance in Z3 to 3, solution becomes feasible. In this case, the optimal path is achieved with α = 0.665. By increasing tolerance in Z3 from 4 to 5, an optimal solution with α = 0.748 is obtained. Further increase in aspiration level is obtained by increasing tolerance on Z_1 It is evident from Table 2 that on decreasing value of parameter S, the aspiration level increase. Further increasing the tolerance i.e. the flexibility on TSP parameters, aspiration levels can further be increased. These results show that by adjusting parameter S and tolerances, an optimal solution to Multi-Criteria TSP can be determined.

Table 2. Solution of Fuzzy multi-objective linear programming problem at S=1

Sol	Z_1,t_1	Z_2,t_2	Z_3,t_3	α	route
1	65,5	16,2	11,1	-	No feasible solution
2	65,5	16,2	11,3	0.6650257	$(x_{02},x_{21},x_{13},x_{30})$
3	65,5	16,2	11,4	0.7153278	$(x_{02},x_{21},x_{13},x_{30})$
4	65,5	16,2	11,5	0.7487948	$(x_{02},x_{21},x_{13},x_{30})$
5	65,6	16,2	11,5	0.7625209	$(x_{02},x_{21},x_{13},x_{30})$
6	65,7	16,2	11,5	0.7725456	$(x_{02},x_{21},x_{13},x_{30})$
7	65,16	16,2	11,5	0.8077841	$(x_{02},x_{21},x_{13},x_{30})$

Table 3. Solution of Fuzzy multi-objective linear programming problem at S=0.1

Sol	Z_1,t_1	Z_2,t_2	Z_3,t_3	α	route
1	65,5	16,2	11,1	-	No feasible solution
2	65,5	16,2	11,3	0.7090877	$(x_{03},x_{31},x_{12},x_{20})$
3	65,5	16,2	11,4	0.7637187	$(x_{03},x_{31},x_{12},x_{20})$
4	65,5	16,2	11,5	0.7968391	$(x_{03},x_{31},x_{12},x_{20})$
5	65,6	16,2	11,5	0.8081107	$(x_{03},x_{31},x_{12},x_{20})$
6	65,7	16,2	11,5	0.8161799	$(x_{02},x_{21},x_{13},x_{30})$
7	65,16	16,2	11,5	0.8435238	$(x_{02},x_{21},x_{13},x_{30})$

The decision parameters λ_1, λ_2 and λ_3 can also be determined using Analytic Hierarchical Process. A quick example is as given below using data in table 4. Entries in the cell are indicative of importance of one parameter over the other. Eigenvalues are calculated in the last column.

Table 4. Use of AHP to determine weight of parameters λ_1, λ_2 and λ_3

	Cost	Time	Distance	Eigenvector
Cost	1	3	5	0.62670
Time	1/3	1	4	0.27969
Distance	1/5	1/4	1	0.09362
Totals				1.00000

The eigenvector (0.62670, 0.27969, and 0.09362) indicates relative importance of one parameter over other. Thus, the cost is the given maximum importance and the distance is the least important. The next step is to check the Consistency of the result by calculating λ_{max}. Multiply on the right the matrix of judgments by the eigenvector, obtaining a new vector e.g. the first row in the matrix gives 1 x 0.62670 + 3 x 0.27969 + 5 x 0.09362 = 1.93384 and the remaining two rows give 0.86305 and 0.28888. Call this vector of three elements (1.93384, 0.86305, 0.28888) as the product $A\omega$ and the AHP theory says that $A\omega=\lambda_{max}\omega$ so we can now get three estimates of λ_{max} by the

simple expedient of dividing each component of (1.93384, 0.86305, 0.28888) by the corresponding eigenvector element. This gives 1.93384/0.62670= 3.085759 along with 3.085752 and 3.085782. Mean of these values is 3.085764 and that is the estimate for λ_{max}. The consistency Index for a matrix is calculated from (λ_{max} -n)/(n-1) at n=3 as 0.042882. The final step is to calculate the consistency Ratio for this set of judgement using the *CI* or the corresponding value from large samples of matrices of purely random judgments using the Table 5 below, derive from Saaty [8].

Table 5. Index of consistency for random judgments (Saaty 1980)

1	2	3	4	5	6	7	8
0	0	0.58	0.90	1.12	1.24	1.32	1.41

For this example, that gives 0.042882/0.58 = 0.073934. As CR<0.1 (0.073934 < 0.1), we are at safe ground. For $\lambda_1 = 0.63, \lambda_2 = 0.28$ and $\lambda_3 = 0.09$ (as obtained above) and S=1, the result is as tabulated below in table 6.

Table 6. Solution of Fuzzy multi-objective linear programming problem with weights obtained from from AHP and S=1

Sol	Z_1,t_1	Z_2,t_2	Z_3,t_3	α	route
1	65,5	16,2	11,1	-	No feasible solution
2	65,5	16,2	11,3	0.7647053	$(x_{02},x_{21},x_{13},x_{30})$
3	65,5	16,2	11,4	0.7784240	$(x_{02},x_{21},x_{13},x_{30})$
4	65,5	16,2	11,5	0.7875514	$(x_{02},x_{21},x_{13},x_{30})$
5	65,6	16,2	11,5	0.8137557	$(x_{02},x_{21},x_{13},x_{30})$
6	65,7	16,2	11,5	0.8328938	$(x_{02},x_{21},x_{13},x_{30})$
7	65,16	16,2	11,5	0.9001673	$(x_{02},x_{21},x_{13},x_{30})$

5 Conclusion

In the present paper, a symmetric TSP is investigated as Fuzzy multi objective problem with vague and imprecise decision parameters in cost, time and distance using nonlinear membership function. The tolerances or flexibility are introduced by decision maker to accommodate this ambiguity. By adjusting these tolerances or flexibility, a range of solutions with diverse satisfaction level are attained from which decision maker chooses one that best meets his requirements within given flexibility in parameters. In supply chain and logistics TSP plays very important role. In vague environment of supply chains; flexibility in various parameters of TSP and hence the best available satisfaction level as solution, can provide some strategic advantage to supply chains.

6 Scope for Future Work

There is a tremendous potential for further work on development of methods to solve TSP problems with vague description of resources using other techniques like Rough Sets. For efficient results, some heuristics may be exploited such as swarm optimization, ant colony optimization etc. There are many other nonlinear membership functions that are capable of representing real life scenario to some extent. These functions can be identified and better results can be produced. Further a problem with relative dependencies among objective function can be analyzed for further research where one objective may enjoy some priority over other as decided by management of corporate firms.

References

1. Angel, E., Bampis, E., Gourvès, L.: Approximating the Pareto curvewith Local Search for BiCriteria TSP (1,2) Problem. Theor. Comp. Sci. 310(1-3), 135–146 (2004)
2. Chaudhuri, A., De, K.: Fuzzy multi objective linear programming for travelling salesman problem. Afr. J. Math. Comp. Sci. Res. 4(2), 64–70 (2011)
3. Ehrgott, M.: Approximation Algorithms for Combinatorial Multi-Criteria Problems. International Transactions in Operations Research 7, 5–31 (2000)
4. Fischer, R., Richter, K.: Solving Multi-Objective Traveling Salesman Problem by Dynamic Programming. Mathematicsche Operations Forschung und Statistic Series Optimization 13(2), 247–252 (1982)
5. Hansen, M.P.: Use of Substitute Scalarizing Functions to Guide a Local Search Based Heuristics,The Case of MOTSP. Journal of Heuristics 6, 419–431 (2000)
6. Melamed, I.I., Sigal, I.K.: The Linear Convolution of Criteria in the Bi-Criteria Traveling Salesman Problem. Computational Mathematics and Mathematical Physics 37(8), 902–905 (1997)
7. Paquete, L., Chiarandini, M., Stützle, T.: Pareto Local Optimum Sets in Bi-Objective Traveling Salesman Problem: An Experimental Study. In: Metaheuristics for Multiobjective Optimization. Lect. Notes Econ. Math. Syst., vol. 535, pp. 177–199. Springer, Berlin (2004)
8. Rehmat, A., Saeed, H., Cheema, M.S.: Fuzzy Multi-objective Linear Programming Approach for Travelling Salesman Problem. Pak. J. Stat. Oper. Res. 3(2), 87–98 (2007)
9. Saaty, T.: The analytical hierachy process. McGraw Hill, USA (1980)
10. Sigal, I.K.: Algorithm for Solving the Two-Criterion Large-scale Traveling Salesman Problem. Computational Mathematics and Mathematical Physics 34(1), 33–43 (1994)
11. Tung, C.T.: A Multi criteria Pareto-optimal Algorithm for the Traveling Salesman Problem. Asia-Pacific Journal of Operational Research 11, 103–115 (1994)
12. Yan, Z., Zhang, L., Kang, L., Lin, G.: A New MOEA for Multi-objective TSP and Its Convergence Property Analysis. In: Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele, L. (eds.) EMO 2003. LNCS, vol. 2632, pp. 342–354. Springer, Heidelberg (2003)
13. Zadeh, L.: Fuzzy Logic and its Applications. Academic Press, New York (1965)
14. Zimmerman, H.J.: Fuzzy programming and linear programming with several objective functions. Fuzzy Sets and Systems 1, 45–55 (1978)

A Novel Approach for Image Retrieval Based on ROI and Multifeatures Using Genetic Algorithm

K.S. Md. Musa Mohinuddin[1], P. Subbaiah[2], and S. Tipu Rahaman[1]

[1] Vaagdevi Institute of Technology and Science,
JNTUA University,
Proddatur, India
{musamohinuddin,shaiktipurahaman}@gmail.com
[2] Acharya College of Engineering,
JNTUA University,
Badvel, India
Subbaiah_nani@sify.com

Abstract. The need for efficient Image Retrieval has increased tremendously in many application areas and in addition to it the present day images are extremely varied with lot of information. Hence the problem of Image Retrieval has grown further complex. Implementing CBIR based on single feature like color, texture or shape does not produce satisfactory results. In our proposed approach, the retrieval is carried out on the user selected region i.e., ROI (Region – Of – Interest) followed by evaluating the low level features. These multi-features are fused based on the similarity score and the fitness function is evaluated by the Genetic Algorithm (GA). In GA the weights of similarity score are optimally assigned. The Corel databases of 1000 images are considered in which image retrieval is done for actual and ROI image. The performance is evaluated by the parameters recall rate and precision rate. From the obtained results, it is evident that our proposed approach outperforms traditional methods.

1 Introduction

The problem of searching for digital images in large databases is an open area that has attracted many researchers. Text query can be applied to multimedia information retrieval, but it has inherent deficiencies. An alternative approach that overcomes text – based is CBIR which is designed to work with actual features of the image. Though CBIR gives satisfactory results over text based, it is mainly based on low – level features of image. The commonly used features are color, texture and shape. In general, algorithms for CBIR are based on single feature which doesn't yield good query accuracy. But algorithms based on high level features will reduce the query efficiency. Representing an image with multi low-level features is expected to achieve better results. This can be done by fusion of multi-features. To increase the accuracy further, optimization techniques are used in fusion of multi feature.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1021–1026.
springerlink.com

2 Related Works

Information fusion can be carried out in two levels (i) Feature level (ii) Decision making level. Information fusion at feature level is done based on different features of an image like color, texture and shape. B.G. Prasad et al. [2] proposed Region based image retrieval using integrated color, shape and location index. In the method proposed by Young Deok Chun et al. [5], image retrieval is done based on combination of multi resolution color and texture features. For information fusion to be carried out at decision making level, different features are obtained and then based on similarity score values the results are being fused according to certain rules. This approach yields better results than the previous method. The method proposed by Anil K. Jain et al. [1] is based on decision making level, which integrated the results of shape and color by combining the associated similarity values with appropriate weights. Xiuqi Li et al. [4] proposed image retrieval based on color, texture and spatial information. Mladen Jovic et al. [3] proposed an image similarity method based on fusion of similarity scores of feature obtained from similarity ranking lists, this method utilizes the advantage of using different integration algorithms for combining the similarity volume score.

3 Proposed Method

In this paper, image retrieval based on ROI and multi feature similarity score fusion using GA is being analyzed. The features colour and texture are considered for evaluating similarity score. For applying GA on these multifeatures, the weights of similarity score are optimally assigned. By our proposed approach the results obtained are much superior to the existing methods. In human eye perception of an image, the features that are mostly considered are color, texture and shape. For high accuracy, these perception features can be increased but the problem is that, which feature has to be given high priority. A relevant solution to this problem is to depend on optimization techniques.

3.1 ROI Selection

An image consists of varied objects, regions etc. The retrieval for on particular object or region does not produce satisfactory results because the low level features are extracted for the entire image. A better approach to increase the accuracy is by concentrating on desired region or object of the query. This method is termed as Region of Interest (ROI).

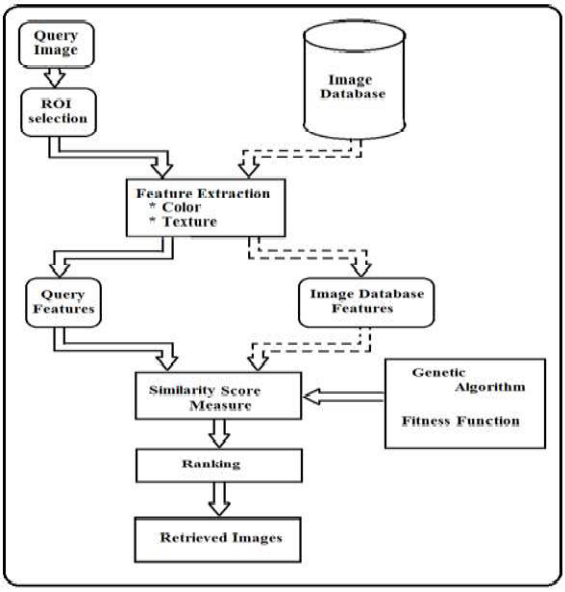


Fig. 1. Proposed Algorithm

3.2 Color Feature Extraction

To extract color feature, HSV color space is considered because it confines to the human eye interpretation and also it has the ability to separate chromatic and achromatic components. In this colour space hue represents the dominant color value of the image, saturation give the percentage of white light added to a pure color and value refers to the perceived light intensity. In our approach to evaluate the color feature the RGB or any color space image is converted into HSV. Then for the query and every image in the database HSV values are computed and their absolute difference is obtained according to equations 1. To reduce the computation complexity the absolute difference values are grouped into single parameter G1.

$$G1 = [abs(qH-dH) + abs(qS-dS) + abs(qV-dV)] \quad (1)$$

Here abs is the absolute difference, qH, dH, qS, dS qV, dV are the Hue, Saturation,Value of the query and database images respectively.

3.3 Texture Feature Extraction

Image Texture gives perceived information about the object by the human eye. Texture is about the spatial arrangement of color or intensities in an image or selected region of an image. In our paper, the query image is converted into gray scale on which the gray level co – occurrence matrix (GLCM) is gained. Then the statistical features like Entropy, Energy, Contrast, Homogeneity and Correlation are computed on the GLCM with the following equations as described. The texture feature evaluation

accuracy would increase when more features are considered. But in some images certain features will be more dominant so they should be given much priority. This not only increases the computational burden but also choosing such priority features is difficult. Hence for simplicity all features are given same importance.

$$\text{Entropy} = -\Sigma(\,p(\,i\,,j\,).*\log2(p(\,i,j\,)\,)\,) \tag{2}$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \tag{3}$$

$$\text{Contrast} = \sum_{i,j} |i-j|^2 p(i,j) \tag{4}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1-|i-j|} \tag{5}$$

$$\text{Correlation} = \sum_{i,j} \frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma_i \sigma_j} \tag{6}$$

Here p(i, j) is the gray-level value at the coordinate (i, j) .The above five statistical feature values are obtained and the absolute difference is evaluated between the query and every image in the database. To reduce the computation complexity the absolute difference values are grouped into single parameter G2.

$$G2 = [\text{abs}(qE - dE) + \text{abs}(qEn - dEn) + \text{abs}(qC - dC) + \text{abs}(qHo - dHo) + \text{abs}(qCo - dCo)\,] \tag{7}$$

Here abs is the absolute difference, qE, dE, qEn, dEn qC, dC, qHo, dHo, qCo, dCo are the Entropy, Energy, Contrast, Homogenity, Correlation of the query and database images respectively.

3.4 Similarity Score Fusion Using Genetic Algorithm

Multifeature similarity score fusion cannot be performed directly because the statistical features are based on different parameters and also assigning weights to the multifeature is a key problem. By eliminating the differences between multifeature, similarity score fusion can be done which requires normalization of the feature values. To assign the weights optimally Genetic Algorithm is being used. The results of multi feature similarity score are computed by,

$$S_{Fi} = \frac{S_{NCi} \cdot W_c + S_{NTi} \cdot W_T}{W_c + W_T}, \tag{8}$$

Where S_{Fi} is the fused similarity score, S_{NCi} is the normalized color feature similarity score, S_{NTi} is the normalized texture feature similarity score, W_C is the weight of color feature similarity score, W_T is the weight of texture feature similarity score. By assigning appropriate values to W_C and W_T , a fine similarity score fusion can be gained. The value of W_C is an integer between 0 an I where I is a positive integer. The value of W_T is I – W_C. According to the weights of W_C and W_T of N individuals we

obtain n groups of image retrieval results. For every group top M images are considered by which the total number of images is MN. After calculating the occurrence frequency of images of every group, the fitness of every individual is evaluated. The occurrence frequency of all images of i^{th} group Gi in all MN images is

$$N_i = \sum_{k=1}^{M} N_{ik} \tag{9}$$

The fitness function is evaluated by

$$P_i = \frac{N_i}{\sum_{l=0}^{N} N_l} \tag{10}$$

Where Pi gives the images in the i^{th} group which possess the images in high proportion of all MN images. In this paper, the number of iterations are taken as three and the obtained results are the optimal solutions.

4 Experments and Analysis

The proposed algorithm is evaluated in two steps. (i) By considering the normal Corel image database of 1000 images which contain 10 domains each with 100 images. (ii) By selecting ROI on the query image for the same database. The parameters precision rate and recall rate are taken into consideration. For the query image, ROI is selected and the output is evaluated as shown in Fig.2.

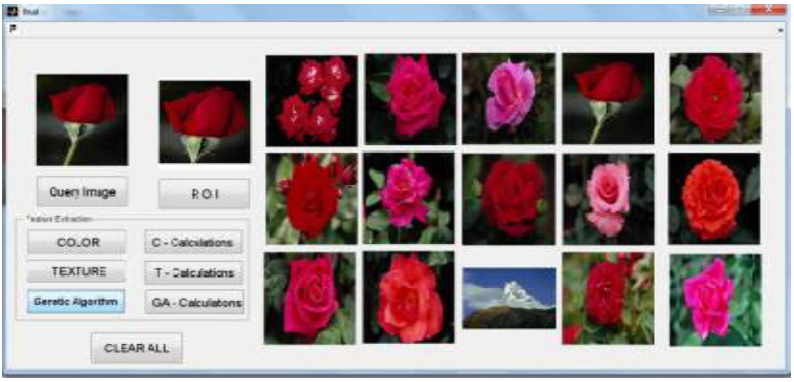


Fig. 2. Retrieval result based on ROI+color + texture feature

For the above images, the precision rate and recall rate based on color, texture, Genetic Algorithm and ROI are tabulated below. Precision rate is the fraction of retrieved images that are relevant to the query, while Recall rate is the fraction of relevant images that are retrieved.

Table 1. Retrieval results with and without ROI

ROI based Feature Extraction techniques	Without ROI		With ROI	
	Precision Rate (%)	Recall Rate(%)	Precision Rate (%)	Recall Rate (%)
Color	39.2	28.4	40.5	29.6
Color+Texture	42.4	37.3	43.4	36.5
Genetic Algorithm	46.7	39.7	48.6	40.9

5 Conclusion

In this paper a novel approach for image retrieval using ROI and Multifeature similarity score fusion is proposed. For the query image multi features are evaluated and by using Genetic Algorithm better retrieval results are obtained. These results are compared with ROI selected query. The obtained results show that our proposed method gives higher accuracy. This approach can be further extended for other low level features like shape, rotation, translation, etc... which would be our future work.

References

[1] Jain, A.K., Vailaya, A.: Image Retrieval using color and shape. Pattern Recognition 29, 1233–1244 (1996)

[2] Prasad, B.G., Biswas, K.K.: Region-Based Image Retrieval using Integrated color, shape and location index. Computer Vision and Image Understanding 94, 193–233 (2004)

[3] Jović, M., Hatakeyama, Y., Dong, F., Hirota, K.: Image Retrieval Based on Similarity Score Fusion from Feature Similarity Ranking Lists. In: Wang, L., Jiao, L., Shi, G., Li, X., Liu, J. (eds.) FSKD 2006. LNCS (LNAI), vol. 4223, pp. 461–470. Springer, Heidelberg (2006)

[4] Li, X., Chen, S.-C., Shyu, M.-L., Furht, B.: Image Retrieval By Color, Texture, And Spatial Information. In: Proceedings of the 8th International Conference on Distributed Multimedia Systems (DMS 2002), San Fransisco, Bay, CA, USA, pp. 152–159 (2002)

[5] Chun, Y.D., Kin, N.C., Jang, I.H.: Content based image retrieval using multiresolution color and texture features. IEEE Transaction on Multimedia 10, 1073–1084 (2008)

Ship Detection from SAR and SO Images

Y. Sreedevi and B. Eswar Reddy

JNTUACE, Anantapur

Abstract. A Ship detection method was proposed in this paper by combining top-down recognition with bottom-up image segmentation, which will work on Synthetic Aperture Radar (SAR) images and Space-borne Optical (SO) images. There are two steps in this method: a hypothesis generation step and a verification step. In the top-down hypothesis generation step, we design an improved Shape Context feature, which is more robust to ship deformation and background clutter. The improved Shape Context is used to generate a set of hypotheses of ship locations and figure ground masks, which have high recall and low precision rate. In the verification step, we first compute a set of feasible segmentations that are consistent with top-down ship hypotheses, and then we propose a False Positive Pruning (FPP) procedure to prune out false positives. We exploit the fact that false positive regions typically do not align with any feasible image segmentation. Experiments show that this simple framework is capable of achieving both high recall and high precision with only a few positive training examples and that this method can be generalized to many ship classes.

Keywords: Ship Detection, SAR images, So Images, Segmentation, Thresholding.

1 Introduction

Ship detection is an important, yet challenging vision task. It is a critical part in many applications such as image search, image auto-annotation and scene understanding; however it is still an open problem due to the complexity of ship classes and images. Current approaches ([1][2] [3][4][5] [6][7] [8] [9][10]) to ship detection can be categorized by top-down, bottom-up or combination of the two. Top-down approaches ([11][2][12]) often include a training stage to obtain class-specific model features or to define ship configurations. Hypotheses are found by matching models to the image features. Bottom-up approaches start from low-level or mid-level image features, i.e. edges or segments ([8][5][9] [10]). These methods build up hypotheses from such features, extend them by construction rules and then evaluate by certain cost functions.

The third category of approaches combining top-down and bottom-up methods have become prevalent because they take advantage of both aspects. Although top-down approaches can quickly drive attention to promising hypotheses, they are prone to produce many false positives when features are locally extracted and matched. Features within the same hypothesis may not be consistent with respect to low-level image segmentation. On the other hand, bottom-up approaches try to keep consistency in low level image segmentation, but usually need much more efforts in searching and grouping.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1027–1035.
springerlink.com

Wisely combining these two can avoid exhaustive searching and grouping while maintaining consistency in ship hypotheses. For example, Bornstein et al. enforce continuity along segmentation boundaries to align matched patches ([2]). Levin et al. take into account both bottom-up and top-down cues simultaneously in the framework of CRF ([3]). Our detection method falls into this last category of combining top-down recognition and bottom-up segmentation, with two major improvements over existing approaches. First, we design a new improved Shape Context (SC) for the top-down recognition. Our improved SC is more robust to small deformation of ship shapes and background clutter. Second, by utilizing bottom-up segmentation, we introduce a novel False Positive Pruning (FPP) method to improve detection precision. Our framework can be generalized to many other ship classes because we pose no specific constraints on any ship class.

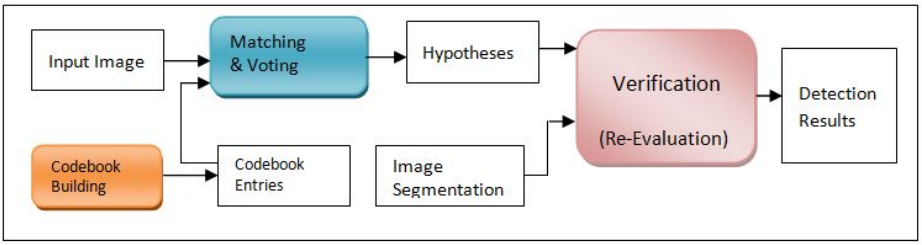


Fig. 1. Method overview. Our method has three parts (shaded rectangles). Codebook building (cyan) is the training stage, which generates codebook entries containing improved SC features and ship masks. Top-down recognition (blue) generates multiple hypotheses via improved SC matching and voting in the input image. The verification part (pink) aims to verify these top-down hypotheses using bottom-up segmentation. Round-corner rectangles are processes and ordinary rectangles are input/output data.

The overall structure of the paper is organized as follows. Sec. 2 provides an overview to our framework. Sec.3 describes the improved SCs and the top-down hypothesis generation. Sec.4 describes our FPP method combining image segmentation to verify hypotheses. Experiment results are shown in Sec.5, followed by discussion and conclusion in Sec.6.

2 Method Overview

Our method contains three major parts: codebook building, top-down recognition using matching and voting, and hypothesis verification, as depicted in Fig.1. The ship models are learned by building a codebook of local features. We extract improved SC as local image features and record the geometrical information together with ship figure-ground masks. The improved SC is designed to be robust to shape variances and background clutters. For rigid ships and ships with slight articulation, our experiments show that only a few training examples suffice to encode local shape information of ships.

We generate recognition hypotheses by matching local image SC features to the codebook and use SC features to vote for ship centers. A similar top-down voting scheme is described in the work of [4], which uses SIFT point features for ship detection. The voting result might include many false positives due to small context

of local SC features. Therefore, we combine top-down recognition with bottom-up segmentation in the verification stage to improve the detection precision. We propose a new False Positive Pruning (FPP) approach to prune out many false hypotheses generated from top-down recognition. The intuition of this approach is that many false positives are generated due to local mismatches. These local features usually do not have segmentation consistency, meaning that pixels in the same segment should belong to the same ship. True positives are often composed of several connected segments while false positives tend to break large segments into pieces.

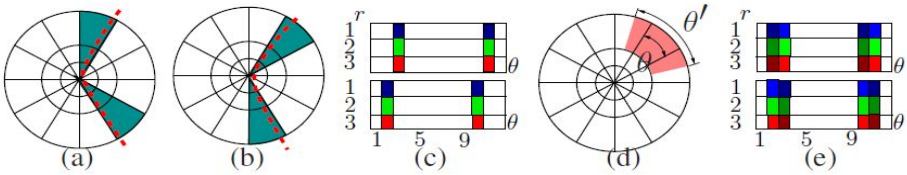


Fig. 2. Angular Blur. (a) and (b) depicts different bin responses of two similar contours. (c) depicts their histograms. (d) Enlarges angular span θ to θ', letting bins be overlapped in angular direction. (e) depicts the responses on the overlapped bins, where the histograms are more similar.

2.1 Top-Down Recognition

In the training stage of top-down recognition, we build up a codebook of improved SC features from training images. For a test image, improved SC features are extracted and matched to codebook entries. A voting scheme then generates ship hypotheses from the matching results.

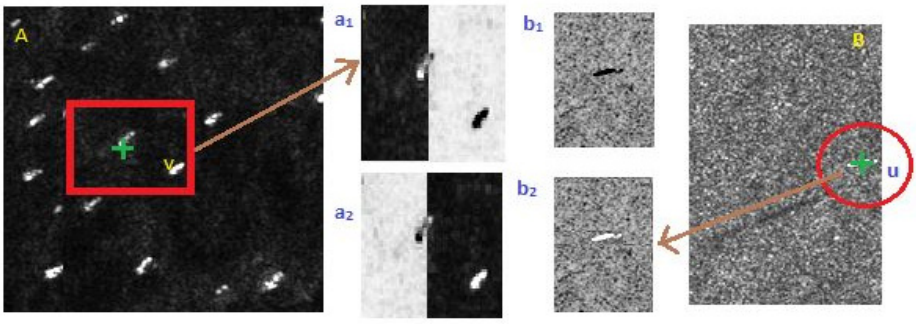
Codebook Building. For each ship class, we select a few images as training examples. Ship masks are manually segmented and only edge map inside the mask is counted in shape context histogram to prune out edges due to background clutter. The Codebook Entries (CE) is a repository of example features: CE ={ ce_i }. Each codebook entry $ce_i = (u_i, \delta_i, m_i, w_i)$ records the feature for a point 'i' in labeled ships of the training images. Here u_i is the shape context vector for point 'i'. δ_i is the position of point 'i' relative to the ship center. m_i is a binary mask of figure-ground segmentation for the patch centered at point 'i'. w_i is the weight mask computed on m_i, which will be introduced later.

Improved Shape Context. The idea of Shape Context (SC) was first proposed by Belongie et al. ([13]). The basic definition of SC is a local histogram of edge points in a radius-angle polar grid. Following works ([14] [15]) improve its distinctive power by considering different edge orientations. Besides SC, other local image features such as wavelets, SIFT and HOG have been used in key point based detection approaches ([4] [12]). Suppose there are n_r (radial) by n_Θ (angular) bins and the edge map E is divided into E_1, … … , E_0 by o orientations (similar to [15]), for a point at p, its SC is defined as $u = \{h_1, \ldots \ldots, h_0\}$,

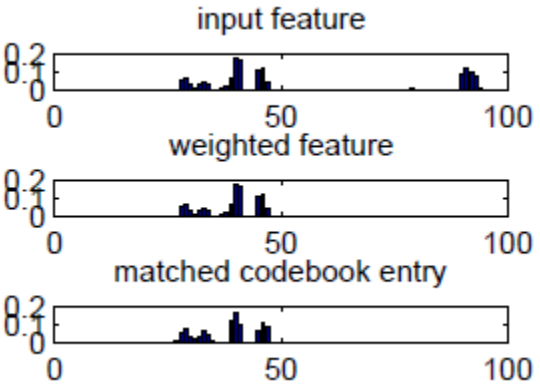
$$\text{where } h_i(k) = \#q \neq p : q \in E_i, \overline{pq} \in \text{bin}(k), k=1,2, \ldots, n_r n_\Theta \quad (1)$$

Angular Blur. A common problem for the shape context is that when dense bins are used or contours are close to the bin boundaries, similar contours have very different histograms (Fig.2-(c)). This leads to a large distance for two similar shapes if L2-norm or $\varkappa^2$ distance function is used. EMD ([16]) alleviates this by solving a transportation problem; but it is computationally much more expensive.

The way we overcome this problem is to overlap spans of adjacent angular bins: bin(k) ∩ bin(k+1) ≠ ø (Fig.2-(d)). This amounts to blurring the original histogram along the angular direction. We call such an extension Angular Blur. One edge point in the overlapped regions are counted in both of the adjacent bins. So the two contours close to the original bin boundary will have similar histograms for the overlapping bins (Fig.2- (e)).With angular blur, even simple L_2-norm can tolerate slight shape deformation. It improves the basic SC without the expensive computation of EMD.



(a)



(b)

Fig. 3. Distance function with mask. In (a), a feature point v has the edge map of a1 around it. Using ship mask b1, it succeeds to find a good match to u in B (ship model patch), whose edge map is b2. a2 is the ship mask b1 over a1. Only the edge points falling into the mask area are counted for SC. In (b), histograms of a1, a2 and b2 are shown. With the mask function, a2 is much closer to b2, thus got well matched.

Mask Function on Shape Context

In real images, ships SCs always contain background clutter. This is a common problem for matching local features. Unlike learning methods ([1] [12]) which use a large number of labeled examples to train a classifier, we propose to use a mask function to focus only on the parts inside ship while ignoring background in matching.

For ce = (u, δ, m, w) and a SC feature f in the test image, each bin of f is masked by figure-ground patch mask m of ce to remove the background clutter. Formally, we compute the weight w for bin k and distance function with mask as:

$$w(k) = \text{Area}(\text{bin}(k) \cap m) / \text{Area}(\text{bin}(k)),\ k = 1,2, \ldots, n_r n_\Theta \tag{2}$$

$$D_m(ce, f) = D(u, w.v) = \|u - w.v\|^2 \tag{3}$$

Where $(\cdot)$ is the element-wise product. D can be any distance function computing the dissimilarity between histograms (We simply use L_2-norm). Figure 3 gives an example for the advantage of using mask function.

Hypothesis Generation. The goal of hypothesis generation is to predict possible ship locations as well as to estimate the figure-ground segmentation for each hypothesis. Our hypothesis generation is based on a voting scheme similar to [4]. Each SC feature is compared with every codebook entry and makes a prediction of the possible ship center. The matching scores are accumulated over the whole image and the predictions with the maximum scores are the possible ship centers. Given a set of detected features $\{f_i\}$ at location $\{l_i\}$, we define the probability of matching codebook entry ce_k to f_i as $p(ce_k|l_i) \propto \exp(-D_m(ce_k, f_i))$. Given the match of ce_k to f_i, the probability of a ship o with center located at c is defined as $p(o,c|ce_k,l_i) \propto$ eexp$(-\|c+\delta_k-l_i\|^2)$. Now the probability of the hypothesis of ship o with center c is computed as:

$$P(o, c) = \sum_{ik} F_{ik} P(o, c \mid ce_k, l_i) P(ce_k, l_i) P(l_i) \tag{4}$$

P (o, c) gives a voting map V of different locations c for the ship class o. Extracting local maxima in V gives a set of hypotheses $\{H_j\} = \{(o_j, c_j)\}$.

Furthermore, figure-ground segmentation for each H_j can be estimated by backtracking the matching results. For those f_i giving the correct prediction, the patch mask m in the codebook is "pasted" to the corresponding image location as the figure ground segmentation. Formally, for a point p in image at location 'pl', we define $P(p = fig|ce_k,l_i)$ as the probability of point p belonging to the foreground when the feature at location l_i is matched to the codebook ce_k : $P(p= fig|ce_k,l_i) \propto \exp(-\|p_l-l_i\|) m_k (\overline{p}_l l_i)$. And we assume that $P(ce_k, l_i|H_j) \propto p(o_j,c_j|ce_k,l_i)$ and $P(f_i|ce_k) \propto p(ce_k|f_i)$ the figure-ground probability for hypothesis H_j is estimated as

$$P(p = fig|H_j) \propto \int_k \exp(\| p_l - l_i \|) m_k (\overline{p}_l l_i) P(f_i \mid ce_k) P(ce_k, l_i \mid H_j) \tag{5}$$

Eq. (5) gives the estimation of top-down segmentation. The whole process of top-down recognition is shown in Fig. 4. The binary top-down segmentation (F, B) of figure (F) and background (B) is the obtained by thresholding $P(p = fig|H_j)$.

2.2 Verification: Combining Recognition and Segmentation

From our experiments, the top-down recognition using voting scheme will produce many False Positives (FPs). In this section, we propose a two-step procedure of False Positive Pruning (FPP) to prune out FPs. In the first step we refine the top-down hypothesis mask by checking its consistency with bottom-up segmentation. Second the final score on the refined mask is recomputed by considering spatial constraints. Combining Bottom-up Segmentation The basic idea for local feature voting is to make global decision by the consensus of local predictions. However, these incorrect local predictions using a small context can accumulate and confuse the global decision. For example, in ship detection, two trunks will probably be locally taken as human legs and produce a human hypothesis; another case is the silhouettes from two standing-by ships.

In ship detection, the top-down figure-ground segmentation masks of the FPs usually look similar to a ship. However we notice that such top-down mask is not consistent with the bottom-up segmentation for most FPs. The bottom-up segments share bigger contextual information than the local features in the top-down recognition and are homogenous in the sense of low-level image feature. The pixels in the same segment should belong to the same ship. Imagine that the top-down hypothesis mask (F, B) tries to pull the ship F out of the whole image. TPs generally consist of several well-separated segments from the background so that they are easy to be pulled out. However FPs often contains only part of the segments. In the example of tree trunks, only part of the tree trunk is recognized as foreground while the whole tree trunk forms one bottom-up segment. This makes pulling out FPs more difficult because they have to break the homogenous segments.

Based on these observations we combine the bottom-up segmentation to update the top-down figure-ground mask. Incorrect local predictions are removed from the mask if they are not consistent with the bottom-up segmentation. We give each bottom- up segment Si a binary label. Unlike the work in [17] which uses graph cut to propose the optimized hypothesis mask, we simply define the ratio $\frac{Area(s_i \rightrightarrows F)}{Area(s_i \rightrightarrows B)}$ as criteria to assign S_i to F or B. We try further segmentation when such assignment is uncertain to avoid the case of under-segmentation in a large area. The Normalized Cut (NCut) cost ([18]) is used to determine if such further segmentation is reasonable. The procedure to refine hypothesis mask is formulated as follows:

Input: top-down mask (F, B) and bottom-up segments $\{S_i, i = 1,\ldots\ldots,N\}$.

Output: refined object mask (F, B).

Set i = 0.

1) If i > N, exit; else i = i+1.
2) If A = $\frac{Area(s_i \rightrightarrows F)}{Area(s_i \rightrightarrows B)}$! k_{up} ,*then* F = F U S_i, goto 1;
   else if A < k_{down}, then F = F – (F ∩ S_i), goto 1. Otherwise goto 3.

3) Segment S_i to (S_i^1, S_i^2). If $\varsigma = NCut(S_i) > \gamma_{up}$, $F = F - (F \cap S_i)$, goto 1;
Else $S_{N+1} = S_i^1$, $S_{N+2} = S_i^1$, $S = S \cup \{ S_{N+1}, S_{N+2}\}$, $N = N + 2$, goto 1.

Re-Evaluation. There are two advantages with the updated masks. The first is that we can recomputed more accurate local features by masking out the background edges. The second is that the shapes of updated FPs masks will change much more than those of TPs, because FPs are usually generated by locally similar parts of other ships, which will probably be taken away through the above process. We require TPs must have voters from all the different locations around the hypothesis center. This will eliminates those TPs with less region support or with certain partial matching score.

The final score is the summation of the average scores over the different spatial bins in the mask. The shape of the spatial bins is predefined. For ships we use the radius- angle polar ellipse bins; for other ships we use rectangular grid bins. For each hypothesis, SC features are re-computed over the masked edge map by F and feature f_i is only allowed to be matched to 'ce_k' in the same bin location. For each bin j, we compute an average matching score E_j $f\_p(ce_k \mid f_i)$, where both `ce_k' and f_i come from bin j

The final score of this hypothesis is defined as:

$$E \quad \int_j E_j^!, where E_j^! \quad \begin{matrix} \uparrow E_j, if E_j \ ! \ \Delta, \\ \rightarrow \\ \downarrow \ \Delta, if E_j \quad 0 \end{matrix} \tag{6}$$

The term α is used to penalize the bins which have no matching with the codebook. This decreases the scores of FPs with only part of true ships, i.e. bike hypothesis with one wheel. Experiments show that our FPP procedure can prune out FPs effectively.

3 Results

Our experiments test different ship classes. These pictures were taken from scenes around campus and urban streets. Ships in the images are roughly at the same scale. For ships, the range of the heights is from 186 to 390 pixels. For our evaluation criteria, a hypothesis whose center falls into an ellipse region around ground truth center is classified as true positive. The radii for ellipse are typically chosen as 20% of the mean width / height of the ships. Multiple detections for one ground truth ship are only counted once.

Angular Blur and Mask Function Evaluation. We compare the detection algorithm on images w/ and w/o Angular Blur (AB) or mask function. For ship and umbrella detection, it is very clear that adding Angular Blur and mask function can improve the detection results. For other ship classes, AB+Mask outperform at high-precision/low-recall part of the curve, but get no significant improvement at high-recall/low-precision part. The reason is that AB+Mask can improve the cases where ships have deformation and complex background clutter. For bikes, the inner edges dominate the SC histogram; so adding mask function makes only a little difference.

In the figure 4, the intermediate results are shown. Depending on the voting scheme, the number of segments with maximum voting are matched and elevated in that particular region. In the figure 5, the final results are shown. Two results are shown. In both the cases, clearly the ships are detected. Number of SAR and SO images are tested and detected the ships in the images.

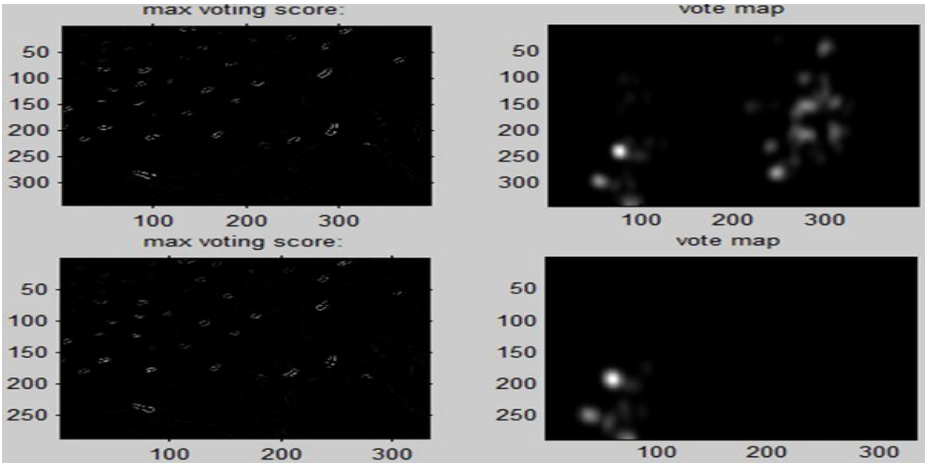
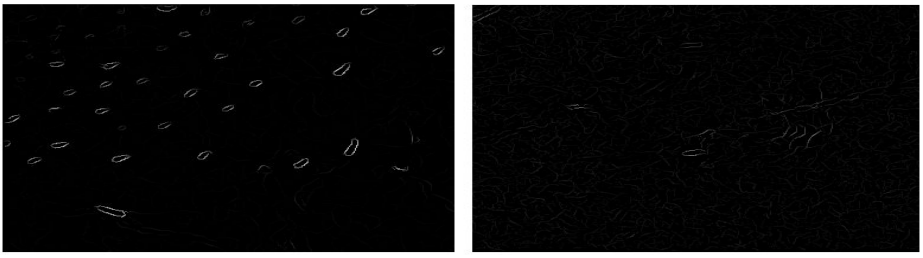


Fig. 4. Iteration Results



(a) (b)

Fig. 5. Sample Output images indicating detected ships

4 Conclusion and Future Scope

In this paper, we developed ship detection method of combining top-down model based recognition with bottom-up image segmentation. Our method not only detects ship positions but also gives the figure-ground segmentation mask. We designed an improved Shape Context feature for recognition and proposed a novel FPP procedure to verify hypotheses. This method can be generalized to many ship classes. Results show that our detection algorithm can achieve both high recall and precision rates. However there are still some FPs hypotheses that cannot be pruned. They are

typically very similar to ships, like a human-shape rock, or some tree trunks. More information like color or texture should be explored to prune out these FPs. Another failure case of SC detector is for very small scale ship. These ships have very few edges points thus are not suitable for SC. Also our method does not work for severe occlusion where most local information is corrupted. The algorithm can be used for different types of objects and images if sufficient masks are provided.

References

[1] Viola, P.A., Jones, M.J.: Rapid ship detection using a boosted cascade of simple features. In: CVPR (2001)

[2] Borenstein, E., Ullman, S.: Class-Specific, Top-Down Segmentation. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part II. LNCS, vol. 2351, pp. 109–122. Springer, Heidelberg (2002)

[3] Levin, A., Weiss, Y.: Learning to Combine Bottom-Up and Top-Down Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part IV. LNCS, vol. 3954, pp. 581–594. Springer, Heidelberg (2006)

[4] Leibe, B., Seemann, E., Schiele, B.: Ship detection in crowded scenes. In: CVPR (2005)

[5] Ferrari, V., Tuytelaars, T., Van Gool, L.: Object Detection by Contour Segment Networks. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006, Part III. LNCS, vol. 3953, pp. 14–28. Springer, Heidelberg (2006)

[6] Kokkinos, I., Maragos, P., Yuille, A.L.: Bottom-up & top-down ship detection using primal sketch features and graphical models. In: CVPR (2006)

[7] Zhao, L., Davis, L.S.: Closely coupled ship detection and segmentation. In: ICCV (2005)

[8] Ren, X., Berg, A.C., Malik, J.: Recovering human body configurations using pairwise constraints between parts. In: ICCV (2005)

[9] Mori, G., Ren, X., Efros, A.A., Malik, J.: Recovering human body configurations: Combining segmentation and recognition. In: CVPR (2004)

[10] Srinivasan, P., Shi, J.: Bottom-Up Recognition and Parsing of the Human Body. In: Yuille, A.L., Zhu, S.-C., Cremers, D., Wang, Y. (eds.) EMMCVPR 2007. LNCS, vol. 4679, pp. 153–168. Springer, Heidelberg (2007)

[11] Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for ship recognition. International Journal of Computer Vision 61(1) (2005)

[12] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: CVPR (2005)

[13] Belongie, S., Malik, J., Puzicha, J.: Shape matching and ship recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. 24(4) (2002)

[14] Mori, G., Belongie, S.J., Malik, J.: Efficient shape matching using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. 27(11) (2005)

[15] Thayananthan, A., Stenger, B., Torr, P.H.S., Cipolla, R.: Shape context and chamfer matching in cluttered scenes. In: CVPR (2003)

[16] Rubner, Y., Tomasi, C., Guibas, L.J.: A metric for distributions with applications to image databases. In: ICCV (1998)

[17] Ramanan, D.: Using segmentation to verify ship hypotheses. In: CVPR (2007)

[18] Shi, J., Malik, J.: Normalized cuts and image segmentation. In: CVPR (1997)

Automatic Speaker Recognition System

P.M. Ghate, Shraddha Chadha, Aparna Sundar, and Ankita Kambale

Rajarshi Shahu College of Engineering, Pune
{pmghate,shradhachadha08,aparnasundar20}@gmail.com
ankee276@yahoo.com

Abstract. The proposed work provides a description of an Automatic Speaker Recognition System (ASR). It particularly documents all the stages involved in the proposed ASR system starting from the preprocessing stage to the decision making stage. The main aim of this work is to achieve a system with high robustness and user friendly. Voice samples from three different users are used as acoustic material. Feature extraction is done by computing Mel Frequency Cepstral Coefficients (MFCC) which is used to create reference template. For the purpose of feature matching, Dynamic Time Warping (DTW) algorithm is used wherein DTW distance is computed between the test signal and the reference signal. Decision is made by comparing the distance with a predefined threshold value.

1 Introduction

The basic application of Speaker verification systems is to provide protection against unauthorized access in a circumstance where a speaker must be correctly recognized. Primarily speaker recognition is broadly classified as:

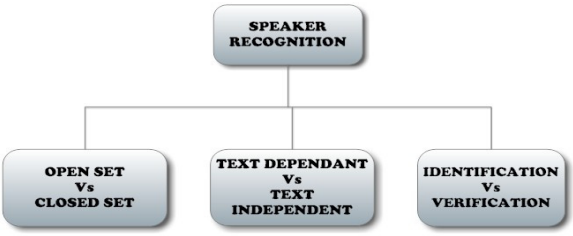


Fig. 1. Classification of Speaker Recognition System

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1037–1044.
springerlink.com

The speaker recognition is comprehensive of two basic stages as shown below:

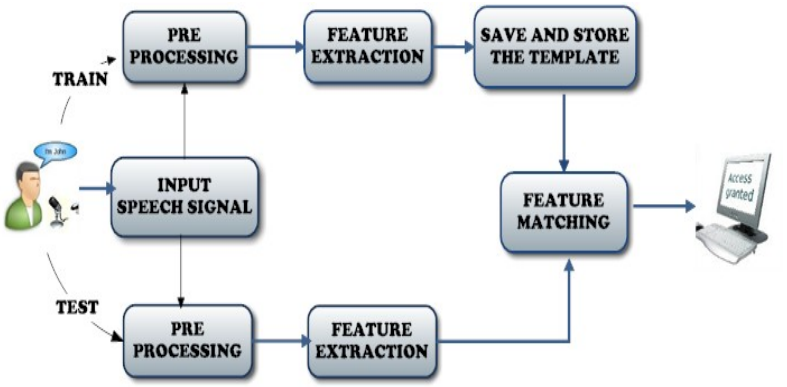


Fig. 2. Basic Block Diagram

2 Algorithm Prototype

2.1 Pre Processing

The first step is Pre Processing of the speech signal. This is done for improving the efficiency and robustness of the system. The steps involved are:

A/D conversion

Analog speech signal is converted in its digital form for further processing. To avoid aliasing effect sufficient sampling rate should be chosen. The sampling frequency should be greater than or equal to 8 KHz to avoid aliasing.

Pre Emphasis

For the human speech signal the higher frequencies get damped while the lower frequencies are boosted. So to increase the energy of the high frequencies we pass the speech signal through a high pass FIR filter which has the transfer function:

$$Y[n] = X[n] - k\,X[n-1] \tag{1}$$

The general value used for k is 0.96.

Noise Gate

The speech signal may contain some background noise which needs to be removed so that it does not affect the feature vectors.

Alignment

The speech signal is aligned to start from zero in the time axis. This will help in the feature matching process since the speech signal will become much closer to each other.

2.2 Feature Extraction

The next step is feature extraction which is used to extract the speaker's voice features from the speech signal. The algorithm used is Mel Frequency Cepstral Coefficients (MFCC). MFCC.Below is the basic block diagram of MFCC:

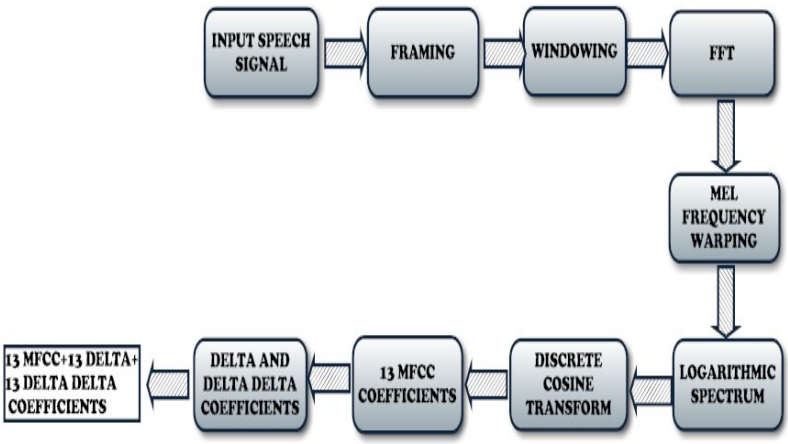


Fig. 3. Block Diagram of MFCC

Framing

As the speech signal is quasi stationary and remains periodic for 10 to 30 msec of time, so the speech signal is divided into frames of 10 to 30 msec. Overlapping is done to avoid edge effect and loss of data. Generally 50% of overlapping is sufficient.

Windowing

Windowing is done to remove the discontinuities of the signal at the extremities. A window function should have narrow main lobe and small side lobe. The window function is applied to each frame. The most commonly used window function is Hamming window. It is defined as:

$$w(n) = 0.54 - 0.46\, cos\left(\frac{2\pi N}{N-1}\right)\ 0 \leq n \leq N-1 \tag{2}$$

$$Y(n) = X(n) \times W(n) \tag{3}$$

Where,

N=No of samples in each frame

Y (n) =Output signal

X (n) =Input signal

W (n) = Window function

Discrete Fourier Transform

This step is applied to convert each frame from time domain to frequency domain. The fastest way to calculate DFT is to use FFT algorithm. The DFT is given by the equation:

$$Y_k = \sum_{n=0}^{N-1} y_n \, e^{\frac{-2\pi kn}{N}} \quad k = 0 \ldots N-1 \tag{4}$$

Mel Frequency Warping

Human hearing perception is based upon the Mel scale which is linearly spaced below 1000 Hz and logarithmic above 1000 Hz.

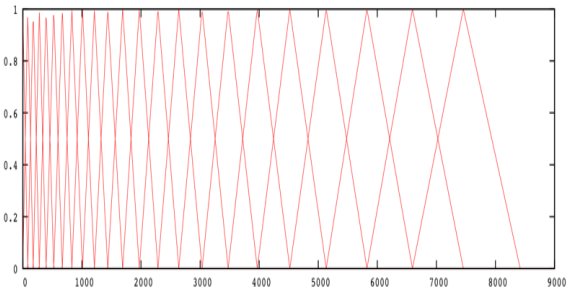


Fig. 4. Mel Frequency Filter Bank from (32 Leaves Blog)

It consists of a set of filters whose frequency response is triangular in shape. The magnitude of each filter is 1 at the center frequency and decreases to zero at the center frequency of adjacent filters. Each filter output is the summation of its filtered spectral coefficients. To calculate the Mel frequency at a given frequency f following equation is used:

$$F(mel) = 2595 \times log_{10}(1 + \frac{f}{700}) \tag{5}$$

The output is obtained as:

$$Y(i) = \sum_{i=0}^{N} S_i H_i \tag{6}$$

Where, S_i =N point magnitude spectrum

H_i =Magnitude response of filter bank

Logarithmic Spectrum

The signal is passed through the Mel Filter to mimic the human hearing. The log value of this signal is then calculated to obtain the logarithmic spectrum.

Discrete Cosine Transform

The log spectrum is then converted into time domain with the help of Discrete Cosine Transform. The output of this conversion is known as MFCC. We get 13MFCC coefficients which is speaker's feature vector.

$$c[n] = \sum_{i=0}^{M} \log\big(Y(i)\big) \times \cos\left(\frac{\pi n}{M}\left(i - \frac{1}{2}\right)\right) \tag{7}$$

Delta and Delta Delta Coefficients

The first order derivative of the feature vector is called as Delta coefficients and second order derivative is called Delta Delta coefficients. This is basically done to add time evolution information. To calculate delta coefficient following equation is used:

$$\Delta f_k[n] = \Delta f_{k+M}[n] - \Delta f_{k-M} \tag{8}$$

And Delta Delta is calculated using the equation:

$$\Delta^2 f_k[n] = \Delta^2 f_{k+M} - \Delta^2 f_{k-M} \tag{9}$$

M is typically 2 to 3 frames.

2.3 Feature Matching

In the testing phase, a test signal is compared with the stored template and a pattern matching algorithm is used to measure the similarity between the two signals. In our proposed work we intend to use Dynamic Time Warping (DTW) algorithm. The main advantage of this algorithm is its ability to efficiently measure the distance between two signals that vary in either time or speed. It finds the optimal path between the two signals that are warped non-linearly by either stretching or shrinking them in time axis.

The main aim of DTW is to compute the minimum distance between the two dynamic signals and measure the similarity between them based on the computed distance.

Consider two vectors A and B of length m and n respectively. DTW finds the path,

$\{(p_1, q_1), (p_2, q_2), \ldots, (p_x, q_x)\}$, that minimizes

$$\sum_{t=1}^{x} |A(p_t) - B(q_t)| \tag{10}$$

Certain constraints are to be applied before computing the distance for reasonable time alignment. These are:

Boundary Condition

Beginning point = (1; 1) and

Ending point = (m; n).

Monotonicity Condition

$$m_1 \leq m_2 \leq \ldots \leq m_x \text{ and } n_1 \leq n_2 \leq \ldots \leq n_x$$

This condition is used to preserve the time ordering of points.

Local Continuity Condition

For any given node (*i, j)* in the optimal path, the fan-in nodes can be (*i*–1, *j*), (*i, j*–1) and (*i*–1, *j*–1). It ensures a monotonically non-decreasing path.

The DTW distance is then calculated using the equation:

$$D(i,j) = |t(i) - r(j)| + min \begin{cases} D(i-1,j) \\ D(i-1,j-1) \\ D(i,j-1) \end{cases} \quad (11)$$

Decision Logic

The final decision is made by comparing the computed DTW distance with a predefined threshold value.

$$score(TS_i, RS_i) \begin{cases} \geq TH, speaker\ accepted \\ < TH\ , speaker\ rejected \end{cases} \quad (12)$$

Where,

TH is the verification threshold

TS_i is the test signal

RS_i is the stored reference signal

Lower the DTW distance, higher is the score. The threshold is set based on False Rejection Rate (FRR) and false acceptance Rate (FAR)

3 Methdology

The following flowchart shows the basic steps involved in speaker recognition.

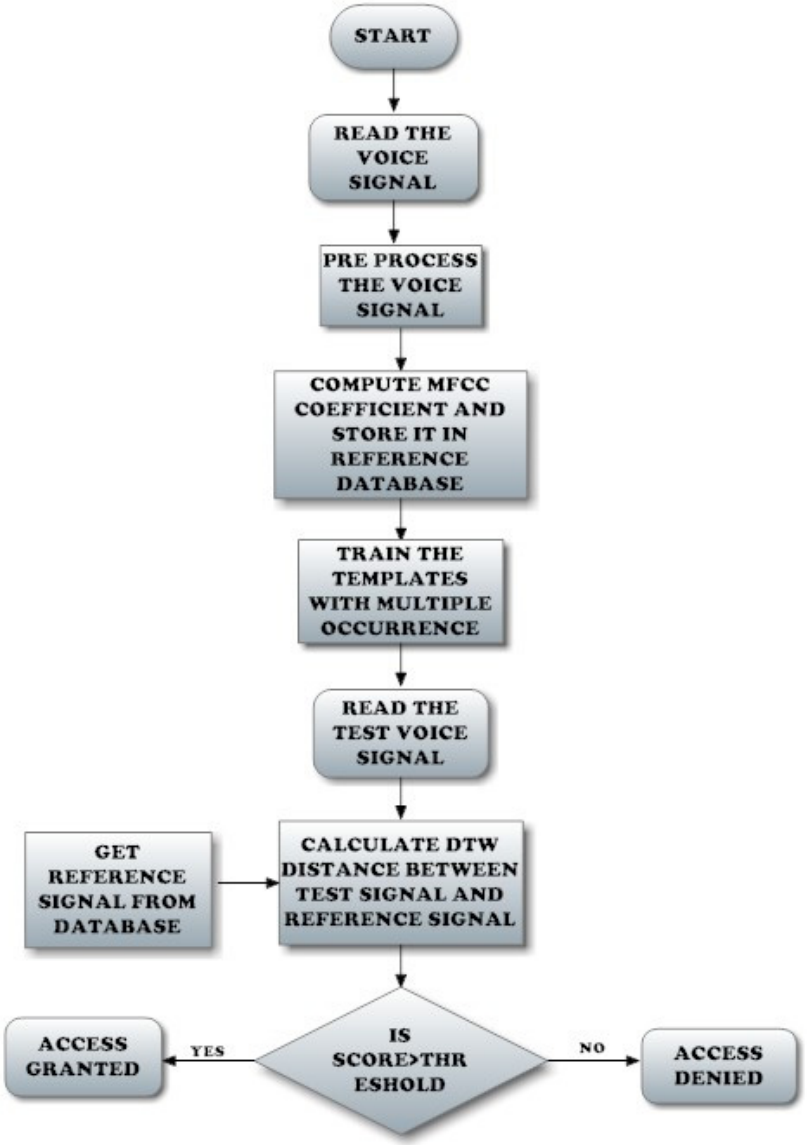


Fig. 5. Flowchart of ASR System

4 Result

PARAMETER	RESULT
ACCURACY	80%
EXECUTION TIME	25 SEC

5 Conclusion

For the proposed work it is observed that out of the various algorithms available, MFCC and DTW provide s considerable accuracy for text dependent systems. In addition to this, training of template by multiple utterances and post processing of the MFCC vectors help to improve the accuracy of the system. The system was tested with limited number of speakers and can be extended further to support large database.

References

[1] Abdulla, W., Chow, D., Sin, G.: Cross-words reference template for DTW-based speech recognition systems. In: Proc. IEEE TENCON, Bangalore, India (2003)
[2] Furui, S.: Digital Speech Processing, Synthesis and Recognition. Marcel Dekker, New York (2001)
[3] Ezzaidi, H., Rouat, J., O'Shaughnessy, D.: Towards combining pitch and MFCC for speaker identification systems. Presented at the Eurospeech Conf., Aalborg, Denmark (2001) Paper No. 2825 (unpublished)
[4] Boves, L., et al.: Design and Recording of Large Data Bases for Use in Speaker Verification and Identification. In: ESCA Workshop on Automatic Speaker Recognition, Martigny (CH), pp. 43–46 (1994)
[5] Staroniewicz, P., Majevski, W.: SVM Based Text-Dependent Speaker Identification For Large Set of Voices. In: Proc. 12th European Signal Processing Conference (EUSIPCO), vol. 1, pp. 333–336 (2004)
[6] Reynolds, D.A.: Experimental evaluation of features for robust speaker identification. IEEE Transactions on Publication, Speech and Audio Processing 2(4), 63964 (1994)

An Approach for Document Image Based Printed Character Recognition

Sushila Aghav[1] and Shilpa Paygude[2]

[1] MIT College of Engineering, Pune
sushila_aghav@yahoo.com
[2] Department of Computer Enginnering,
MIT, Pune,
shilpa.paygude@mitpune.edu.in

Abstract. Document image analysis analyzes the document images to extract the text and graphics information from image. Printed character recognition is important in the context of document image analysis. Machine learning Approach such as pattern recognition and matching can be applied to document image based printed character recognition. In this paper we discussed the Template Matching approach to printed character recognition. Template matching is found to be an effective technique to recognize printed character as compared to neural network and other classification techniques.

1 Introduction

The objective of document image analysis is to recognize the text and graphics components in images of documents, and to extract the intended information as a human would.

Document Image based Text contains useful information for automatic annotation, indexing, and structuring of images [8] Extraction of this information involves detection, localization, tracking, extraction, enhancement, and recognition of the text from a given image. However, variations of text due to differences in size, style, orientation, and alignment, as well as low image contrast and complex background make the problem of automatic text extraction extremely challenging. While comprehensive surveys of related problems such as face detection, document analysis, and image indexing can be found, the problem of text information extraction is not well surveyed. A large number of techniques have been proposed to address this problem.

Two categories of document image analysis can be defined [1] as, Textual part processing, which deals with the text components of a document image and graphics processing which deals with graphics part of document image. There are various Textual processing tasks like: Determining the skew (any tilt at which the document may have been scanned into the computer), Finding columns, paragraphs, text lines, and words, Recognizing the text (and possibly its attributes such as size, font etc.).Details of document image processing for text are explained in [8].Text Extraction and Recognition plays vital role in Document Image analysis. Text localization, text line detection, character segmentation, character recognition is the parts of Text extraction and recognition. Character recognition is the complex process in text extraction and

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1045–1050.
springerlink.com

recognition, in document image analysis context. Details of character recognition are explained in chapter 2. Character recognition uses classification technique to classify (recognize) the input character, the one which is extracted from image.

2 Document Based Image Processing

Important Document Image Processing phases are Color Image to Gray scale conversion, Grayscale to binary conversion, Skew Correction, Image Binarization (Foreground, Background separation),Text and Non Text Region Separation, Text Extraction, Character Segmentation, Character recognition.

The document image captured by camera or scanned image first converted in grayscale format to minimize the processing complexity. Image is then binarized using Image Binarization techniques. Binarization process separates the background from foreground i.e. it separate informative region from non informative part. Details of Binarization are explained in [8].If the image is skewed, the text extraction and character recognition will not be accurate. Skew detection and correction is done details of the skew detection and correction is explained in [1].The informative regions are processed further to separate text part from non text part. Text extraction extracts text from text region. Extracted text further is input to character segmentation phase wherein the each character is separated and extracted. Text part of the document image is identified and localized to extract the text of the images. For this text line localization, Text segmentation techniques can be used. Details of these techniques are explained in [8].Once the text line are detected and segmented the text regions are extracted and character segmentation will be done by using various techniques like Connected component, Region growing algorithm. The character segmentation is explained in [1].Character recognition recognizes the extracted character by comparing it with the provided character template. In the following section the character recognition phases are surveyed.

3 Printed Character Recognition in Document Image

Machine learning is found to be very useful in pattern recognition. One of the very popular applications of machine learning is Character recognition, which is recognizing character codes from their images. This is an example where there are multiple classes, as many as there are characters we would like to recognize. There are many possible images corresponding to the same character as there are different font styles.

Machine Printed Character Recognition carried out in following in the phases like: Template Character Learning, Individual Character Extraction, Character Feature Extraction, and Comparison of features of learned and extracted character.

3.1 Machine Learning of Character Templates

Character template is represented by an Image. For each character, multiple images are provided as a input to this phase. Character template Images are learned to extract features of the character.

Generating the learned set is quite simple. It requires that an image file with the desired characters in the desired font be created, and a text file representing the characters in this image file. If a character such as pi, has a multi-character translation, delimiter should be placed around the translation. Once the learned set has been read in from the image file and its properties recognized, it can be written out to a "learn" file. This file stores the properties of the learned characters in abbreviated form, eliminating the need for retaining the images of the learned characters, and can be read in very quickly.

Table 1. Character Templates

A	A	A	A	A
n	n	n	N	N
C	C	C	C	C
D	D	D	D	D
E	E	E	E	E
F	F	F	F	F
G	G	G	G	G
H	H	H	H	H
1	1	1	1	1
5	5	5	5	5

3.2 Individual Character Extraction

Character extraction divided into two phases: Text line segmentation, Detection of Connected Component [5]

3.2.1 Text Line Segmentation

Various Text line segmentation techniques are [1]

1. Projection-based
2. Hough Transform,
3. Smearing methods,
4. Grouping methods,
5. Active Contour methods.
6. Graph-based methods

The projection-based approaches are making use of the structural characteristics of the documents. They are top-down techniques, simple and easy in implementation. Hough Transform is also a popular methodology in the area of text line segmentation. It describes parametric geometric shapes and identifies geometric locations that suggest the existence of the sought shape. Serious drawback of this method is the computational complexity. The smearing methodology is a bottom-up technique. It is the process of converting a set of background pixels located between foreground pixels into foreground pixels whether their amount is less than a certain threshold. Smearing methods strengthen by local techniques, solve specific problems and overlapping

touched connected component. Moreover, these methods work successfully with documents that contain characters of variable height.

The grouping methods [1] are also bottom-up. From the lower level, the pixel, starts a process of grouping according to specific constrains designed to result to a layer of text lines. The process is relatively easy in the case of printed documents, but it may be proved to be difficult and problematic in manuscripts.

The Active Contour methods use the difference between the foreground and the background through characteristics such as brightness or color that occurs at the border contours of the object. The edge is a curved line from which derive all the properties and characteristics that describe the specific category of shapes, in our case text lines.

The representation of document images by graphs is an important tool of the line segmentation procedure. The graph is constructed as vertices of pixel or more complex connected components. The vertices are normally associated with weighted edges that depict distances between connected components. After the modeling of the document image, the treatment method can be chosen.

3.2.2 Detection of Connected Component

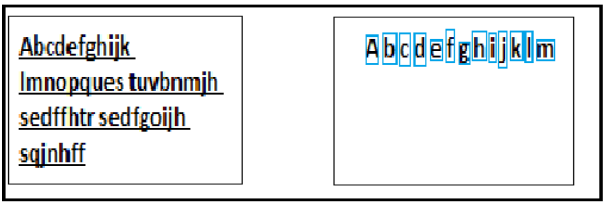


Fig. 1. (a) Text Line Segmentation (b) Connected Component Detection

To extract the connected components from each line, starting at the upper right corner of each line, removes touching intervals of black pixels from the image until nothing more connected can be found. The extraction routine then looks upward and downward to see if there are possible "extra parts", such as the dot on an 'i', hanging directly above or below the component.

3.3 Feature Extraction

Extracted Features of each character will be stored in a vector. For this, the character can be divided into equal sized region and intensity values of the pixel in particular region are considered as feature.

3.4 Comparison of Features of Learned and Extracted Character

After the segmentation of Character from document Image ,the character recognition phase recognize each individual character by calculating the correlation analysis between the template images of various character and segmented character. Template Character Features are compared with the extracted character feature [1]. Correlation analysis is done on feature vectors on template and character image.

Co-relation Function is shown in (1)

$$C = \Sigma \, |t(x, y) - t'(x, y)| \text{ for all pixel } N \quad (1)$$

As Shown in (1), Correlation analysis finds out the similarity between the feature vectors of two images. The similarity coefficient is between 0 and 1.

If the result of correlation analysis is nearer to 1 the match is good enough to recognize the character. And character said to be recognized.

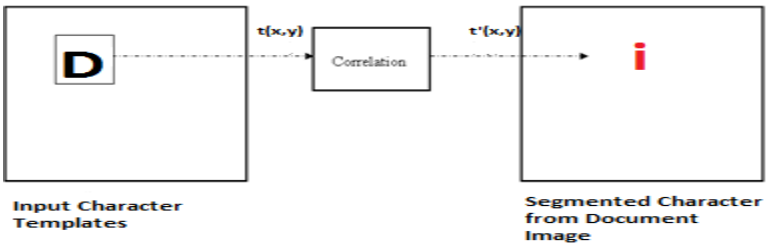


Fig. 2. Template Matching using Correlation Analysis

4 Conclusion

In this paper we reviewed various machine printed character recognition phases. There are various techniques like ANN, Template matching, which can be applied at Character recognition phase to get correct and efficient recognition results. Template Matching is an efficient technique for Printed character Recognition in document image.

The result of the character recognition can be improved by extending feature vector size and adding more character features.

References

1. Mollah, A.F., Majumder, N., Basu, S., Nasipuri, M.: Design of an Optical Character Recognition System for Camera-based Handheld Devices. IJCSI International Journal of Computer Science Issues 8(4), No.1 (June 2011)
2. Kavallieratou, E., Daskas, F.: Text Line Detection and Segmentation: Uneven Skew Angles and Hill-and-Dale Writing. Journal of Universal Computer Science 17(1), 16–29 (2010)
3. Nagy, G.: State of Art of Document Image Processing. In: 2008 SSDI P, GN, Bangalore (2008)
4. Park, J., Dinh, T.N., Lee, G.: Binarization of Text Region based on Fuzzy Clustering and Histogram Distribution in Signboards. In: World Academy of Science, Engineering and Technology 43 (2008)
5. Sushma, J., Padmaja, M.: Text Detection in Color Images. IEEE, IAME (2009); 978-1-4244-4711-4/09©2009

6. Gao, J., Yang, J.: An Adaptive Algorithm for Text Detection from Natural Scenes. In: Proceedings of the 2001 IEEE Conference on Computer Vision and Pattern Recognition (December 2001)
7. Jung, K., Kim, K.I., Jain, A.K.: Text Information Extraction in Images and Video: A Survey. Pattern Recognition 37, 977–997 (2004)
8. O'Gorman, L., Kasturi, R.: Document Image Analysis. Library of Congress, Number 97-17283, ISBN 0-8186-7802-X

Flexibility in Supplier Selection Using Fuzzy Numbers with Nonlinear Membership Functions

Atul Kumar Tiwari[1], Cherian Samuel[1], and Anunay Tiwari[2]

[1] Institute of Technology, Banaras Hindu University Varanasi
atultiitb@gmail.com
csamuel.mec@itbhu.ac.in
[2] Tapasthali Vidyashram Society, Varanasi
tiwari.anunay@gmail.com

Abstract. Supplier selection is a multi-criteria decision making problem. It can critically influence the competitiveness in the entire supply chain network. It simultaneously takes into account several quantitative and qualitative factors. In this paper, a fuzzy approach is utilized for the supplier selection problem, using some important criteria as discussed in the supply chain literature, e.g. net purchasing cost, quality, service and commit date deliveries. These parameters are defined using fuzzy numbers with nonlinear membership functions to handle the vagueness, uncertainty and the inexactness in the data and to capture a real life like scenario. Further they are weighted using Analytic Hierarchical Process, for their relative importance so as to match the organization's ultimate goal. A multi objective program is modeled to determine promising suppliers and the ordered quantities in a multi-product, quantity discounted environment. Finally a single objective fuzzy linear program is devised to solve the proposed model efficiently. The model developed in this paper is illustrated with the help of a numerical example.

Keywords: supply chain, supplier selection, exponential fuzzy numbers, multi product, volume discount, multi criteria, multi objective, Analytic Hierarchical Process.

1 Introduction

Supplier selection is one amongst the most critical issues being faced by corporate houses and industries in modern supply chains that are operating in extremely competitive business environments. Selecting better suppliers certainly increases strategic position of the manufacturer/buyer in its supply chain. The buyers may bargain for better purchase prices, better qualities, better service and better delivery schedules from their suppliers to serve their customers more effectively and efficiently and hence they can make a more profitable business. However supplier selection is no easy task as there can be numerous factors that might need decision makers' attention, few of which may be contradicting. Some of these factors may be quantitative in nature while others qualitative. Besides it, in an environment of multi products procurement and quantity discounts, modelling of such problems become

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1051–1060.
springerlink.com

extremely difficult. In this paper we investigate for supplier selection model in such an environment. The remaining of the paper is organized as follows. Section 2 cites literature review. In section 3, a Multi Objective Program is developed for optimizing criteria considered in this paper. In section 4, a numerical example is taken to illustrate the model while formulating an equivalent fuzzy LP as in section 3. Paper concludes with the summary and future research scope in section 5.

2 Literature Review

Many previous studies on supplier selection and evaluation defined numerous evaluation criteria and selection frameworks for supplier selection. For example, Dickson [1] surveyed buyers for identifying and listing those factors which they had considered in selecting their suppliers. Out of the 23 factors considered, Dickson concluded three criteria namely quality, delivery, and performance history to be most important. In an another study, Weber et al.[2], derived key factors thought to influence supplier selection decisions most. These factors were obtained from 74 related articles that had appeared since Dickson's well-known study. Based on this rigorous review of vendor selection and evaluation methods, they concluded price to be the highest-ranked factor which was followed by delivery and then quality. Weber et al. considered some more factors e.g. geographical location, which he declared to enjoy more importance over the factors suggested by Dickson. Raja et al [3] used mixed integer programming model to select suppliers and determine the order quantity. The model considers uncertainty in demand and is modeled to optimize total purchasing cost and cost of receiving poor quality. Amid et al. [4,7] formulated a fuzzy multi-objective linear programming model for selection of suppliers. The model could handle the vagueness and imprecision of input data, and helps the decision maker to calculate the optimal order quantity from each supplier. They included three objective functions with different weights assigned to each objective in their model and they developed an algorithm to solve this model. Amid et al. [5] formulated a fuzzy multi-objective mixed integer linear programming model to solve the vendor selection problem. The approach is very similar to that in Amid et al.[4] in terms of the of objective functions considered in the model, the specific criteria used to evaluate the suppliers and the solution methodology that is used to solve the model. The only difference being that quantity discount was considered in this paper. The price discount schedule was based on the total quantities ordered. Amid et al. [6] developed a weighted max-min fuzzy model to handle the vagueness and imprecision in the information. The weights used in the problem are the weights calculated in their previous paper by using analytic hierarchy process (AHP). The paper is very similar to the previous two except the application of the max-min fuzzy operator which was coined by Zadeh. Saaty and many other authors [8,15,16,17,18] have widely used Analytic Hierarchy Process (AHP) in different MCDM problems for its ease to use, simplicity and the greater flexibility. However the output of AHP in this context is only to assign relative weights to the criteria considered. Therefore different authors integrate different approaches with AHP to handle supplier selection effectively. Chan and Kumar [9,15] used a fuzzy AHP for supplier selection. In this approach, the linguistic preferences were changed into the triangular fuzzy numbers for the pairwise

comparison scale and a fuzzy synthetic extent analysis method was used to represent decision makers' comparison judgment and decide the final priority of different criteria. Bayrak et al. [10] utilized a fuzzy supplier selection approach to rank the technically efficient suppliers according to both predetermined and the product related performance criteria. Method is based on finding fuzzy suitability indices for the most efficient supplier alternatives and then ranking these fuzzy indices to select the best supplier. Li et al. [11] considered a supply contracting problem under uncertainty of price and demand in a dynamic market. They compared the selection of a long term contract supplier over the periodic purchase from the spot market and then developed a stochastic dynamic programming for a time horizon to incorporate the purchasing commitments. This paper differs significantly to the papers published prior to it. First we employ nonlinear membership functions, more precisely the exponential functions for fuzzy numbers to model supply parameters. These are more realistic in capturing and addressing real life scenarios. To the best of our knowledge such functions have not been used earlier in the literature in context of supplier selection in supply chains. Further, Amid et al. [5], and Wang et al. [12] discuss supplier selection in quantity discount environment for a single product but in this paper we consider a multiproduct discounted scenario for supplier selection. Adding to it, we have modelled MOP considering optimization based on four criteria, total purchasing cost service level, quality level and late deliveries. This combination of criteria is unique and more inclined to industrial applications.

3 The Multi Objective Supplier Selection Model

Supplier selection model is a multi-objective optimization model. There may be a number of criteria that simultaneously are supposed to be at their best, some of which may be contradicting to other e.g. quality has to be maximised and total purchasing cost has to be minimised. Therefore a trade-off is required between the criteria. All the objectives of the model do not take their best solution individually but they provide best compromised optimal solution altogether.

3.1 Assumptions

Following set of assumptions, index, parameters and decision variables are considered for the formulation of multi-objective model. Assumptions of the model are as follows.

1. Lead times are known with certainty.
2. Demand for different variety of products is known to lie in a specific given range.
3. Replenishment policy and inventory decisions are not being considered.

3.2 Indices and Parameters

i	index for suppliers	$i = 1, 2 \ldots i$
j	index for price levels for quantity discounts	$j = 1, 2 \ldots J$

v index for variety of products $v = 1, 2...V$

c_{ivj} Price of unit item of product v at price level j from supplier i.

D_v Demand of product v

Q_i Acceptable Quality level from supplier i

S_i Service level available (in percentage) from supplier i

M_i Percentage of missed items on committed delivery dates from supplier i

CAP_{iv} Capacity of product v available at all price levels from supplier i.

Decision Variables

x_{ivj} Order quantity given to supplier i for product type v at price level j.

Binary Decision Variables

$$Y_{ivj} = \begin{cases} 1 \text{ if } x_{ivj} \neq 0 \\ 0 \text{ if } x_{ivj} = 0 \end{cases}$$

MOP Model

Multi Objective mixed integer linear Program (MILP) is as follows Minimize

$$\sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} c_{ivj} x_{ivj} \tag{1}$$

$$\sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} (1 - Q_i) x_{ivj} \tag{2}$$

$$\sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} (1 - S_i) x_{ivj} \tag{3}$$

$$\sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} M_i x_{ivj} \tag{4}$$

Subject to

$$D_v = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ivj} \tag{5}$$

$$D = \sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} x_{ivj} \tag{6}$$

$$\sum_{j=1}^{J} x_{ivj} \le CAP_{iv} \tag{7}$$

$$\mathbb{Q}_{i,j-1} Y_{ivj} \le x_{ivj} < \mathbb{Q}_{i,j} Y_{ivj} \tag{8}$$

$$\sum_{j=1}^{J} Y_{ivj} \le 1 \forall i, v \tag{9}$$

$$Y_{ivj} = \{0,1\} \tag{10}$$

$$x_{ivj} \ge 0,\ x_{ivj} \in Integers \tag{11}$$

Equations (1) to (4) are the objective functions which are to be minimised. Equation (1) is to minimise the total purchasing cost. Equation (2) denotes minimisation of poor quality materials. Here Q_i is acceptable quality from supplier i in time period t. Equation (3) is to ensure minimization of poor service from potential suppliers. Here S_i is a service standard available in percentage per article. Equation (4) is to minimize the number of article with missed commit dates. M_i is Percentage of missed items on committed delivery dates from supplier i. A total of these four objective functions have been defined subject to constraints (5) to (11). Constraint (5) ensures total order placed to all suppliers for product v at all price levels and should be equal to the total demand for this product. (6) is to ensure that total order placed to all suppliers for all products at all price should be equal to the total demand for all products. Capacity constraint is imposed with equation (7) and it means that for each product v, the total order placed to any supplier should be less than the capacity of supplier at all price levels. Equation (8) to (10) is to describe quantity discount schedule. Order quantity given to supplier i for product type v at price level j should fall in one and only one of the price discount schedule as in (8). $\mathbb{Q}_{i,j}$ is quantity specified by supplier i where price breaks at level j. Equation (9) ensures that at most one price level is selected for each product, from each supplier. Equation (10) describes nature of variable Y_{ivj} is binary and constraint (11) is for non negativity of decision variable x_{ivj}. This is hereby declared to be an integer variable.

4 Numerical Illustration

We have designed a numerical problem to illustrate the proposed model. Consider a firm which requires two types of parts/products from its suppliers e.g. A Pizza house may require 8" and 10" pizza bases from confectionaries (their suppliers).

The suppliers provide all quantity price discount schedule for various products and they have capacity allocation for each kind of product they supply to the individual buyers. Data as collected in Table 1 is used to carry on the specific problem.

Table 1. Collected data for numerical example

Supp -lier	Variety of product	Quantity discount schedule	per unit price (in Rs)	%of good quality	% of good service	% of missed deliveries	Capa -city
1	1	Q<5000	10	95	85	0.5	20K
		5000<=Q< 8000	9				
		Q>=8000	8				
	2	Q<5000	11				20K
		5000<=Q< 8000	10				
		Q>=8000	9				
2	1	Q<4000	9.5	90	80	0.1	25K
		4000<=Q< 7500	9				
		Q>=7500	8				
	2	Q<4000	10.5				20K
		4000<=Q< 7500	10				
		Q>=7500	9				
3	1	Q<6000	9	85	10	0.25	30K
		6000<=Q< 10000	8.5				
		Q>=10000	8				
	2	Q<6000	10				25K
		6000<=Q< 10000	8.5				
		Q>=10000	7.5				

First MOLP Model is formulated using the optimizing functions and the constraints as in equations (1) to (11) and each objective is optimized one by one relaxing other objectives using software Lingo 11.0 on an Intel® 1.60GHz Processor with 1 GB RAM.. Doing so, we find the best value for the the objective being considered. Further we find the values of other objectives at this solution. Carrying out the same procedure for each of objectives we have a set of solution as table 2. This data is used to fuzzify the objectives using the best value and the worst value for each of the objectives. The fuzzy non-linear exponential membership functions for each of the objectives are constructed in equation (12) to (15) to capture a real life scenario of vagueness and imprecision in the information.

Table 2. Data obtained from LINGO for membership functions

Objective functions	Z1*,Z2, Z3, Z4			Z1, Z2*,Z2, Z3			Z1, Z2, Z3*,Z4			Z1,Z2, Z3, Z4*		
	X113=20K	X213=25K	X223=20K	X113=20K	X123=20K	X213=25K	X113=20K	X123=20K	X213=10K	X113=20K	X123=20K	X213=10K
	X313=30K	X323=25K		X223=20K	X313=29K	X322=6K	X223=15K	X313=30K	X323=25K	X223=15K	X313=30K	X323=25K
Z1	967500*			1003000			982500			982500		
Z2	13750			11750*			12750			12750		
Z3	17500			18500			16500*			16500		
Z4	688			738			588			588*		

$$\mu_{Z_1} = \begin{cases} 1, & Z_1 \le 967500 \\ \dfrac{e^{S(967500-Z1)/35500} - e^{-S}}{1-e^{-S}}, & 967500 \le Z_1 \le 1003000 \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$\mu_{Z_2} = \begin{cases} 1, & Z_2 \le 11750 \\ \dfrac{e^{S(11750-Z_2)/2000} - e^{-S}}{1-e^{-S}} & 11750 \le Z_2 \le 13750 \\ 0, & \text{otherwise} \end{cases} \quad (13)$$

$$\mu_{Z_3} = \begin{cases} 1, & Z_3 \le 16500 \\ \dfrac{e^{S(16500-Z_3)/2000} - e^{-S}}{1-e^{-S}} & 16500 \le Z_3 \le 18500 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

$$\mu_{Z_4} = \begin{cases} 1, & Z_4 \le 588 \\ \dfrac{e^{S(588-Z_4)/150} - e^{-S}}{1-e^{-S}} & 588 \le Z_4 \le 738 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

The decision parameters $\lambda_1, \lambda_2, \lambda_3$ and λ_4 to assign weights to these criteria is determined by using Analytic Hierarchical Process. A quick example is as given below using table 3. Entries in the cell are indicative of importance of one parameter over the other. Eigen values are calculated in the last column. The eigenvector for the above matrix comes out to be (0.287908, 0.154689, 0.080951, 0.476452) (For more on AHP, please refer Saaty [8]) indicating Delivery to be most important and Service to be least. These eigenvectors correspond to the weights of these parameters and are found to be consistent on consistency index.

Table 3. Matrix for relative importance of parameters

	Purchasing Cost	Quality	Service	Delivery	Eigenvector
Purchasing Cost	1	2	4	1/2	0.287908
Quality	1/2	1	2	1/3	0.154689
Service	1/4	1/2	1	1/5	0.080951
Delivery	2	3	5	1	0.476452
Totals					1.00000

Now we formulate a crisp single objective linear programming with above fuzzy numbers and least satisfaction level of each of the objectives λ.

$$\text{Max} \quad \lambda = 0.288\lambda_1 + 0.154\lambda_2 + 0.081\lambda_3 + 0.476\lambda_4 \tag{16}$$

Subject to,

$$\lambda_1 \le \frac{e^{S\left(967500 - \sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} x_{ivj} c_{ivj}\right)\Big/ 35500} - e^{-S}}{1 - e^{-S}} \tag{17}$$

$$\lambda_2 \le \frac{e^{S\left(11750 - \sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} (1-Q_i)\, x_{ivj}\right)\Big/ 2000} - e^{-S}}{1 - e^{-S}} \tag{18}$$

$$\lambda_3 \le \frac{e^{S\left(16500 - \sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} (1-S_i)\, x_{ivj}\right)\Big/ 2000} - e^{-S}}{1 - e^{-S}} \tag{19}$$

$$\lambda_4 \le \frac{e^{S\left(588 - \sum_{i=1}^{I}\sum_{v=1}^{V}\sum_{j=1}^{J} (1-S_i)\, x_{ivj}\right)\Big/ 150} - e^{-S}}{1 - e^{-S}} \tag{20}$$

along with constraints (5) to (11).

Software LINGO is used to solve is problem. The compromise solution is as tabulated below in table 4 and 5. Maximum satisfaction for all the objectives together is 0.745.

Table 4. Compromise solution of the crisp LP formulated

Max λ	Z1	Z2	Z3	Z4
0.745	967500	12750	16500	588

Table 5. Optimal Order Quantity

Product	Order Quantity
X113	20000
X123	17000
X213	25000
X313	30000
X323	30000

5 Summary and Concluding Remarks

Supplier selection is a multi-criteria decision making process. Few of these criteria may be conflicting. Prioritizing the criteria is a must to meet the goal of the organization. Finally based on these priority suppliers are selected for different products and quantities. Exponential fuzzy numbers are used as membership functions for selected criteria so as to capture real life decision making situation where data are vague and not known precisely. Then we formulated multi objective optimization programme to select suppliers and quantity to be ordered from them considering compromised total purchase cost, quality level, service level and on-time deliveries. A multi supplier, multi-product, total quantity discount environment is considered for the same. This approach reflects a real world scenario. This multi-objective programme is then used to fuzzify objectives and then transformed to a crisp single objective LP programme. Finally, A numerical example with three suppliers, two variety of products and three discounted price levels is illustrated based on the model developed. Non-linearity in membership function of objective functions and its resembling to the real life is still open for future research. Moreover a multi horizon stochastic set up of supplier can be taken up for further investigations.

References

1. Dickson, G.M.: An analysis of vendor selection systems and decisions. Journal of Purchasing 2, 5–17 (1966)
2. Weber, C.A., Current, J.R., Benton, W.C.: Vendor selection criteria and methods. European Journal of Operational Research 50, 2–18 (1991)
3. Kasilingam, R.G., Lee, C.P.: Selection of vendors–A mixed-integer programming approach. Computers & Industrial Engineering 31(1), 347–350 (1996)

4. Amid, A., Ghodsypour, S.H., O'Brien, C.: A weighted max–min model for fuzzy multi-objective supplier selection in a supply chain. International Journal of Production Economics, 1–7 (2010)
5. Amid, A., Ghodsypour, S.H., O'Brien, C.: Fuzzy multi objective linear model for supplier selection in a supply chain. Int. J. Production Economics 104, 394–407 (2006)
6. Amid, A., Ghodsypour, S.H., O'Brien, C.: A weighted additive fuzzy multi objective model for the supplier selection problem under price breaks in a supply chain. Int. J. Production Economics 121, 323–332 (2009)
7. Kumar, M., Vart, P., Shankar, P.: A fuzzy goal programming approach for supplier selection problem in a supply chain. Computer and Industrial Engineering 46, 69–85 (2004)
8. Saaty, T.: The analytical hierachy process. McGraw Hill, USA (1980)
9. Chan, F.T.S., Kumar, N., Tiwari, M.K., Lau, H.C.W., Choy, K.L.: Global supplier selection: a fuzzy-AHP approach. International Journal of Production Research 46(14), 3825–3857 (2008)
10. Bayrak, M.Y., Çelebi, N., Taşkin, H.: A fuzzy approach method for supplier selection. Production Planning & Control 18(1), 54–63 (2007)
11. Li, S., Murat, A., Huang, W.: Selection of contract suppliers under price and demand uncertainty in a dynamic market. European Journal of Operational Research 198, 830–847 (2009)
12. Wang, T.Y., Yang, Y.H.: A fuzzy model for supplier selection in quantity discount environments. Expert Systems with Applications 36, 12179–12187 (2009)
13. Zadeh, L.A.: Fuzzy sets. Information and Control 8(3), 338–353 (1965)
14. Wang, Y.-M., Chin, K.-S., Leung, J.P.-F.: A note on the application of the data envelopment analytic hierarchy process for supplier selection. International Journal of Production Research 47(11), 3121–3138 (2009)
15. Chan, F.T.S.: Interactive selection model for supplier selection process: an analytical hierarchy process approach. International Journal of Production Research 41(15), 3549–3579 (2003)
16. Kuo, R.J., Lee, L.Y., Hu, T.-L.: Developing a supplier selection system through integrating fuzzy AHP and fuzzy DEA: a case study on an auto lighting system company in Taiwan. Production Planning & Control 21(5), 468–484 (2010)
17. Zaim, S., Sevkli, M., Tarim, M.: Fuzzy Analytic Hierarchy Based Approach for Supplier Selection. Journal of Euromarketing 12(3), 147–176 (2003)
18. Sen, S., Basligil, H., Sen, C.G., BaraCli, H.: A framework for defining both qualitative and quantitative supplier selection criteria considering the buyer-supplier integration strategies. International Journal of Production Research 46(7), 1825–1845 (2008)

Fuzzy Based Interference Reduction in Cognitive Networks

Lavanya G., Pandeeswari S., Shanmugapriya R.K., and Umamaheswari A.

Department of Information Technology,
Sri Krishna College of Technology, Kovaipudur, Coimbatore-42
lavanya_joyce@yahoo.co.in,
{pandeeswaris91,rkpriyait,aumas.it}@gmail.com

Abstract. Cognitive networks are wireless network consisting of cognitive radios, primary user, and secondary user. These networks become significant for the prevailing apparent lack of spectrum under the current spectrum management policies. Cognitive user in the network should communicate without having the primary user communication. The main issues in these networks are high interference and low received signal power at the primary receiver side. In this paper, we analyze the cognitive network interference and signal power variation in order to provide an optimal solution to reduce the interference caused due to cognitive radios. This paper exposes our interference reduction techniques which employ fuzzy inference system, to solve the issue. The proposed system reduces the complexity in existing interference reduction techniques. The results rendered by our system helps in ease of spectrum allocation.

1 Introduction

With the emergence of new wireless applications and devices, there is a dramatic increase in the demand for radio spectrum. This problem is solved by choosing cognitive wireless networks. Because of opportunistic spectrum access has the possibility to improve spectrum utilization [8], it allows the reuse of unused bandwidth [6]. It implies that the under-utilized portion of the licensed primary user is for reuse, provided that the transmission of secondary user is facilitated. The cognitive radios [2] should accurately detect and access the idle spectrum [4]. The main challenge to opportunistic spectrum access lies in finding balance in conflicting goals of satisfying performance requirements of secondary user while minimizing the interference to the active primary users and the secondary users. Cognitive networks can adapt their operational parameters in response to user needs or changing environmental conditions. These networks can learn from dynamic adaptations and exploit knowledge to make future decisions. Cognitive networks are the future evolving area, which are required because of their allowable focus on parameter other than configuring and managing networks. They can be characterized by using their self-attributes such as self-managing, self-optimizing, self-monitoring, self-repair, self-production, self-adaptation, self-healing to adapt dynamically to changing

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1061–1068.
springerlink.com

requirements or component failures while taking into account the end-to-end goals. Cognitive networks comfort to provide better production against security attacks and intruders by analyzing feedback from various layers of the network. In cognitive network, if the secondary user is reused by the unused portion of the primary user it may cause interference [10] to the primary user during the transmission. The issue of interference in the primary receiver due to the cognitive network is addressed in our proposed system.

Fuzzy technique is employed in the proposed system to reduce the interference in simple way. In the proposed system the primary users and secondary receiver are uniformly distributed in a circular disk. The secondary transmitter is located at the center of the disk is allowed to transmit concurrently with the primary transmitters. The secondary transmitter is equipped with multiple channels which causes interference on the primary receiver. Due to simultaneous transmission of both primary and secondary transmitter the received signal power will be low. The two main issues to be addressed in the cognitive networks are high interference and low received signal power. The interference in the cognitive network can be reduced in two ways.

Simon Yiu et al has developed a Beamformer [9] to maximize the cognitive user's signal-to-interference ratio (SIR), mathematical tools from random matrix theory is derived for both lower and upper bounds on the average interference at the primary receivers and the average SIR of the cognitive user [7]. Patrick Mitran addresses the size of the primary network is fixed and they have analyzed how quickly the interference threshold limit of the primary network can be reduced as a function of secondary network size.

Here the tradeoff is determined in the regime that the interference decreases sufficiently fast for Rayleigh fading. The issue of interference in the primary receiver due to the cognitive network is addressed in our proposed system. Fuzzy technique is employed in the proposed system to reduce interference in simple way. This implies that the interference to the primary user can be minimized based on dynamic allocation [4] of number of cognitive transmitters. An interference solution is provided for cognitive network using fuzzy inference system.

In this paper, the following section deals with cognitive network architecture and system model, fuzzy based interference reduction system, simulation results and the conclusion for the proposed system.

2 Network Architecture

Consider a cognitive network [2] with primary users and multiple cognitive users as shown in Fig.1 Notations used for the system description are shown in Table 1. Number of primary users can range as $\mathbf{1} \leq i \leq N_P$**.** Each primary user and cognitive user requires a transmitter and receiver for its transmission.

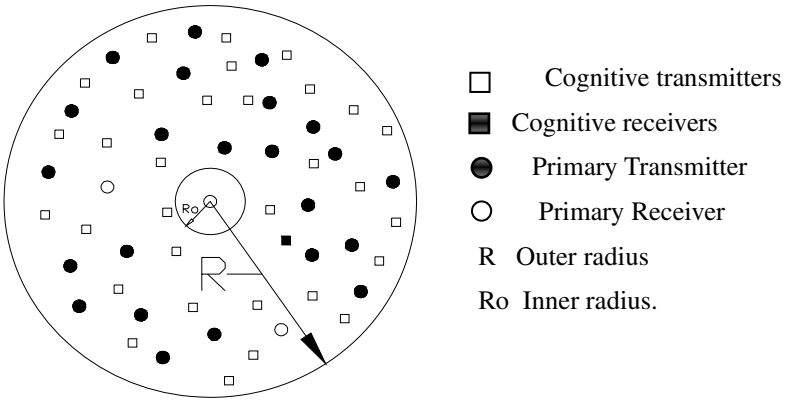


Fig. 1. Network Model

At the same time, we assume that each receiver (either primary or secondary) has a protected radius of $R_0 > 0$ without any interfering transmitter inside. These assumptions impede infinite interference at any receiver in the network. We consider only the interference at the primary receivers created by the cognitive transmitter. The average total interference created by the cognitive transmitter C_T and average Cognitive signal to interference ratio (CSIR) are given by the equation (1) & (2) respectively

$$F\,[I_{TOT}] = [\Sigma B e \sigma_c^2] \tag{1}$$

B represents transmitting power constant.

$$F\,[CSIR] = \frac{\text{power of the received signal at } C_R}{I+\sigma_C^2} \tag{2}$$

where the function is taken over the distribution of C_R and P_R .

σ_c^2 represents the noise variance at C_R

The network accepts the cognitive transmitter C_T equipped with N_T uncorrelated channels whereas the cognitive receiver C_R and the primary receivers P_R, $1 \le i \le N_P$ are equipped with a single channel. Channel allocation is denoted by N_T x 1 channel vector from C_T to P_T as ω_0 and from C_T to C_R as ω_1.

The path loss exponent is considered to be $\alpha \ge 2$, which are independent and identically distributed zero mean complex Gaussian random variables with unit variance.

Table 1. Notations

Symbol	Description
I_{TOT}	Total Interference
N_P	Number of primary users
N_{CT}	Number of cognitive users
C_T	Cognitive Transmitter
C_R	Cognitive Receiver
P_T	Primary Transmitter
P_R	Primary Receiver
SIR	Signal to interference ratio
CSIR	Cognitive signal to interference ratio
I	Interference
B	Transmitting power constant
ω_0, ω_1	Channel capacity
E	Channel lower limit
R	Random distance between primary transmitter and cognitive receiver

The distance R_d represents the coverage area distance of the primary receivers. Finally in order to provide theoretical bounds for the considered network, it is assumed that C_T has global state information of the network, i.e., complete knowledge of ω_0 and ω_1.

The cognitive transmitter C_T employs a fuzzy logic vector ω with dimension N_Tx1 for transmission of its data symbol X. The corresponding received signal at C_R and P_R are given in equation (3)

$$\text{Power of received signal} = 1/(r)^2 \tag{3}$$

3 Mathematical Models of Interference and CSIR

We analyze the relation of interference and CSIR variation with change in coverage area distance [3] and also received signal power by varying R_d and N_{ct} mathematically. Let the data symbol X in equation (4) are taken from M-ary symbol alphabet. Therefore, the instantaneous total interference [10] created by C_T is given by

$$F_I(x) = P(I<x) \tag{4}$$

The generic interference I [1] created by the constant number of primary receiver is given in equation (5)

$$I=BL\gamma \tag{5}$$

where γ - path loss
B - Transmitting power constant
L - Shadowing variable

The primary signal strength is calculated using the equation(6)

$$S = ALPr_p^{-\gamma} \tag{6}$$

Finally, the total interference [1] is expressed and calculated using the equation (7)

$$F_I(x) = 1 - F_z(\omega_1|\sigma_x) + \frac{1}{R^2 - R_0^2}[(R^2 F_z(\omega_1|\sigma_x) - R^2 F_z(\omega_0|\sigma_x)) - (\beta|x)^{\frac{2}{\gamma}} r_0^{\frac{2\sigma_x^2}{\gamma^2}} [F_z\left(\omega_1 + \frac{2\sigma_x^2}{\gamma}\middle|\sigma_x\right) - F_z\left(\omega_0 + \frac{2\sigma_x^2}{\gamma}\middle|\sigma_x\right)]] \tag{7}$$

Where, $\omega_0 = \ln(\gamma R^{-\delta})$, $\omega_1 = \ln(\gamma R_0^{-\delta})$

In equation (7), R and R_0 is the outer and inner radius of the circle.

We presuppose that escalating number of channels and radius of the primary coverage area improve the signal power, one would expect the higher average CSIR and a lower average interference as a relation expressed in equation (8),

$$\text{CSIR= power of received signal / } I_{TOT} \tag{8}$$

Based on the predefined equations and formulae, the analysis was conceded to determine the variation of Interference, CSIR and received signal power, with respect to R_d. The result of analysis is exposed in section (5).

4 Fuzzy Based Interference Reduction System

Fuzzy Logic Controller (FLC) is based on fuzzy logic and constitutes a way of converting linguistic control strategy into an automatic by generating a rule base which controls the behavior of the system. Fuzzy provides a remarkably simple way to draw definite conclusions from vague ambiguous or imprecise information. It is suitable for any applications that involves in decision making or control of any parameter in a technical system.

As the cognitive networks [2] have lot of non linear parameters within itself, controlling any parameter of the network becomes a complex process. Using Fuzzy inference system in controlling the dynamic nonlinear parameters has high regard compared to other classical controlling techniques. Pros of using fuzzy logic are simplicity of control, low cost and the possibility to design with/without knowing the exact mathematical model of the process. Fuzzy logic incorporates an alternative way of thinking which allows modeling complex systems using higher level of abstraction originating from the knowledge and experience. Fuzzy logic can be described simply as computing words rather than numbers or control with sentence rather than equations.

Based on the analysis of interference, CSIR and power variation with respect to R_d and N_{CT}, fuzzy inference system is designed for reducing the interference in the network. The general block diagram for the fuzzy based interference reduction system is shown in Fig.2 In the proposed system, we assume the network with the parameters given in table 2, for the primary receiver in the cognitive networks. The primary receiver is allowed to receive the signal throughout. Instantaneous CSIR and received signal power is measured and fed as input to the fuzzy logic controller. Fuzzy inference system in the Fuzzy logic controller provides the allowable number of cognitive users as the output to the spectrum access controller. Based on the decision

of the rules in the inference system, the output is offered to the spectrum access controller. Strength of cognitive radios will be dynamically changing according to the signal interference and received signal power.

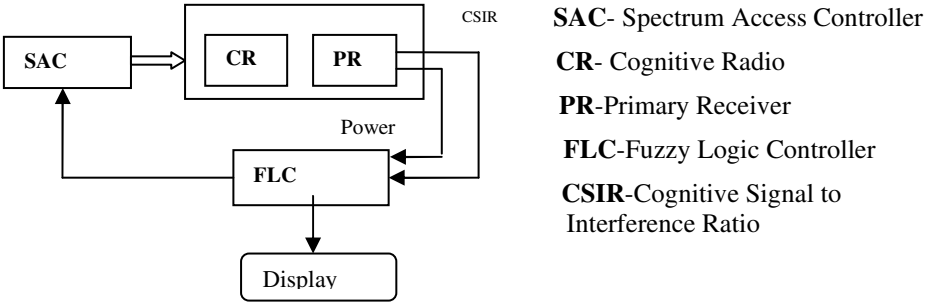


Fig. 2. Block diagram of fuzzy based interference reduction systems

The FIS editor is used to handle the high level issues of the system. Two input fuzzy inference system producing single output is designed based on the analysis of mathematical model referred in section 3. Fuzzy inference system for the proposed system with CSIR and signal power as input to the system, consisting three member functions are designed with the following perspective. The member functions consist of range, display range and numerical parameters. Each member functions are derived based upon the analysis in the section 3. The member function is chosen to be a Gaussian function. Each input consists of three member function with the fuzzy variables low, medium and high. Member functions for CSIR and received signal power is assigned as shown in fig. 4(a) and 4(b) respectively.

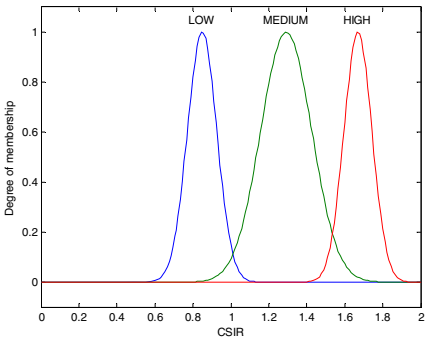


Fig. 4(a). Membership function of CSIR

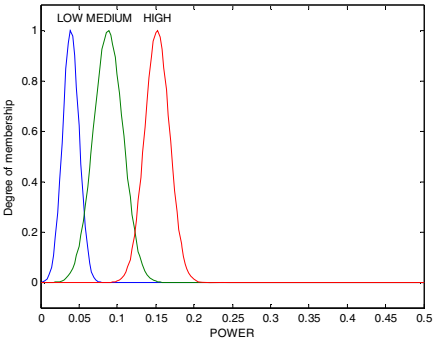


Fig. 4(b). Membership function of received signal power

Based on the description of input variable defined in the FIS editor. The rules for fuzzy logic system is framed in the designed FIS. Finally, The designed FIS is evaluated to give the desired number of cognitive users allowed for a particular state in the network.

5 Simulation Results

The mathematical analysis was carried out in MATLAB. Using the obtained results in the above simulation, fuzzy variables are assigned with the range, which decides the member function of the inputs in FIS. The designed fuzzy inference system shows the optimum number of cognitive radios that can be allowed for transmission in the cognitive networks in order to maintain a low level of consistent interference. The variation of interference is represented in cognitive signal to interference ratio(CSIR) in our system. The FIS provides the relation of CSIR and received signal power against the number of cognitive users. The crisp results of FIS are shown in Fig.5.

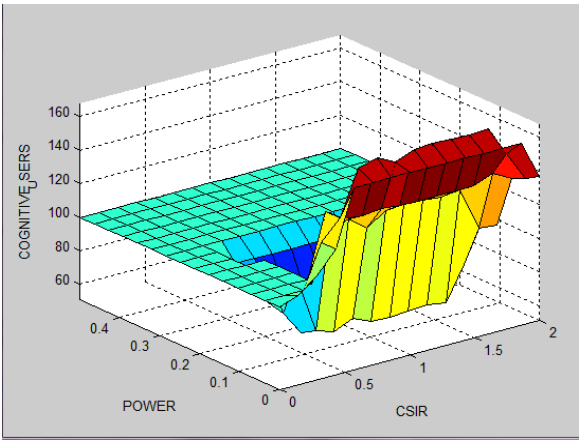


Fig. 5. Surface graph of FIS

Based upon the measured CSIR and the received signal power, the FIS endow with the optimum number of cognitive users to be allowed.

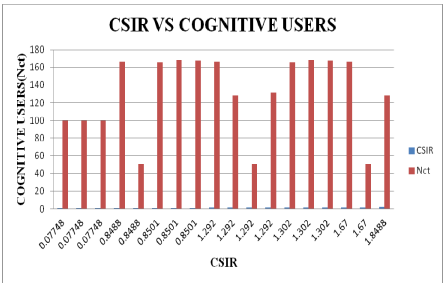


Fig. 6(a). CSIR vs Cognitive Users

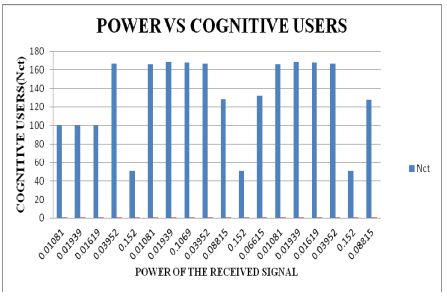


Fig. 6(b). Power vs Cognitive Users

Hence, the obtained variation of optimum number of cognitive users from the FIS system with respect to the input CSIR and received signal power is shown in the Fig6(a) and6(b) respectively.

6 Conclusion

In this paper, we consider a cognitive network consisting of single primary receiver with multiple primary and secondary transmitters. With the primary licensed transmitter and secondary cognitive users are involved in communication simultaneously. We have designed a simple and feedback interference reduction technique for the cognitive networks. In our proposed system, number of cognitive users can be change dynamically due to the interpretation of interference caused to primary user using fuzzy logic closed loop control system. The result of the proposed system shows by controlling the number of cognitive users in the network, interference, CSIR ratio and power is maintained or controlled to the desired level. We also suggest the future work of our proposed system to include the parameter for channel fading in the fuzzy logic controller.

References

1. Rabbachin, A., Quek, T.Q.S., Shin, H., Win, M.Z.: Cognitive network interference. Submitted to IEEE Journal on Selected Areas in Communications 29(2) (February 2011)
2. Jovicic, A., Viswanath, P.: Cognitive Radio: An Information-Theoretic Perspective. Submitted to the IEEE Transactions on Informatiom Theory (April 2006)
3. Cabric, D., Mishra, S.M., Brodersen, R.W.: Implementation issues in spectrum sensing for cognitive radios. In: Proc. IEEE Asilomar Conf. on Signals Systems and Computers, Pacific Grove, CA, pp. 772–776 (November 2004)
4. Ghasemi, A., Sousa, E.S.: Spectrum sensing in cognitive radio networks: Reqiurements, challenges and design trade-offs. IEEE Commun. Mag. 46(4), 32–39 (2008)
5. Ghasemi, A., Sousa, E.S.: Fundamental limits of spectrum-sharing in fading environments. IEEE Trans. Wireless Commun. 6(2) (February 2007)
6. Mishra, S.M., Ten Brink, S., Mahadevappa, R., Brodersen, R.W.: Cognitive technology for ultra-wideband/WiMax coexistence. In: Proc. IEEE DySPAN, Dublin, Ireland (2007)
7. Yiu, S.: Interference and Noise Reduction by Beamforming in Cognitive Networks. Submitted to IEEE Transactions on Communications 57(10) (October 2009)
8. Shared spectrum company, Comprehensive spectrum occupancy measurements over six different locations (August 2005)
9. Veen, B.D.V., Buckley, K.M.: Beamforming: a versatile approach to spatial filtering. IEEE ASSP Mag., 4–24 (1998)
10. Win, M.Z.: A mathematical model for network interference. In: IEEE Communication. Theory Workshop, Sedona, AZ (May 2007)

Managing Traffic Flow Based on Predictive Data Analysis

Dhara J. Patel[1], Snoeji Varghese John[2], and Fibinse Kaliangra[2]

[1] Institute of Technology, Nirma University
dharamtechec@gmail.com
[2] Motorola Mobility India Pvt. Ltd
{snoeji,fibinse.xavier}@gmail.com

Abstract. In this paper, we propose a solution to reduce traffic congestion by utilizing the existing traffic technologies and social data. The amount of social data will increase in the upcoming years as mobile devices and Internet begin serving a larger population. The availability of the driver's location data from mobile devices and feeds from micro-blogging sites play a key role in arriving at the solution. The concepts used to implement the solution are image processing algorithms, making efficient use of traffic camera feeds and social data. This solution will govern the existing traffic signal system and adapt it in real-time to streamline the traffic flow. The main challenge solved by this mechanism is to create smooth flowing traffic between two traffic signals while maintaining the fairness to other traffic conditions, thus improving the average traffic speed. This approach evaluates traffic variables obtained from the real time camera feed with sentiment analysis on social-data to create dynamic traffic timing. The end result of implementing the proposed solution will help in reducing travel time, vehicle idling time and reduce accident occurrences arising due to driver's mental fatigue.

Keywords: Social data, machine learning, image processing, traffic jams, Data Mining, Data Science.

1 Introduction

Traffic congestion arises due to a variety of limitations, such as:

1. Spiking growth in traffic in areas where the infrastructure does not grow to match the traffic requirement.
2. Increasing occurrences of accidents.
3. Static traffic control systems which are incapable of adapting to changing traffic requirements.

The gist of the problem would be - poor average speed of the vehicles. Improving this attribute could help achieve a better traffic flow between two points.

This paper shows that this can be achieved with the technologies already under use in developing countries, namely - traffic camera monitoring and growing use of low-cost mobile.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1069–1074.
springerlink.com

At any cross-section with say, two incoming and two outgoing roads, less than two traffic lights function simultaneously taking one functioning traffic signal visualized as traffic being fed from one source to multiple roads.

The proposed algorithm is explained below with the help of one traffic producing signal (T1) and one traffic receiving signal T2),:

Let's assume that the traffic signal at T1 is GREEN and the one at T2 is RED.

2 Proposed Solution

1. Collect the below mentioned traffic parameters:-

   a. Average speed (Av) of traffic leaving from T1 at green light,
   b. Vehicle density of the stationary vehicles at T2.

2. Predict the optimum time(using the below mentioned mail algorithm) to begin traffic flow at the downstream traffic signal T2, so as to prevent growth of stationary traffic.

3. Supplement information to the predictive algorithm in step- 2, with real-time geographical-location data available from mobile devices and feeds from micro-blogging sites.

4. Map location data from micro-blogging feeds which help in reducing the errors in prediction by providing exact location of the driver.

5. Highlight the level of traffic on the roads on a virtual map to notify incoming vehicles users of the real-time status of the roads.

The main goal of the above solution is to increase the clearing rate (CL) at each traffic signal.

The CL at each signal is directly proportional to a product of Number of times the signal can turn Green in a time-frame and the average speed of the traffic that it lets through.

$$CL = k * N * Asp$$

k = constant,

N= Number of times the signals go green,

Asp=Average speed of the traffic passing through. The ideal state would be to have the highest value of N and Avg.

Higher value of value of N is obtained by making the traffic signal dynamic and improving image processing of the video camera feed. The camera is placed at an angle at which it can view the both the stationary traffic and that is slowly joining the rear of stationary traffic.

2.1 Predictive Algorithm

1. Sample the camera video feed at more than 4 fps in burst mode.
2. Pass sampled images through the image process algorithm to calculate average speed (Avg) of traffic and vehicle density (VD), in both upper and lower half of the frames. The upper half of the feed characterizes the incoming traffic and the lower half is for the traffic closest to the signal.
3. Find average, of the values of Average speed and Vehicle density for each burst of images sampled from the camera feed.
4. The values obtained from the image processing algorithm are passed to the signal-decision algorithm (SDA) mentioned in point 6
5. Micro-blog feeds from mobile device application and from the general micro-feed are used in the following manner:

   5.1. They are mapped per user to find his/hers displacement upon receiving the next feed. This provides general direction of travel and displacement.

   5.2. Feeds are also clustered based on geo-location to find traffic density per road. c. The ratio of positive to negative sentiments around a particular signal to find the acceptance level of the crowd using the signal.
6. The Signal Decision Algorithm (block diagram in Fig-2) decides based on type of sentiment clustered around the particular signal if the signal should change from the present state.

   6.1. If sentiment-ratio is poor, check average speed (SocialAvg) from social data.

   6.2. If average speed (SocialAvg) is low:

   a. and it agrees with average speed (Avg) obtained from camera feed, increase time for which Green signal remains and then jump to step c.

   b. doesn't agree with average speed (Avg) obtained from camera feed, continue with present time-slot for Green signal

   c. Check if average speed (SocialAvg) has improved and is in accordance with that obtained from camera feed. If yes, wait for two Green signal cycles to make sure that positive sentiment has stabilized. If average speed (SocialAvg) has not improved, return to step 6.1

   6.3. If average speed (SocialAvg) is high/normal then Ignore the low sentiment-ratio and turn signal Green.

   If sentiment ratio is high, turn signal Green.Pass values down to the downstream signal T3 when upstream signal turns Green.

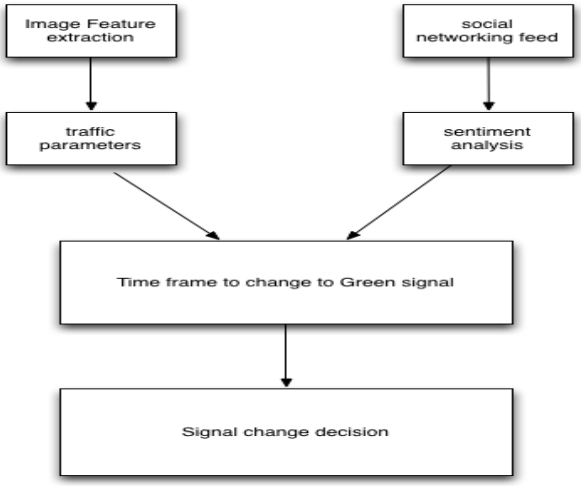


Fig. 2. Signal Decision algorithm block diagram

2.2 Algorithm to Analyze Micro-blog Feeds

1. Phone app sends micro-blog update on regular basis containing geo-location data to server. Data sent in this manner is from a limited set of words, which are segregated as positive and negative words[3].
2. Apply Maximum entropy and Naive Bayes to classify the data generated by non-phone app user with emoticons as noise labels, as described in classification by Distant Supervision [3].

Maximum Entropy:

In this formula, c is the class, d is the tweet, and λ is a weight vector. The weight vectors decide the significance of a feature in classification. A higher weight means that the feature is a strong indicator for the class

$$c* = argmac_c P_{NB}(c|d)$$

$$P_{NB}(c|d) := \frac{(P(c)\sum_{i=1}^{m} P(f|c)^{n_i(d)})}{P(d)}$$

The weight vector is found by numerical optimization of the lambdas so as to maximize the conditional probability.

Naive Bayes

Naive Bayes is a simple-to-use classification algorithm which finds immense use in text classification.

$$P_{ME}(c|d,\lambda) = \frac{\exp[\Sigma_i \lambda_i f_i(c,d)]}{\Sigma_{c'} \exp[\Sigma_i \lambda_i f_i(c,d)]}$$

In this formula, f represents a feature and ni(d) represents the count of feature fi found in tweet d. There are a total of m features.

Parameters P (c) and P (f |c) are obtained through maximum likelihood estimates, and add-1 smooth- ing is utilized for unseen features. [3]

From the extracted features from the image we can identify the speed and density of the traffic on the road.

To calculate traffic parameters, we evaluate the tracked trajectories of the cars. First, we calculate the velocity for each car. This enables us to decide which car is moving on which carriage way.

Thereby, we can count the number of cars on each road side. Knowing the length of the road segment which we observed, we can determine the traffic density

Traffic Density :

$$D = \frac{n}{l}$$

Traffic Speed:

$$\overline{v} = \frac{\sum_{i=0}^{n} v_i}{n} \text{ with } v_i \text{ being the speed of each car.}$$

n being the number of cars on the chosen carriage way, and l the length of the road segment.

The average speed per carriage way can easily be derived as another important parameter for traffic monitoring is traffic flow. This parameter describes the number of cars passing a fixed position in a certain time interval, which makes it hard to derive directly from aerial images.

A vague guess, however, is possible by multiplying the density and the average velocity. While aerial images show advantages to determine the above parameters, induction loops are better suited for calculating the traffic flow [4].

3 Conclusions

As demonstrated, using sentiment analysis of micro-blogging feeds, on top of the data collected from the traffic-image study, we can adapt traffic-signal-timings more effectively.

This approach will help us create a smoother vehicle flow than possible with conventional methods with least expenses incurred.

References

[1] Tubaishat, M., Shang, Y., Shi, H.: Adaptive Traffic Light Control with Wireless Sensor Networks. University of Missouri, Columbia
[2] Zeller, K., Hinz, S., Rosenbaum, D., Leitloff, J., Reinartz, P.: Traffic Monitoring Without Single Car Detection From Optical Airborne Images
[3] Go, A., Bhayani, R., Huang, L.: Twitter Sentiment Classification using Distant Supervision
[4] Hinz, S., Lenhart, D., Leitloff, J.: detection and tracking of vehicles in low framerate aerial image Sequences. Article

Digital Filter Approach for ECG in Signal Processing

Sonal K. Jagtap and M.D. Uplane

Department of Electronics, Shivaji University, Kolhapur, MS-416004
{sonalkjagtap,mduplane}@gmail.com

Abstract. Signal processing has a rich background and its importance is extended in fields as biomedical engineering, acoustics, sonar. Electrocardiogram (ECG) is used for the primary diagnosis of coronary heart diseases which shows electrophysiology of the heart and changes like arrhythmia as well as conduction defects. ECG in signal processing is one of the important research area in Biomedical signal processing. Recent advances in computer hardware and digital filter approach in signal processing have made it feasible to use ECG signals to communicate with a computer. So quality diagnosis of ECG is a technological challenge. This paper introduces comprehensive survey of digital filtering methods to cope with the noise artefacts in the ECG signal. The aim of this paper is to extract important features of ECG using signal processing techniques. In this paper, approaches of different digital filter for ECG in signal processing are discussed. Thus nowadays, importance of signal processing appears to be no visible sign of saturation.

Keywords: Signal Processing, ECG, Biomedical, Digital Filter.

1 Introduction

Processing of the ECG signal using digital filter involves initial sampling of the signal from electrodes on the body surface. Next, the digital ECG must eliminate or suppress low-frequency noise that results from baseline wander, movement, and respiration and higher-frequency noise that results from muscle artifact and power-line or radiated electromagnetic interference. As a result, the ECG signal at the body surface must be filtered and amplified by the electrocardiograph.

1.1 ECG in Signal Processing

Digital filters can be designed to have linear phase characteristics, and this avoids some of the distortion introduced by classic analog filters. Once filtered, individual templates are constructed for each lead from data sampled generally from dominant complexes, from which amplitude and duration measurements are made. Measurement error has an important effect on the accuracy of ECG diagnostic statements.

Many researchers have worked on the problems of Power Line Interference (PLI) and Baseline Wander in the diagnosis of ECG signal. Different methods are suggested for removing interferences. Cramer E, McManus C. D., Neubert D. Van Alste J.A.,

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1075–1082.
springerlink.com

Van Eck W., Herrmann O.E. have introduced global filtering of AC interference in the digitized ECG as a new concept. Cramer E, McManus C. D., Neubert D. recommended two digital filters. One is based on summation method and other uses a least squares method. Real ECGs and artificial signals used in analysis by applying each predictive filter and compared both the methods (1987). Challis R.E., Kitney R.I. have developed a digital filter using pole zero techniques. The Chebyshev and butterworth filters were developed. It is found that both types work satisfactory in the ECG signal. De Pinto V. evaluated two digital filters and found very effective in reducing signal contamination(1992). Gaydecki P. has established a simple but highly integrated digital signal processing system for real time filtering of biomedical filters (2000). Frankel R.A., Pottala E. W., Bowser R. W., J.J. described a digital filter for suppression of baseline wander. In the article authors suitable for preserving accuracy in the ST segment of the ECG signal (Oct 1991).Van Alste JA, van Eck W., Herrmann O.E. suggested the linear filtering method for baseline wander reduction (1986). Thus a great deal of research has focused on ECG as one of the biomedical signal. Despite the improvements that have been achieved in this area, filtering of ECG still poses some challenges. In this paper, the approach of different digital filters for preprocessing of ECG signals regarding their real-time applications has been reviewed.

2 Literature Survey

Real-time signal processing based on both general purpose microprocessors and fast digital signal processors (DSPs) is a technique that emerged over 20 years ago, and is now widely considered one of the fastest growing application areas in the field of digital technology.

2.1 Digital Filter Approach in Signal Processing

Digital filter Application includes biomedical signal analysis, image analysis, image coding and decoding techniques. [1-4]. Typically for filtering, the analog waveform is first digitized by an analogue to digital converter (ADC), and the binary values are transmitted to a DSP device that performs a real-time convolution operation in discrete space using either a finite impulse response (FIR) or infinite impulse response (IIR) algorithm. The processed data are then sent to a digital to analog converter (DAC) that outputs a filtered analogue signal. In order to meet the requirements of the Sampling Theorem with respect to the incoming waveform, and to eliminate quantization noise in the processed signal, an anti-aliasing filter is included before the (analog to digital converter) ADC and similarly, a reconstruction filter is included after the DAC.

Filters constructed using DSP technology offer many advantages over traditional analog methods. Most important, they are inherently flexible, since changing the characteristics of the filter merely involves changing the program code or filter coefficients; with an analogue filter, physical reconstruction is required. Furthermore, they are immune to the effects of ageing and environmental conditions, since the filtering process is dependent on numerical calculations, not mechanical characteristics of the

components. This makes them particularly suited for very low frequency signals. For the same reason, the performance of digital filters can be specified with extreme precision, in contrast to analog filters where a 5% figure is considered excellent. It would be a mistake, however, to assume that there is no scope remained in the field of digital filters.

2.2 Basic DSP Filter Theory

The (linear) process of filtering in time t is encapsulated in the convolution integral

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau \tag{1}$$

Where $y(t)$ is the output (filtered) signal, $x(t)$ is the incoming signal, t is the time-shift operator and $h(\tau)$ is the impulse response of the filter. In discrete space, this equation may be implemented using either an FIR or IIR solution. In the former case, the infinite response is truncated, which yields an expression of the form

$$y[n] = \sum_{k=0}^{M} h[k]x[n-k] \tag{2}$$

with the z-transform of the impulse response, i.e., the transfer function H(z), being given by

$$H(z) = \frac{Y(z)}{X(z)} = \sum_{n=0}^{\infty} h[n] \backslash z^{-n} \tag{3}$$

In contrast, IIR filters rely on recurrence formulae, where the output signal is given by

$$y[n] = \sum_{k=0}^{N} a[k]x[n-k] - \sum_{k=1}^{M} b[k]y[n-k] \tag{4}$$

and the transfer function is given by

$$H(z) = \frac{a[0] + a(1)z^{-1} + \cdots + a(m)z^{-m}}{1 + b(1)z^{-1} + \cdots + b(n)z^{-n}} = \frac{\sum_{m=0}^{M} a[m]z^{-m}}{1 + \sum_{n=1}^{N} b[n]z^{-n}} \tag{5}$$

There are important consequences and behaviors associated with these two approaches to digital filtering, which are summarized in Table 1. One of the most important criteria in assessing the performance of a filter is its stability. As equations (2) and (3) show, FIR filters are unconditionally stable since there is no recursion or feedback in the convolution process. In contrast, IIR filters always feedback a fraction of the output signal (see second term in eq. (4)), which necessitates careful attention to design if stability is to be ensured. This may be viewed another way: Eq. (5) shows that the transfer function is the ratio of two polynomials in ascending negative powers of z. Thus high-order polynomials are associated with very small denominator terms and hence the risk of an ill-conditioned division. It is for this reason that IIR filters are sensitive to the word-length of the DSP device. In general, the higher the order of the filter, the greater the risk of instability, so high-order IIR filters are often designed by cascading together several low-order sections.

Table 1. Common Properties of FIR and IIR filters

Property	FIR	IIR
Stability	Yes	No
Immunity to DSP word-length	Good	Poor
Linear phase	Yes	No
Arbitrary frequency response	Yes	No
Design ease	Straightforward	Labour intensive
Computational load	High	Low
Direct analog equivalent	No	Yes

With high-performance audio system filters, linear-phase is desirable; in the processing of biomedical signals, this property is essential. Linear phase means that any time delay experienced by one frequency component is experienced by them all in equal measure; hence the shape of the filtered signal is preserved. Linear phase is guaranteed if the impulse response of the filter is symmetrical, i.e. it obeys the relationship given by

$$h(n) = h(N - n - 1), n = 0,1, (N - 1)/2 \qquad (N\ odd) \tag{6}$$

With IIR filters, it is impossible to achieve pure linear phase, especially in the transition bands. A number of other properties also make FIR filters the desirable choice in many applications; for example, they can be made to have arbitrary frequency responses by specifying this in the Fourier domain and taking the inverse transform to obtain the impulse response. This is known as the Frequency Sampling Method, and due to its simplicity, is very widely used. Although it is theoretically possible to generate arbitrary IIR filters, in practice the computational burden in calculating the filter coefficients makes this totally impractical. IIR filters are commonly designed by calculating the poles and zeros for a particular filter, and accepting those which lie within the unit circle of the z-plane. This can be a complicated procedure, so equations have been established to obtain the poles and zeros of commonly used filters, e.g., Butterworth and Chebyshev types.

From this critique it might appear that FIR filters are overwhelmingly superior to IIR filters. In fact, both are widely used. The principal advantage that the IIR type has, is computational efficiency. In order to realize a filter with a sharp cut-off, the IIR uses many fewer terms than the FIR. Hence for a processor with a given power, IIR filters are more effective and use less memory resource. Moreover, analogue filters can be readily transformed into equivalent IIR digital filters, with similar performances. This is in general not true for FIR filters. The system described below makes use of FIR techniques although it can be adapted, with no changes in the hardware, to run both FIR and IIR types.

2.3 Performance Measures

Performance measures are important factors in the analysis of ECG signal. Generally termed as Signal to Noise ratio and Mean Square Error which are explained as below.

2.3.1 Signal to Noise Ratio (SNR)

The SNR value is to determine corresponding function for filtering the ECG signal. The output SNR is given by eq. (7)

$$SNR = 10\log_{10}\frac{\sum_{n=0}^{N} x[n]^2}{\sum_{n=0}^{N}(x_{dn}[n] - x[n])^2} \tag{7}$$

2.3.2 Mean Square Error (MSE)

The MSE value is estimated between the original ECG signal and filtered ECG signal is given by eq. (8)

$$MSE = \frac{1}{N}\sum_{n=0}^{N}(x_{dn}[n] - x[n])^2 \tag{8}$$

3 Methods

3.1 Low Frequency Filtering

Traditional analog filtering, a 0.5-Hz low-frequency cutoff introduces considerable distortion into the ECG, particularly with respect to the level of the ST segment. This distortion results from phase nonlinearities that occur in areas of the ECG signal where frequency content and wave amplitude change abruptly, as occurs where the end of the QRS complex meets the ST segment. Digital filtering provides methods for increasing the low-frequency cutoff without the introduction of phase distortion.

This can be accomplished with a bidirectional filter by a second filtering pass that is applied in reverse time, i.e. from the end of the T wave to the onset of the P wave. This approach can be applied to ECG signals that are stored in computer memory, but it is not possible to achieve continuous real-time monitoring without a time lag. Alternatively, a zero phase shift can be achieved with a flat step response filter, which allows the reduction of baseline drift without low frequency distortion.

3.2 High Frequency Filtering

The digital sampling rate (samples per second) determines the upper limit of the signal frequency that can be faithfully represented. According to the Nyquist theorem, digital sampling must be performed at twice the rate of the desired high-frequency cutoff. Because this theorem is valid only for an infinite sampling interval, the 1990 AHA report recommended sampling rates at 2 or 3 times the theoretical minimum. A series of studies have now indicated that data at 500 samples per second are needed to allow the 150-Hz high-frequency digital filter cutoff that is required to reduce amplitude error measurements to1% in adults. Greater bandwidth may be required for accurate determination of amplitudes in infants.

3.3 Signal Pre-processing

Filtering aims at simplifying subsequent processing operations without losing relevant information. An important goal of preprocessing is to improve signal quality by

improving the so-called Signal-to-Noise Ratio (SNR). A bad or small SNR means that the interferences are occurring in the original signal, which makes relevant information hard to detect. In general, a good or large SNR, simplifies the detection and classification task.

When measuring the ECG, all of the signals do not come from the electrical activity of the heart. Many potential changes seen in the ECG may be from other sources. These changes in the ECG that are of non-cerebral origin are called artefacts and their sources may be the equipment or the subject. The surrounding electrical equipment may induce 50-Hz or 60-Hz component in the signal. It is removed by filtering by notch filter.

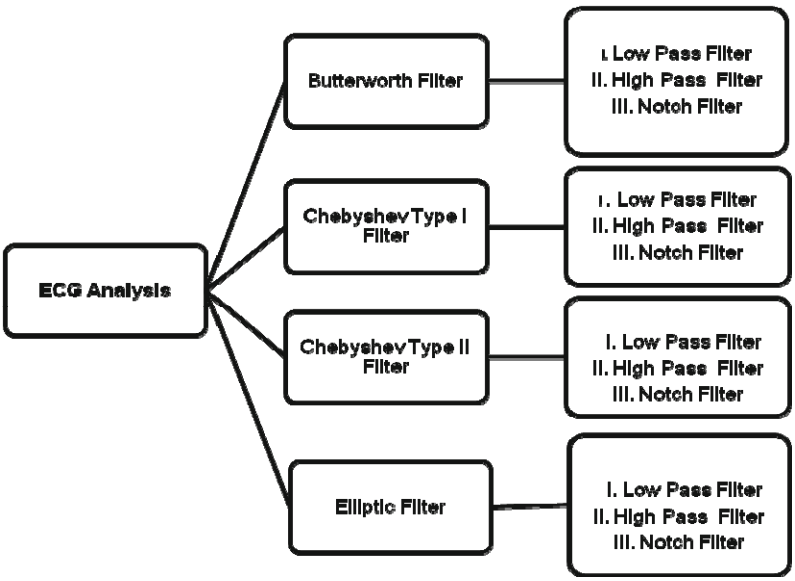


Fig. 1. Signal Processing Techniques for ECG Analysis

4 Simulink Model Used in the Present Work

The model used in the present work for the filtering of the ECG signal using IIR filters as shown in the fig. . In the model digital inputs indicates the data from the patients or from the stored digital sample file. All the filters are designed in simulink and cascaded to get the results.

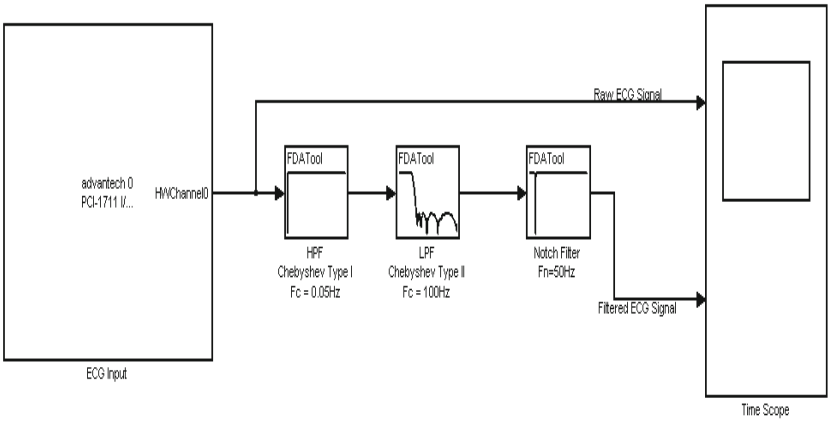


Fig. 2. Final Simulink Model Used in the ECG Filtration System

All the filters are designed in simulink and cascaded to get the results. Final the filters are cascaded to achieve the combined effect of these filters. Filters are cascaded and final model is built in SIMULINK to remove the noise from the ECG. Based on performances, final model includes Chebyshev Type I low pass filter, Chebyshev type II high pass filter and a notch filter centered at 50Hz. Result of this final model is shown in the Fig. 9.

5 Results and Discussion

ECG filtration using Low pass, High Pass and Notch filters were carried out separately in the work. Finally the filters are cascaded to achieve the combined effect of these filters. Fig. 3 shows the raw ECG signal before filtration. Also the fig. 4 shows ECG signal after application of the cascaded filters.

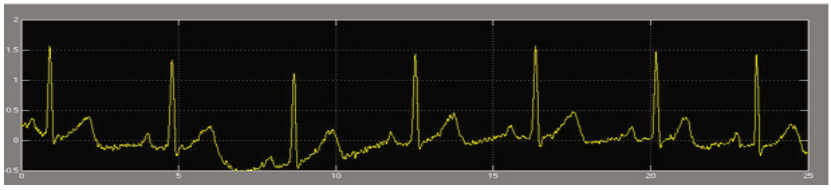


Fig. 3. Raw ECG Signal Before Filtration

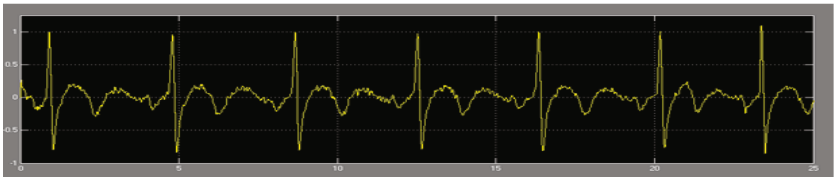


Fig. 4. ECG Signal After Filtration

6 Conclusion

- Filtering is an important step in the processing of the surface ECG signal. This study investigated some of the different factors influencing digital filtering approach in filtration of the ECG and compared the performances of all IIR filters.
- Present work shows that the performance of cascaded filters is better for filtration of ECG signal. At low coverage, the IIR filters removed noise better than any other filters.
- QRS complex is modified and baseline wander is reduced at a greater extent.
- This work leads to optimum solution without losing relevant information in ECG signal.

References

1. Cramer, E., McManus, C.D., Neubert, D.: Estimation and removal of power line interference in the electrocardiogram: a comparison of digital approaches. Comput. Biomed. Res. 20(1), 12–28 (1987)
2. Challis, R.E., Kitney, R.I.: The design of digital filters for biomedical signal processing, Part 3: The design of Butterworth and Chebyshev filters. J. Biomed. Eng. 5(2), 91–102 (1983)
3. De Pinto, V.: Filters for the reduction of baseline wander and muscle artifact in the ECG. J. Electrocardiol. 25 (suppl.), 40–48 (1992)
4. Frankel, R.A., Pottala, E.W., Bowser, R.W.J.J.: A filter to suppress ECG baseline wander and preserve ST-segment accuracy in a real time environment. J. Electrocardiol. 24(4), 315–323 (1991)
5. Gaydecki, P.: A real time programmable digital filter for biomedical signal enhancement incorporating a high-level design interface. Physiol. Meas. 21(1), 187–196 (2000)
6. Ifeachor, E.C., Jervis, B.W.: Digital signal processing: a practical approach. Addison-Wesley, Wokingham (1993)
7. Carr, J.J., Brown, J.M.: Introduction to Biomedical Equipment Technology, 4th edn. Pearson Education (2007) ISBN 81-7758-883-4
8. Webster, J.G.: Medical Instrumentation, 3rd edn. John Wiley & Sons Inc. (2005) ISBN 9971-51-270-X
9. Kligfield, P., Gettes, L., Bailey, J., et al.: Recommendations for the standardization and Interpretation of the Electrocardiogram. J. Am Coll. Cardiol. 49, 276–281 (2007)
10. Sornmo, L.: Time-varying digital filtering of ECG baseline wander. Med. Biol. Eng. Comput. 31(5), 503–508 (1993)
11. Van Alste, J.A., Van Eck, W., Herrmann, O.E.: ECG baseline wander reduction using linear phase filters. Comput. Biomed. Res. 19(5), 417–427 (1986)

A Pattern Recognition Approach of Japanese Text Recognition for Template Matching

Soumendu Das and Sreeparna Banerjee

West Bengal University of Technology, India
sdphotoes@gmail.com, sreeparnab@hotmail.com

Abstract. Handwritten character recognition is a difficult job in the field of document retrieval and analysis. However the steps included in this procedure could be error prone because the Japanese language has over 3000 characters which can be classified as syllabic characters, or Kana, and ideographic characters, called Kanji. In addition there is no concept of delimiters like space, used to separate and placed between two different words. In organizations also, handwritten signatures and their identification play an important role in the term of security. Time consuming and cost effective computer generated signature identification is very complex process when it comes to Japanese handwritten text recognition, due to presence of similarly shaped, homomorphic characters in Japanese language. This note surveys some earlier attempts and presents a template to character recognition.

Keywords: Japanese character recognition, Handwritten character recognition, Template matching for pattern recognition.

1 Introduction

Nowadays, retrieving desired documents from the huge document databases plays a major task for offices. Optical character readers come in to play to convert proper documents to electronic documents, but these are error prone. Hence there exists a necessity of efficiently combining optical character recognition along with document retrieval techniques. Document retrieval technique in Japanese is further complicated by the fact that the text comprises of both the syllabic / phonetic characters (Kana) as well as the ideographic characters (Kanji) and similar shape definition of several Japanese characters. Furthermore, Japanese text is not separated by delimiters such as spaces. Homomorphism or similar shape definition for different Japanese characters also poses problems especially in sans serif fonts. The next section discusses the Japanese language model briefly. Section 3 gives a brief survey of earlier approaches. Our proposed algorithm is outlined in section 4 and the analysis and the future works is included in section 5.

2 Japanese Languages Model

Japanese text consists of more than 3000 characters and many among them are complex and similar in shapes, and the text is not separated by delimiters such as spaces. Japanese writing system has three different characters sets, namely, Hiragana, Katakana and Kanji.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1083–1089.
springerlink.com

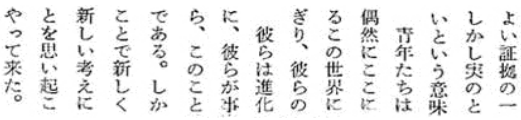


Fig. 1. Sample Japanese Text

For Japanese words, Hiragana (see Fig2a) is used mostly for grammatical morphemes. Katakana (see Fig2b) is used for transcribing foreign words, mostly western, borrowing and non-standard areas. In addition, diacritic signs like dakuten and handakuten are used (see Fig3 and 4).

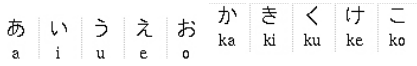


Fig. 2a. Hiragana Script

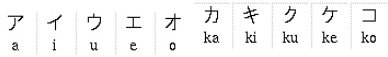


Fig. 2b. katakana script

Dakuten are used for syllables with a voiced consonant phoneme. The dakuten glyph (゛) resembles a quotation mark and is directly attached to a character (Foljanty 1984).

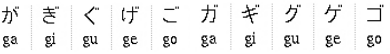


Fig. 3. Hiragana and Katakana Dakuten Alphabets

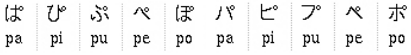


Fig. 4 Hiragana and Katakana Handakuten Alphabets

Handakuten are used for syllables with a homomorphism. The glyph for a 'maru' is a little circle (°) that is directly attached to a character (Foljanty 1984). Kanji are content bearing morphemes. In Japanese text Kanji are written according to building principles like Pictograms (graphically simplified images of real artifacts), ideograms (combinations of two or more pictographically characters) and phonograms (combinations of two Kanji characters).

3 Earlier Attempts

3.1 Document Retrieval Strategies

Character recognition can be done using a spelling checker which is capable of integrating characteristic patterns of recognition errors which differ from normal typing errors. There are also approaches like [1] including linguistic knowledge about the content of documents [2] (in addition to syntactic and lexical knowledge) and the category utilizes vocabulary derived, in order to improve the word recognition rate [3]. The document processing system "Transmedia Machine" is used for document search with out optical character recognition [4]. Character images of scanned documents are encoded into two binary features for each character succeeded by a "string matching"

based on incomplete codes. Word level encoding [5] has been proposed as a more reliable alternative. Searching for text passages in document image database and subsequent pattern matching using a number of feature descriptors has also pattern been proposed [6]. In order to make the search system tolerant of recognition error, multiple candidates have been used in the search process [7]. Optical character recognition keeps multiple candidates for ambiguous recognition and outputs them as a result text. Segmentation ambiguities [8] can also be included, with multiple hypotheses in both character segmentation and recognition represented as a network of hypotheses. Itoh [10] proposed a method to overcome the problem of huge character set with different entropies in Japanese language and n-gram based character recognition by using a clustering scheme based on different parts of speech of Kanji and also by homogenizing the entropies of different Kana and Kanji characters. Miyao and Maruyama [11] attempt to overcome this difficulty by synthesizing virtual examples from a small number of real samples of huge character set. Kim [12] proposed an offline effective algorithm to overcome the problem for large scale character recognition for large set characters like Korean and Chinese to overcome problems like absorbing variations of the same characters among different writing styles using the algorithm of template matching and improvement strategies.

3.2 Recognition by Feature Vector of the Character

Barners and Manic [16] proposed an algorithm of producing feature vectors of the pictorial text. The idea is to identify the character by its pattern and features uniquely, introducing with designing a neural network and not by training. Applying this on dakuon and handakuon characters, they showed that the center of gravity doesn't change even the character is rotated. The center of gravity moves proportionally with the additional pixels and produces a set of unique feature characteristics.

The Size-Translation-Rotation-Invariant Character Recognition and Feature vector Based STRICR-FB algorithm is based on the Kohonen Winner Take All [13, 14], type of unsupervised learning. The algorithm comprises of two phases; Construction of Character Unique Feature Vectors which calculates distance between characters and in an expanded form of the Euclidean distance defined in [15]. The post-processing phase is passing the character unique feature vectors through a neural network for character recognition. A test set of random characters is then used to determine the effectiveness of this artificial neural network. The experiment by random characters produces three sets of results; among which rotation set produced 96.2% accuracy rate and that of random character set is 93%.

4 Proposed Process Descriptions

The experiment below is showing how target and template pictorial Japanese text is being matched with one another. The template image is collected from the pool of template set of trained pictorial images. The steps that we have followed are as following:

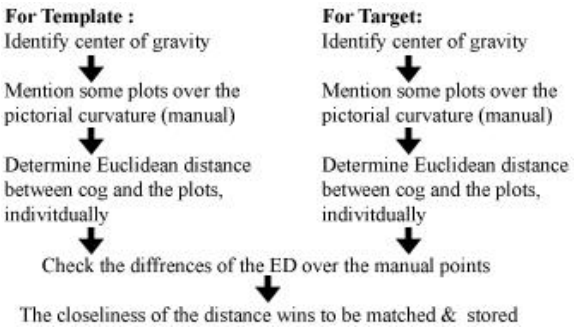


Fig. 5. The processes applied on both target and template pictorial Japanese text and measured the closeness, separately. Also the target pictorial text is rotated and then processed.

A. Processing steps for template image

//Find the COG of the template image (i_{cog}, j_{cog})

$$i_{cog} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} i.C_{ij} \right) / \left(\sum_{i=1}^{m} \sum_{j=1}^{n} C_{ij} \right) \quad (1)$$

$$j_{cog} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} j.C_{ij} \right) / \left(\sum_{i=1}^{m} \sum_{j=1}^{n} \right) \quad (2)$$

//Plot some of the feature points (pixels over template image)

a1(x1, y1), a2(x2, y2), a3(x3, y3). (3)

Find Euclidian distance

$$ed1_{ij} = \sqrt{[(x1 - i_{cog})^2 + (y1 - j_{cog})^2]} \text{ //max Euclidean dist.} \quad (4)$$

$$ed2_{ij} = \sqrt{[(x2 - i_{cog})^2 + (y2 - j_{cog})^2]} \text{ //with cog and point 'a'} \quad (5)$$

$$ed3_{ij} = \sqrt{[(x3 - i_{cog})^2 + (y3 - j_{cog})^2]} \text{ //with cog and point 'b'} \quad (6)$$

B. Processing steps for target working image

The steps (4), (5) and (6) are repeated same for the target image to be checked with the defined template image based on the distance between the center of gravity and that with each of the visually assigned feature points on intersecting areas.

//Find the COG of the target image (i_{cog}, j_{cog}) (7)

//Finding the same spots in target image t1(tx1, ty1), t2(tx2, ty2), t3 (tx3, ty3),

$$ed\_t1_{ij} = \sqrt{[(tx1 - i_{cog})^2 + (ty1 - j_{cog})^2]} \text{ //max Euclidean dist.} \quad (8)$$

$$ed\_t2_{ij} = \sqrt{[(tx2 - i_{cog})^2 + (ty2 - j_{cog})^2]} \text{ //with point 'a'.} \quad (9)$$

$$ed\_t3_{ij} = \sqrt{[(tx3 - i_{cog})^2 + (ty3 - j_{cog})^2]} \text{ //with point 'b'.} \quad (10)$$

//Check the differences between $(ed1_{ij}, ed2_{ij}, ed3_{ij})$ and $(ed\_t1_{ij}, ed\_t2_{ij}, ed\_t3_{ij})$

//Comparing the value of (4), (5) and (6) with (8), (9) and (10) respectively, we get the differences and hence we can reveal the result based on this features.

Fig. 6. The distance between COG and some of the pixel in target image (left) and that of the template image (right) for the Japanese character "あ" are shown

Table 1. The COG and the feature point analysis of template and target image of character "あ"

Char : あ	COG	Ed1	Ed2	Ed3
Template	(68, 76)	57.3847	31.7805	34.9285
Target	(61, 72)	66.7308	24.0416	65.3911

Table 2. The COG and the feature point analysis of template and target image (after rotation) of character "あ"

Char : あ	COG	Ed1	Ed2	Ed3
Template	(68, 76)	57.3847	31.7805	34.9285
	(50, 54)	44.6430	13.6015	26.9258
Target	(86, 91)	67.4759	22.4722	58.0086

All are having same computer generated text features with the same font style and size. And the target image is containing a handwritten Japanese text, after scanning and then digitized for preprocessing ahead.

Fig. 7. Pixel by pixel matching visualization of the character "あ", before and after rotation of the target hand written Japanese pictorial text. Fig (right) shows the template and target image matching interface.

5 Analysis and Future Work

The following are the interfaces of the application prepared for recognizing Japanese text, shown in two consecutive steps of 1) segmenting, recognizing and identifying and 2) template matching, as follows. The following are the interfaces of the application, have been simulated using MATLAB.

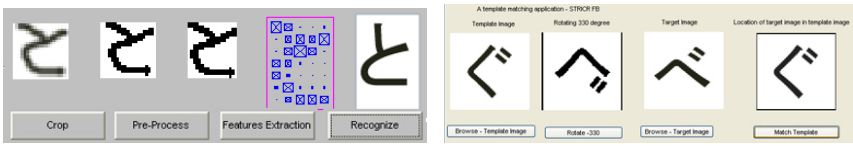


Fig. 8. The Interface for recognizing Japanese hiragana "to" (と). The figure (right) shows the template and target image (after rotation) matching interface. Character ぐ(gu) and べ(be) gets matched after being rotated.

Many characters in Japanese text look very similar to some other characters. At the time of character recognition many of them get matched and provide wrong results by template and target pictorial image matching. Characters like は(ha), ぐ(gu), じ(ji) and が(ga) etc. get matched with the characters respectively with な(na), べ(be), づ(zu) after rotations. Also there is a presence of noise which can be removed by using Gabor filter to reduce the tendency of getting wrongly matched characters including disruptions from documents containing high contrast or illumination, rough surface, unwanted spots etc.

6 Conclusions

Japanese text based documents retrieval is difficult to pursue due to huge character set of the text and no spaces between two words. Japanese text comprises of over 3000 characters based on syllabic characters (Kana) and ideographic character (Kanji). This paper surveys earlier approaches to character recognition and presents a template based recognition approach.

References

1. Dahl, D.A., Norton, L.M., Taylor, S.L.: Improving OCR accuracy with linguistic knowledge. In: Proc. Second Ann. Symp. Document Analysis and Information Retrieval, pp. 169–177 (1993)
2. Niwa, N., Kayashima, K., Shimeki, Y.: Postprocessing for character recognition using keyword information. In: IAPR Workshop Machine Vision Applications, pp. 519–522 (1992)
3. Hull, J.J., Li, Y.: Word recognition result interpretation using the vector space model for information retrieval. In: Proc. Second Ann. Syrup. Document Analysis and Information Retrieval, pp. 147–155 (1993)
4. Tanaka, Y., Torii, H.: Transmedia machine and its keyword search over image texts. In: Proc. RlAO 1988, pp. 248–258 (1988)
5. Trenkle, J.M., Vogt, R.C.: Word recognition forinformation retrieval in the image domain. In: Proc. Second Ann. Symp. Document Analysis and Information Retrieval, pp. 105–122 (1993)
6. Hull, J.J.: Document image matching and retrieval with multiple distortion–invariant descriptors. In: Proc. IAPR Workshop on Document Analysis Systems, pp. 383–399 (1994)
7. Fujisawa, H., Hatakeyama, A., Nakano, Y., Higashino, J., Hananoi, T.: Document storage and retrieval system. U.S. Patent 4985863 (1986)
8. Senda, S., Minoh, M., lkeda, K.: Document image retrieval system using character candidates generated by character recognition process. In: Proc. Second Int. Conf. Document Analysis and Recognition, pp. 541–546 (1993)
9. Marukawa, K., Hu, T., Fujisawa, H., Shima, Y.: Document retrieval tolerating character recognition errors—evaluation and application. Pattern Recognition 30(8), 1361–1371 (1997); Oriental Character Recognition
10. Itoh, N.: Japanese language model based on bigrams and its application to on-line character recognition. PR 28(2), 135–141 (1995)

11. Maruyama, K.-I., Maruyama, M., Miyao, H., Nakano, Y.: Handprinted Hiragana recognition using upport vector machines. In: Proceedings. Eighth International Workshop on Frontiers in Handwriting Recognition, pp. 55–60 (2002), doi:10.1109/IWFHR.2002.1030884
12. Kim, S.H.: Performance Improvement Strategies on Template Matching for Large Set Character Recognition. In: Proc. 17th International Conference on Computer Processing of Oriental Languages, Hong Kong, pp. 250–253 (April 1997)
13. Kohonen, T.: Self-organized formation of topologically correct feature maps. Biological Cybernetics 43, 59–69 (1982)
14. Kohonen, T.: Self-Organization and Associative Memory, 2nd edn. Springer, New York (1988)
15. Hung, D., Cheng, H., Sengkhamyong, S.: Design of a Hardware Accelerator for Real-Time Moment Computation: A Wavefront Array Approach. IEEE Transactions on Industrial Electronics 46(1) (February 1999)
16. Barnes, D., Manic, M.: STRICR-FB, a Novel Size-Translation- Rotation-Invariant Character Recognition Method. In: 2010 3rd Conference on Human System Interactions (HSI), pp. 163–168. Univ. of Idaho, Moscow (2010)

Extraction of Bacterial Clusters from Digital Microscopic Images through Statistical and Neural Network Approaches

Chayadevi M.L.[1] and Raju G.T.[2]

[1] JSSATE, Bangalore-60
chayadevi1999@gmail.com
[2] RNSIT, Bangalore -98
gtraju19990@yahoo.com

Abstract. The field of bioinformatics shows a tremendous growth at the crossroads of biology, medicine, and informatics. Applying data mining techniques in the medical field is a very challenging undertaking due to the idiosyncrasies of the medical profession. Using conventional methods, it is very difficult to determine the exact number of microorganisms in a microscopic picture. In this paper, the emphasis is on the automatic detection of microbes using automated tools and extraction of bacterial clusters through statistical and neural network approaches. Also, Multiscan approaches with freeman chain code and contour detection for the bacterial patterns in the images have been presented. Experimental results shows that the bacterial cluster patterns obtained through neural network approach are better than the statistical approach.

1 Introduction

Biomedical informatics is ultimately aimed at organizing, storing and processing information on molecular and cellular processes, tissues and organs, individuals, population and society to support the definition of suitable decision making strategies in health care.[7]. The detection and analysis of patterns in microscopic images needs to be automated since it is very difficult to determine the exact number of microorganisms in microscopic images manually using conventional microscopic methods. In manual method, microorganisms are counted under microscope by a medical expert that is time consuming. Bacteria can only be seen under microscope, they are unicellular and colourless. Stains have been applied on the cultured bacteria to identify them under microscope. The size of bacteria is in *millionth p*art of a meter. The usual bacterial detection and identification methods include analysis of morphological, physiological, biochemical and genetic data. In this paper, we propose *freeman chain code* or contour technique for bacterial shape detection/identification and the statistical and neural network approach for clustering bacterial patterns.

2 Related Work

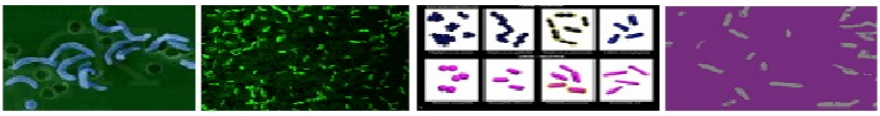
Bioinformatics is an evergreen, ever challenging field. Extracting the molecular data of a microorganism is a challenging work. Microorganism detection and counting

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1091–1099.
springerlink.com

microbes is the major role in clinical pathology before any treatment. Overview of current practices, challenges, tools and technologies in data mining with bioinformatics are provided in [3]. Shillabeer and Roddick's work[9] discussed several inherent conflicts between the traditional methodologies of data mining approaches in medicine. Khalid Raza [4] discussed the Application of Data mining in Bioinformatics and some of the important areas of Bioinformatics such as Sequence analysis, Genome annotation etc., Yuanyuan Shen et al.[14], Woolf P. J., Wang Y.[15] discussed some of the bioinformatics tools such as blast ,cs-blast. Arati kadav et al., Youfang Cao[13], Nakul Soni[5] et al., addressed some of the software used in field of health care sectors such as Public health and Biosurveillance etc., Massana et al. discussed the pre-processing edge detector for bacterial image under epifluorescence conditions and concluded that Marr-Hildreth operator functions with a high degree of independence for exposure and lighting characteristics. P.S. Hiremath et al.[6] defines the classification of bacterial cells in digital microscopic images with k-NN, neural network and fuzzy classifier techniques to classify Bacilli bacteria. Riries Rulaningtya et al.[8] addressed the automatic classification of tuberculosis bacteria with neural networks. S. Prabakar et al.[10] discussed the development of image processing scheme for bacterial classification based on optimal discriminate feature. In this paper we propose Self Organizing Feature Maps(SOM) and K-Means clustering algorithms for grouping the similar bacteria from large number of bacterial samples over a. large repository of bacterial images.

3 Bacterial Image Database

Digital images of bacteria were collected across the hospitals. These digital images of the bacteria are captured through a CCD camera which is mounted on top of micro-scope. Fig.1 shows the sample images of the different bacteria.



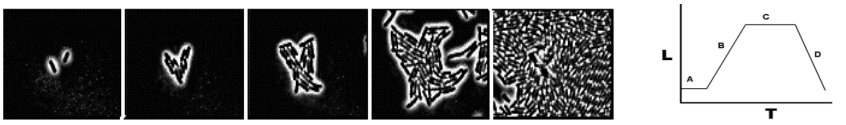
(a) (b) (c) (d)

Fig. 1. Images of the different bacteria (a)V_cholerae (b) Vibrio(c) Gram stain bacteria (d) Bacilli

Bacteria are basically colourless. In the Fig. 1, background colour is represented by stain colour and the usual shapes of bacteria are: *spiral, rod shaped, and circular*. *Rod shaped* bacteria are called *Bacillli*, *Round shaped* are called *Cocci* and the *kidney shaped* curved bacteria are *Vibrio*. Microbes are 0.1-0.5 micron in breadth and 1-4 micron in length. They are identified by their shape. Bacteria develop in 4 phases: *Lag phase, Exponential or Log phase, Stationary phase, and the Death phase*.

During the *Exponential phase*, bacteria undergo fission and doubles. The total number of bacteria per unit time is directly proportional to the population of bacteria. It doubles and increases in each consecutive period if the growth is not limited.

This is clearly shown in the Fig. 2. The slope of the line in Fig. 2 indicates the specific growth rate of the microbes. *L* represents the log numbers.



(a) (b)

Fig. 2. (a) The bacterial growth during exponential phase(b) Plot of growth (L) versus time (*T*)

In conventional methods, 50 to 100 *ml* sample of sputum, blood or urine is placed on glass slide and stained with stains like: gram stain (gram positive or gram negative) for ordinary bacteria, Z.N stain for TB bacteria, Leishman Stain for Blood samples etc., for identification of bacteria. Conventional methods (microscopic reading) employed for microorganism detection is time-consuming, very tedious, subject to poor specificity and requires highly trained personnel.

4 Methodologies for Identification of Bacteria

Bacterial image database has to be pre-processed for eliminating blur, irregular and noisy images. In this work, *binarization* and *thresholding* methods are used for removal of noise. Once the pre-processing is done, the features of the bacteria are extracted that includes a feature set of 81 features. Some of the main features are perimeter, circularity, major axis, minor axis, eccentricity and tortuosity. Out of these features, the potential features that help in identifying/recognizing the bacteria are selected. Finally, by applying image segmentation techniques, the recognition and count of microorganisms are detected.

In our research work, we proposed the *freeman chain contour algorithm* for recognizing and counting of individual type of bacteria [2]. The algorithm is summarized below.

Step1. Scan the segmented image one by one from left to right and top to bottom. When the scan line hits the microorganism (presence of the microorganism is represented by black pixel), the pixel values of the microorganism are extracted and transferred into to an array.

Step2. With pixel position (I, J), the continuity of the microorganism is checked in the neighborhood of the pixel and the directions of each pixel is stored.

Step3. At the end of the contour, if a pixel reaches the initial pixel point of that microorganism with the 8-neighbor connectivity then it is counted as one microorganism. Then microorganism count is incremented.

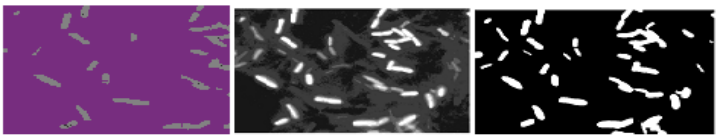
Step4. Contour traversal of the microorganism is obtained along with the directions of traversal (8-connectivity indicates 8 directions).

Step5. The directions of each micro-organism is stored and compared with shape of the microorganism.

Step6. After identifying the microbes, they are set to the background color or to some number so that same microbe is not encountered in the next scan.

Step7. Repeat the entire process till the entire image is scanned.

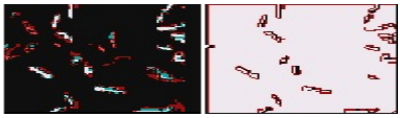
Fig. 3 shows the process of obtaining binary image of the original image during pre-processing.



(a) (b) (c)

Fig. 3. Binary images (a) original image (b) grey level image (c) binary image

In the proposed algorithm, pixel discontinuities and change in image intensities are used for edge detection of bacteria. This algorithm provided a single pixel width of the edge of bacteria. Fig. 4 shows the result of edge detection of bacteria.



(a) (b)

Fig. 4. (a) Original image in grey threshold (b) the edges of single pixel width

The edge of the bacteria which is of primary importance in shape detection is traced using contour traversal. Chain code for square using 4-connectivity is 3-0-1-2, whereas for 8-connectivity, it is 6-6-0-0-2-2-4-4. In this research work, we use 8-connectivity, to have more accuracy and efficiency.

5 Clustering Algorithms

Clustering pertains to unsupervised learning, where the data with class labels are not available. It basically involves enumerating C partitions, optimizing some criterion, over t-iterations, so as to minimize the inter-cluster *distance* (*dissimilarity*) or maximize the intra-cluster *resemblance* (*similarity*). Majority of the techniques that have been used for pattern discovery from bacteria are clustering and classification methods. In medical applications, clustering methods can be used for the purpose of cardiac image segmentation and to track the volume of left ventricle during complete cardiac cycle. It is also used to classify the patients as normal, hyperthyroid, hypothyroid using unsupervised clustering methods etc. A clustering algorithm takes as input a set of input vectors and gives as output a set of clusters thus mapping of each input

vector to a cluster. Clusters can be labeled to indicate a particular semantic meaning pertaining to all input vectors mapped to that cluster. Statistical technique *K-means* and Neural network based clustering technique *SOM-Kmeans* is presented in this paper. Both *K-Means* and *SOM-Kmeans* clustering algorithms clusters *N* data points into *k* disjoint subsets S_j. The geometric centroid of the data points represents the prototype vector for each subset. SOM is a close cousin of K-Means that embeds the clusters in a low dimensional space right from the beginning and proceeds in a way that places related clusters close together in that space. Experimental results are provided to show performance in terms of *intra-cluster* and *inter-cluster* distances for both *K-Means* and *SOM-Kmeans* clustering algorithms.

6 Experimental Results and Discussions

Experiments have been conducted on the bacterial image database of 320 images collected from reputed hospitals. Using the proposed algorithms, microorganism detection and counting is carried out not only for circular bacteria (circular bacteria is called as cocci), but also for bacilli, corny bacteria, vibriyo etc. Overall count of the microorganism is the sum of the microorganisms present in each category (cocci, bacilli, corny bacteria). The results are compared with manual count taken by the doctors. The results obtained by proposed method are more accurate compared to human visual counting or conventional methods. Fig. 5 shows the count of bacteria in sample images. Table 1 presents the comparison between conventional and proposed methods for five sample images of Fig.6. Fig. 7 shows the different shaped bacteria. Table 2 shows the sample feature values extracted for clustering purpose.

Table 1. Comparison between Conventional and Proposed methods

Image Sample	Conventional Method	Proposed Method	% error w.r.t Conventional Method
1	18	34	47
2	30	55	41
3	14	21	33
4	22	32	30
5	20	30	33

Fig. 5. Count of Bacteria in sample images

(a) (b) (c) (d) (e)

Fig. 6. Five Sample images for bacterial detection

(a) (b) (c) (d) (e)

Fig. 7. (a) rod shaped bacteria (b) filamentous iron oxidizing bacterium (c) bacteria spores (d) filaments bacteria (e) curved rod bacteria

Table 2. Sample feature values extracted for clustering purpose

Perimeter	Circularity	Compactness Factor	Major axis	Minor axis	Eccentricity	Tortuosity
28.38905	1.09090512	0.91667	11.12029	6.2973	1.76587379	0.39171054
184.9859	1.34801774	0.74183	70.55497	44.230	1.59515113	0.3814072
95.41247	1.32784491	0.7531	37.76043	20.46736	1.84490965	0.3957599
68.19114	1.1912278	0.83947	26.64424	15.24395	1.74785669	0.3907287
125.1152	1.2603824	0.79341	48.68019	28.32607	1.71856491	0.3890829
41.80209	1.16260144	0.86014	16.11325	9.7192	1.6578714	0.3854651
41.20818	1.14035488	0.87692	16.26154	8.9259	1.82183349	0.3946192
43.88518	1.26125672	0.79286	18.15725	7.7835	2.33274235	0.4137444
35.47754	1.1000011	0.90909	13.54605	8.4594	1.60129962	0.3818204

Results in Fig. 8 shows the performance of *K-Means* and *SOM-Kmeans* clustering algorithms. Here, the quality measures considered are functions of average *Inter-Cluster* and the *Intra-Cluster* distances. Also used are the internal evaluation functions such as *Cluster Compactness* (*Cmp*), *Cluster Separation* (*Sep*) and the combined measure of *Overall Cluster Quality* (*Ocq*) to evaluate the *Intra-Cluster* homogeneity and the *Inter-Cluster* separation of the clustering results. Experimental simulations are performed using MATLab.

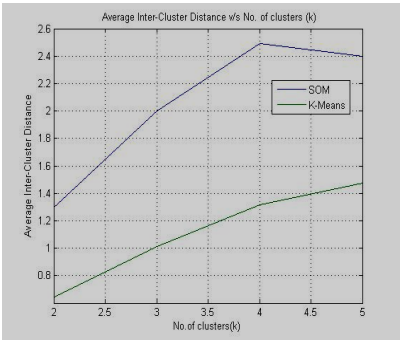
Cluster Compactness is used to evaluate the intra-cluster homogeneity of the clustering result and is defined as:

$$Cmp = \frac{1}{C}\sum_{i=1}^{C}\frac{v(c_i)}{v(X)}$$

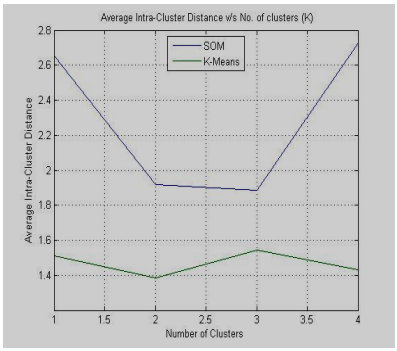
Where C is the number of clusters generated on the data set X, $v(c_i)$ is the deviation of the cluster c_i, and $v(X)$ is the deviation of the data set X given by:

$$v(X) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d^2(x_i, \bar{x})}$$

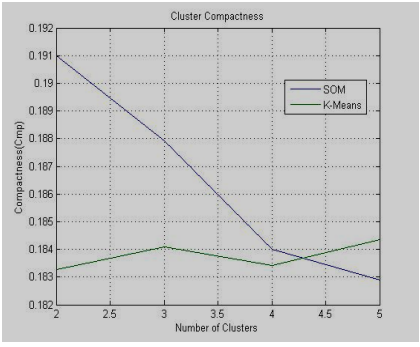
Where $d(x_i, x_j)$ is the Euclidean distance(L_2 norm), is a measure between two vectors x_i and x_j, N is the number of members in X, and $\bar{x}$ is the mean of X.



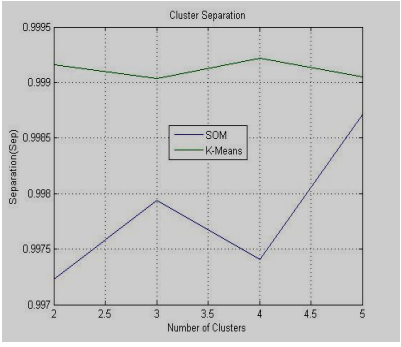
(a)



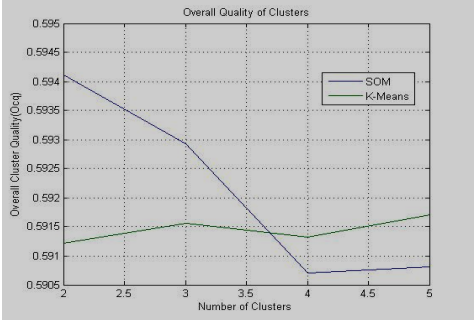
(b)



(c)



(d)



(e)

Fig. 8. Performance of *K-Means* and *SOM-Kmeans* clustering algorithms(a)Average Inter-cluster Distance v/s No. of k clusters (b)Average Intra-cluster distance v/s No. of clusters (c) Cluster Compactness (d) Cluster Separation (e) Overall quality of clusters

Cluster Separation is used to evaluate the intra-cluster separation of the clustering result and is defined as:

$$Sep = \frac{1}{C(C-1)} \sum_{i=1}^{C} \sum_{j=1, j \neq i}^{C} \exp\left(- \frac{d^2(x_{ci}, x_{cj})}{2\sigma^2} \right)$$

Where C is the number of clusters generated on the data set X, σ is the standard deviation of the data set X, and $d(x_{ci}, x_{cj})$ is the Euclidean distance, is a measure between centroid of x_{ci} and x_{cj}. Similar to *Cmp*, the larger the *Sep* value, the larger the overall dissimilarity among the output clusters. It is observed from the Fig.9.(d) that *Sep* value SOM and K-Means algorithms decreases with the increase in number of clusters.

The Overall cluster quality (*Ocq*) is used to evaluate both intra-cluster homogeneity and inter-cluster separation of the results of clustering algorithms. *Ocq* is defined as:

$$Ocq(\beta) = \beta * Cmp + (1 - \beta) * Sep$$

Where $\beta \in [0, 1]$ is the weight that balances the measures *Cmp* and *Sep*. A β value of *0.5* is often used to give equal weights to the two measures for overcoming the deficiency of each measure and assess the overall performance of a clustering system. Therefore, the lower the *Ocq* value, the better the quality of resulting clusters. It is observed from Fig.9(e) that, the *Ocq* value of SOM algorithm is lower compared to K-Means indicating the clusters formed by SOM are with better quality. The time complexity K-Means is *quad log time O* $(n*k*log_2 n)$ and SOM is *polynomial log time O* $(n*k*log_2 n)$ with varying number of iterations. Results of clustering algorithms may be used in the applications to evaluate medical report and to identify different types of microorganisms present in the sample, to count the number of microbes present in each cluster for the purpose of effective treatment, to detect the outlier etc.

7 Conclusions

Bacterial detection and identification process along with the clustering of bacteria have been presented. The results of bacterial clusters may be used in the areas of medical image analysis in microorganisms. This method will assist the doctors in deciding the intensity of the diseases/ infection which is dependent on microorganism count. This method can save critically ill patients. Future research directions in this regard concern with the development of adaptive predictive systems that use hybrid approach such as use of statistical, neural, and Bayesian learning algorithms for implementation of tool box with lesser cost and time and affordable by the hospitals.

References

1. Hans, C., Merchant, F.A., Shah, S.K.: ICVGIP 2010. ACM (2010)
2. Chayadevi, M.L., et al.: Towards Automating the Counting & Recognition of Microorganisms: A Simple Multiscan Approach. In: International Conference on Cognition and Recognition (2008)

3. Raju, G.T., Chayadevi, M.L.: Data Mining in Bioinformatics and its Applications–a Survey. In: ICDECS 2011 (December 2011) ISBN 978-93-81583-17-3
4. Raza, K.: Application of Data mining in Bioinformatics. Indian Journal of Computer Science and Engineering 1(2), 114–118 (2010)
5. Soni, N., Gandhi, C.: Application of Data Mining to Health Care. International Journal of Computer Science and its Applications (2010)
6. Hiremath, P.S., Parashuram, B.: Automatic Identification and classification of Bacilli Bacterial cell growth phases. IJCA special issue on Recent Trends in Image Processing and Pattern Recognition, RTIPPR (2010)
7. Bellazzi, R., Ferrazzi, F., Sacchi, L.: Predictive data mining in clinical medicine: a focus on selected methods and applications. In: Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 1 (2011)
8. Rulaningtya, R., Suksmono, A.B., Tati: Automatic Classification of Tuberculosis Bacteria using Neural n/w. In: International Conference on Electrical Engineering and Informatics (2011)
9. Shillabeer, A., Roddick, J.: Establishing a Lineage for medical Knowledge Discovery. ACM International Conference Proceeding Series 70, 29–37 (2007)
10. Prabhakar, S., Porkumaran, K., Samson Isaac, J.: Development of Image Processing scheme for bacterial. IEEE Classification Based on Optimal Discriminant Feature (2010)
11. Vanitha, V.: Classification of Medical Images using Support Vector. IACSIT press, Singapore (2011)
12. Abe, Y., Sagawa, T., Sakai, K., Kimura, S.: Enzyme-linked immunosorbent assay (ELISA) for human epidermal growth factor (hEGF). Clinica Chimica Acta 168, 87–95 (1987)
13. Cao, Y., et al.: Prediction of protein structural class with RoughSets. BMC Bioinformatics (2006)
14. Shen, Y., Liu, Z., Ott, J.: Support vector machines with L1 penalty for detecting gene-gene interactions. International Journal of Data Mining and Bioinformatics (2011)
15. Woolf, P.J., Wang, Y.: A fuzzy logic approach to analyzing gene expression data. Physiological Genomics 3(1), 9–15 (2000)

Analysis of Brain Activity for Motor Task Using Simultaneous EEG - fMRI

Sandhya M., Rose Dawn, and Rajanikant Panda

Cognitive Neuroscience Center,
National Institute of Mental Health and Neuroscience,
Bangalore, India
drsandym@gmail.com
drrosedawn@yahoo.com

Abstract. Integration of electroencephalogram (EEG) Signal and functional magnetic resonance imaging (fMRI) has opened a new avenue in Computational Neuro Cognitive Science studies as these two are complimentary methods for the analysis of brain activity. We are proposing a novel approach for the correlation between the EEG signal and fMRI to evaluate the blood oxygen level dependent (BOLD) signal changes and electrophysiological activity with high gain of spatio temporal resolution of the brain activity during a task. In this study, simultaneous EEG-fMRI was performed for hand motor task. sLORETA (Standardized low-resolution brain electromagnetic tomography) was reconstructed using segmented EEG for right-hand, left-hand movement and resting state. We estimated condition-specific effects by using statistical parametric maps (SPMs) with general linear model (GLM) approach .By correlating the sLORETA images of EEG and the fMRI images on MNI template during motor task we could demonstrate similar brain activity in both. However with the advantage of simultaneous EEG-fMRI signal acquired at same time points we had a better insight on temporo-spatial information of brain activity during a given task. This method has a huge utility value both in a clinical setting and cognitive studies for example in cryptogenic lesions on imaging with epileptic form discharges detected on EEG combined approach can explain the focal point of seizure and path of its spread to rest of the neuronal substrates on an anatomical image.

Keywords: EEG, fMRI, BOLD, sLORETA, SPM, GLM, Motor Task.

1 Introduction

Functional neuroimaging has been proved to be well suited to investigate the neuronal substrates involved during certain cognitive functions in the brain by detecting perfusion changes. This has lead to a better understanding about where cognitive processes take place in the brain and their role in brain function. Therefore, investigation of functional interactions as well as information about the neuronal direction of these interactions has become an issue in cognitive neurosciences in recent times. Structural and functional Magnetic Resonance Imaging (MRI) techniques have made major contributions to the understanding of the brain in health and disease. MRI

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1101–1107.
springerlink.com

is non-invasive, allows the examination of brain structure and function with high spatial resolution but it lacks the temporal resolution at the level or speed at which cognitive neuronal processing occurs. EEG signature provides good neuronal information with high temporal resolution but lacks the spatial information at the level of sub cortical neuronal substrates such basal ganglia. In order to utilize the information provided by these two complimentary methods EEG-correlated fMRI has been used to map neuro haemodynamic changes that occur with neuronal activity [10,17] also acquisition of EEG during fMRI provides an additional monitoring tool for the analysis of brain state fluctuations. This technique when extended in epilepsy offers a new opportunity to localize the generators of interictal epileptiform discharges (IEDs) and capture the perfusion changes occurring using blood oxygen level-dependent (BOLD) contrast [3] at time points of the seizure activity provided by EEG. Initial EEG-fMRI experiments investigating spikes used EEG triggering to capture fmri images following pre specified events [12] but more recently, it has become possible to acquire good-quality EEG and fMRI simultaneously giving good quality uninterrupted EEG recordings and BOLD time course information [6,7, 8,9,11].

In this paper, we use simultaneous EEG-fMRI in humans and employ a motor task to elicit evoked activity in motor cortex. As scalp EEG measures the activity of multiple distributed neuronal processes, we used a low-resolution brain electromagnetic tomography-LORETA approach to isolate activity that was primarily related to the motor task. The resulting time series was then used as a surrogate for neuronal activity. The fMRI images processed useing a standard statistical parametric mapping (SPM). We then correlated the fMRI data onto low-resolution brain electromagnetic tomography of the EEG data.

2 Materials and Methods

Simultaneous EEG fMRI was obtained for motor task at 3T MRI. The subjects had no history of neuro- psychiatric illness. The subjects were right handed. The paradigm included alternate hand movement with periods of rest intermittently.

2.1 Paradigm

The task paradigm included 3 cycles of self paced motor task of bilateral hands alternating with rest periods as cued by the investigator, after an initial rest period of 30 seconds. For the motor task subject was instructed to oppose all the fingers simultaneously left hand followed by right hand in a sequence. During each motor task period the subject was cued first to use the left hand for 30 seconds and then to use the right hand for 30 seconds. Subjects were asked to perform the task as rapidly as they could. Task performance was visually monitored during the functional MR imaging studies. Subjects were instructed to keep their eyes open throughout the imaging study, to concentrate on task as quickly and evenly as possible when given visual cues, to avoid head movement, and to refrain as much as possible from higher cognitive processes during the rest periods. The condition for successive blocks

alternated between rest and the task, starting with rest followed by left hand and then right hand task. This rest-task paradigm will yield 3 sets of rest and activity involving right and left hand respectively. During the activation period, the patient will perform rapid self-paced motor task of all the fingers of right and left hand alternatively.

2.2 EEG Data Acquisitions

EEG data were recorded using a 32-channel MR compatible EEG system (Brain Products, Gilching, Germany). MRI compatible electrode cap(BrainCap MR, Germany), which was fed via short cables to the amplifiers inside the scanner room. The EEG cap consisted of 31 scalp electrodes placed according to the international 10-20 system electrode placement and one additional electrode dedicated to the electrocardiogram (ECG), which was placed on the back of the subject, approx. around 15 cm below the shoulder and approx. 1 cm left of the midline . Data were recorded relative to an FCz reference and a ground electrode was located at Iz (10–5 electrode system, (ostenveld and Praamstra, 2001). Data were sampled at 5000 Hz, with a bandpass of 0.016–250 Hz along with 50 Hz notch filtering. The impedance between electrode and scalp was kept below 5 kΩ. To synchronize the sampling clocks of the MR and EEG systems we used the Sync Box (BrainProducts), thus making the times of fMRI volume acquisition available for later gradient artifact removal. We placed the EEG amplifiers and the battery (PowerPack, BrainProducts) inside of the magnet's bore (approx. half m distance between EEG cap connectors and the amplifiers), aligned with the b0 field and the wires that connected the cap with the amplifier were fixed in their position with sand bags to avoid vibrations. The digital output from the amplifiers was fed via optical cables into a dedicated laptop positioned outside the scanner. EEG was recorded using the Brain Recorder software (Version 1.03, BrainProducts). To prevent the head movement of subject we placed sufficient head padding at the scanner and for the cable movement was present by fixation of the cap braid at the head coil with one or two pieces of tape.

2.3 fMRI Data Acquisition

Functional MR-images were acquired using a 3T scanner (Skyra, Siemens, Erlangen, Germany). The subject's head was positioned within a prototype radio-frequency quadrature bird cage coil with foam padding to provide comfort and to minimize head movements. Preliminary anatomic images included a sagittal localizer First, a T1-weighted three-dimensional high resolution imaging was performed to facilitate localization of fMRI activation. All axial sections were oriented parallel to the ac-pc (anterior commissure-posterior commissure) line After obtaining the anatomical MR images, echo-planar images (EPI) using BOLD contrast was obtained, 95 volumes were obtained applying the following EPI parameters: 34 slices, 6 mm slice thickness without any inter-slice gap, FOV 192×192 mm, matrix 64×64, repetition time 3000ms, echo time 35ms, refocusing pulse 90°, matrix- 256 x 256 x 114, voxel size-1 x 1 x 1mm. The acquisitions were grouped in blocks of 10 dynamics each.

3 Data Analysis

The EEG-fMRI data was analyzed in the fallowing three steps for getting activation areas of brain for motor task:

1) EEG data analysis 2) fMRI data analysis 3) correlation of output of two data set (EEG -fMRI).

3.1 EEG Data Artifact Removals and Preprocessing

Raw EEG data were processed offline using BrainVision Analyzer version 2 (Brain Products, Gilching, Germany). Gradient artifact correction was performed using modified versions of the algorithms proposed by Allen et al. [1] where a gradient artifact template is subtracted from the EEG using a baseline corrected sliding average of 20 MR-volumes. Data were then down-sampled to 250 Hz and low-pass filtered with an IIR filter with a cut-off frequency of 70 Hz. Following gradient artifact correction, the data were corrected for cardio ballistic artifacts. The cardio ballistic artifact is cardiac pulse artifact due to the static B0 field of the MR-tomography, This name is thought to be predominantly caused by cardiac-related body and electrode movement due to expansions and contraction of scalp arteries between the systolic and diastolic phase [2, 8, 13]. Presumably to a lesser extent, there may also be fluctuations in the Hall-voltage (a potential which is created across a conductor with a current flow perpendicular to a surrounding magnetic field) due to the pulsating speed changes of the blood in the arteries [13]. For this, an average artifact subtraction method [2] was implemented in Brain Vision Analyzer2. This method involves subtracting the artifact on a second by second basis using heartbeat events (R peaks) detected in the previous 10 s. As such it requires accurate detection of R peaks which is aided by the employment of a moving average low pass filter and a finite impulse response high pass filter [2]. In the present study, the R peaks were detected semi-automatically, with manual adjustment for peaks misidentified by the software. To average the artifact in the EEG channels, the R peaks are transferred from the ECG to the EEG over a selectable time delay. The average artifact was then subtracted from the EEG. After that the ocular artifact corrected using infomax ICA with biased infomax technique, the vertical electrooculography components was detected using global field power. Once gradient, cardio ballistic and ocular artifacts had been removed, the data were then inspected for artifacts resulting from muscular sources or any other electro-physiological artifact and any epoch containing a voltage change of more than 150 μV was rejected. After that the EEG data was segmented as per left hand, right hand movement and rest condition for further study.

3.2 EEG Post Processing

On the basis of the scalp-recorded electric potential distribution, sLORETA was used to compute the cortical three-dimensional distribution of current density for all segmented data sets (left hand, right hand movement and rest condition) [15, 16]. Low resolution electromagnetic tomography (LORETA) assumes that the smoothest of all activity distributions is most plausible (‘‘smoothness assumption’’) and therefore, a particular current density distribution is found [15]. This method followed

by an appropriate standardization of the current density, producing images of electric neuronal activity without localization bias. Computations were made in a realistic head model using the MNI152 template [10] with the three-dimensional solution space restricted to cortical gray matter. sLORETA images represent the electric activity at each voxel in neuroanatomic Talairach space as the squared standardized magnitude of the estimated current density.

3.3 fMRI Analysis

Our study aims to detect hemodynamic response during bilateral hand motor task in healthy controls. The fMRI analysis was performed using statistical parametric mapping (SPM8; Welcome Department of Cognitive Neurology, London). The first five functional image frames of each time series were discarded to allow for signal equilibration, giving a total of 90 frames used in analysis. After that the data were realigned for motion correction by registration to the mean image. The images were then normalized to the MNI space [18]. Finally images were smoothed with a Gaussian kernel of 6mm. SPM combines the general linear model (GLM) [4] and Gaussian field theory to draw statistical inferences from bold response data regarding deviations from null hypotheses in 3 dimensional brain space. The realigned, normalized and smoothed data were modeled using a boxcar function convolved with a canonical haemodynamic response function [5, 14]. Reference functions were performed for left hand motor task and right hand motor task separately.

3.4 Comparison between fMRI and EEG Localizations

This comparison has been done in calculating the Euclidean distances between the LORETA current density on MNI space and the BOLD images on MNI space in the same anatomical structures.

4 Result

The results obtained from EEG and fMRI were rendered on the MNI template were compared and correlated, Subjects performed motor task. Significant activation with cluster size 5 and $p<0.01$ uncorrected in motor cortex was considered significant. Activation was noted in the motor cortex on fMRI. Similarly increase in the beta frequency band was noted on EEG in the same location on EEG analysis. State-of-the-art techniques allow EEG activity up to high frequencies in the gamma range to be acquired simultaneously with fMRI data. The utilization of fMRI evidence to better constrain solutions of the inverse problem of source localization of EEG activity is an exciting possibility. Nonetheless, this approach should be applied cautiously since the degree of overlap between underlying neuronal activity sources is variable and, for the most part, unknown. So multimodal integration of EEG signal data with fMRI is possible to get good spatiotemporal information and thus it increases the diagnostic confidence.

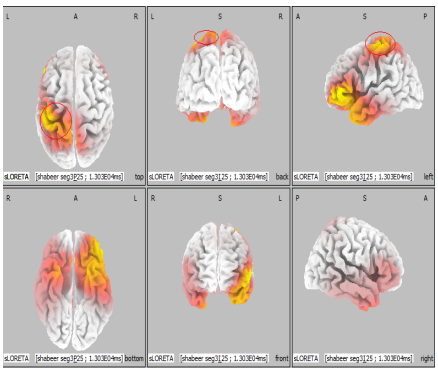


Fig. 1.

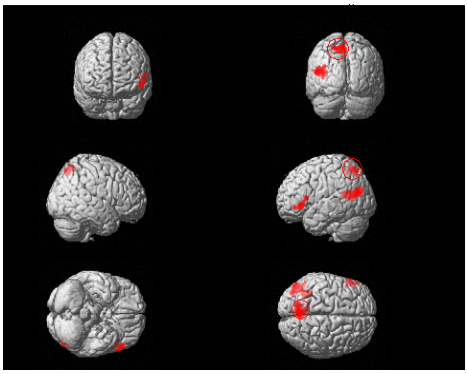


Fig. 2.

Fig 1. EEG-LORETA source localization for showing activation localization at in the right hand motor task, Fig 2. Mean fMRI activation map during the motor task with the functional image overlaid onto the 3D anatomical image in Talairach space showing activation localization in the right hand motor task.

5 Conclusion

Our findings in a simple motor task show that it is possible to obtain the areas activated for motor task using simultaneous EEG–fMRI correlation study to get high resolution of both the spatiotemporal information. By correlating fMRI-BOLD and EEG we may identify more accurately which regions contribute to changes of the electrical response. The clinical utility of this can be considered for demonstrating seizure-related EEG-BOLD signal in the motor cortex even in MRI negative seizure cases.

References

1. Allen, P.J., Josephs, O., Turner, R.: A method for removing imaging artifact from continuous EEG recorded during functional MRI. Neuroimage 12, 230–239 (2000)
2. Allen, A.J., Polizzi, G., Krakow, K., Fish, D., Lemieux, L.: Identification of EEG events in the MR scanner: the problem of pulse artifact and a method for its subtraction. Neuroimage 8, 229–239 (1998)
3. Diehl, B., Salek-haddadi, A., Fish, D.R., Lemieux, L.: Mapping of spikes, slow waves, and motor tasks in a patient with malformation of cortical development using simultaneous EEG and fMRI. Magnetic Resonance Imaging 21, 1167–1173 (2003)
4. Friston, K.J., Homes, A., Worsley, K., Poline, J., Frith, C., Frackowiak, R.S.: Statistical parametric maps in functional imaging: a general linear approach. Human Brain Mapping 2, 189–210 (1994)
5. Friston, K.J., Ashburner, J., Kiebel, S.J., Nichols, T.E., Penny, W.D.: Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press (2007)

6. ois Lazeyras, F., Zimine, I., Blanke, O., Perrig, S.H., Seeck, M.: Functional MRI With Simultaneous EEG Recording: Feasibility and Application to Motor and Visual Activation. Journal of Magnetic Resonance Imaging 13, 943–948 (2001)
7. Gerloff, C., Grodd, W., Altenmüller, E., Kolb, R., Naegele, T., Klose, U., Voigt, K., Dichgans, J.: Co-registration of EEG and fMRI in a Simple Motor Task. Human Brain Mapping 4, 199–209 (1996)
8. Goldman, R.I., Stern, J.M., Engel, J., Cohen, M.S.: Acquiring simultaneous EEG and functional MRI. Clin. Neurophysiol. 111, 1974–1980 (2000)
9. Yuan, H., Liu, T., Szarkowski, R., Rios, C., Ashe, J., He, B.: Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements. NeuroImage 49, 2596–2606 (2010)
10. Mazziotta, et al.: A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philos. Trans. R Soc. Lond. B Biol. Sci. 29; 356(1412), 1293–1322 (2000)
11. Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., Möller, H.J., Juckel, G., Hegerl, U.: Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI. NeuroImage 28, 49–58 (2005)
12. Mulert, C., Jager, L., Schmitt, R., Bussfeld, P., Pogarell, O., Moller, H.J., Juckel, G.: Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. NeuroImage 22, 83–94 (2004)
13. Muri, R.M., Felblinger, J., Rosler, K.M., Jung, B., Hess, C.W., Boesch, C.: Recording of electrical brain activity in a magnetic resonance environment: distorting effects of the static magnetic field. Magnetic Resonance in Medicine 39, 18–22 (1998)
14. Penny, W., Holmes, A.: Random effects analysis. In: Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W. (eds.) Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier, London (2006)
15. Pascual-Marqui, R.D., Michel, C.M., Lehmann, D.: Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. International Journal Psychophysiol. 18, 49–65 (1994)
16. Pascual-Marqui, R.D., Lehmann, D., Koenig, T., Kochi, K., Merlo, M.C., Hell, D.: Low resolution brain electromagnetic tomography (LORETA) functional imaging in acute, neuroleptic- naive, first-episode, productive schizophrenia. Psychiatry Res. 90, 169–179 (1999)
17. Rosaa, M.J., Kilnera, J., Blankenburgb, F., Josephsa, O., Penny, W.: Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMR. NeuroImage 49(2), 1496–1509 (2010)

K–Means Clustering Microaggregation for Statistical Disclosure Control

Md. Enamul Kabir, Abdun Naser Mahmood, and Abdul K. Mustafa

University of New South Wales,
Australia
Abdun.Mahmood@unsw.edu.au,
abdul.mustafa@jcu.edu.au

Abstract. This paper presents a K-means clustering technique that satisfies the bi-objective function to minimize the information loss and maintain k-anonymity. The proposed technique starts with one cluster and subsequently partitions the dataset into two or more clusters such that the total information loss across all clusters is the least, while satisfying the k-anonymity requirement. The structure of *K*– means clustering problem is defined and investigated and an algorithm of the proposed problem is developed. The performance of the *K*– means clustering algorithm is compared against the most recent microaggregation methods. Experimental results show that *K*– means clustering algorithm incurs less information loss than the latest microaggregation methods for all of the test situations.

1 Introduction

Microaggregation is a family of Statistical Disclosure Control (SDC) methods for protecting microdata sets that have been extensively studied recently [1, 2, 7, 9]. The basic idea of microaggregation is to partition a dataset into mutually exclusive groups of at least k records prior to publication, and then publish the centroid over each group instead of individual records. The resulting anonymized dataset satisfies k-anonymity [6], requiring each record in a dataset to be identical to at least $(k-1)$ other records in the same dataset.

The effectiveness of a microaggregation method is measured by calculating its information loss. *k*-anonymity [5, 6, 8] provides sufficient protection of personal confidentiality of microdata, while ensuring the quality of the anonymized dataset, an effective microaggregation method should incur as little information loss as possible. To minimize the information loss due to microaggregation, all records are partitioned into several groups such that each group contains at least *k* similar records, and then the records in each group are replaced by their corresponding mean such that the values of each variable are the same. Such similar groups are known as clusters.

The reminder of this paper is organized as follows. We introduce a problem of microaggregation in Section 2. Section 3 introduces the basic concept of microaggregation. We present a brief description of our proposed microaggregation method in Section 4. Section 5 shows experimental results of the proposed method. Finally, concluding remarks are included in Section 6.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1109–1115.
springerlink.com

2 Problem Statement

The algorithms for microaggregation works by partitioning the microdata into homogeneous groups so that information loss is low. The level of privacy required is controlled by a security parameter k, the minimum number of records in a cluster. This work presents a new clustering-based method for microaggregation which finds the minimal information loss clustering for an increasing number of clusters. The method works by calculating the maximum number of clusters by K =int⌊^{k} ⌋, where n is the total number of records in the dataset and k is the anonymity parameter for k-anonymization. Recall that in a k-anonymous clustering each cluster must have k or more instances. It is easy to prove that a clustering which satisfies k+1 anonymity also satisfies k anonymity. Therefore, the premise of this work is to find a k+i anonymous clustering which has the lowest information loss. The trivial solution is to form a single cluster with all the records in the dataset and calculate the information loss. Clearly, this cluster is k-anonymous (assuming k << n), however, the information loss may be high. Observe that in the rare case where every instance in the dataset is identical, this method can find the k-anonymous clustering in the quickest possible manner. For the general case, total information loss would decrease as the number of cluster increases. Note, in the rare case that all the instances are completely different such that they belong to their own clusters, total information loss would be zero since there is no information loss due to each cluster represented by one instance. However, this would certainly breach k-anonymity requirement since k must be greater than 1. Consequently, the problem is to design a technique that can take advantage of this k-anonymity property by checking fewer clusters first, which is a different approach taken from existing methods. The proposed method is explained in Section 4 and compared against the most recent widely used microaggregation methods in Section 5. The experimental results demonstrate that the proposed microaggregation technique outperforms all of the compared techniques for at least one of the benchmark datasets and has comparable results with these techniques for the other dataset.

3 Background

Consider a microdata set T with p numeric attributes and n records, where each record is represented as a vector in a p-dimensional space. For a given positive integer $k \leq n$, a microaggregation method partitions T into K clusters, where each cluster contains at least k records (to satisfy k-anonymity), and then replaces the records in each cluster with the centroid of the cluster. Let n_i denote the number of records in the ith cluster, and x_{ij} , $1 \leq j \leq n_i$, denote the jth record in the ith cluster. Then, $n_i \geq k$ for $i = 1$ to K, and $\sum_{i=1}^{K} n_i = n$ The centroid of the ith cluster, denoted by $\overline{x}_i$ is calculated as the average vector of all the records in the ith cluster.

In the same way, the centroid of T, denoted by $\overline{x}$, is the average vector of all the records in T. Information loss is used to quantify the amount of information of a dataset that is lost after applying a microaggregation method. In this paper we use the most common definition of information loss by Domingo-Ferrer and MateoSanz [1] as follows:

$$IL = \frac{SSE}{SST}$$

where *SSE* is the within-cluster squared error, calculated by summing the Euclidan distance of each record *xi j* to the average value $\overline{x}_i$ as follows:

$$SSE = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x}_i \right)' \left(x_{ij} - \overline{x}_i \right)$$

and *SST* is the sum of squared error within the entire dataset *T*, calculated by summing the Euclidean distance of each record $x_{i\,j}$ to the average value $\overline{x}$ as follows:

$$SSE = \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(x_{ij} - \overline{x} \right)' \left(x_{ij} - \overline{x} \right)$$

4 The Proposed Approach

This section presents the proposed *K*-means based anonymization technique to solve the dual objective of minimum information loss and *k*-anonymity. The proposed approach builds one cluster at the first instance and subsequently adding more clusters such that *k*-anonymity requirements and information losses are guaranteed.

4.1 Clustering Technique

One of the most widely used clustering algorithms is Lloyd's *K*-means algorithm [12]. Given a set of *N* records (n_1,n_2, · · ·n_n), where each record is a *d*-dimensional vector, the *K*-means clustering partitions the *N* records into *K* clusters ($K < N$) S = (S_1,S_2, · · · ,S_k) such that intra cluster distance is minimized and inter cluster distance is maximized. The number of clusters to be fixed in *K*-means clustering. Let the initial centroids be (w_1,w_2, · · · ,w_k) be initialized to one of the *N* input patterns. The quality of the clustering is determined by the following error function.

$$E = \sum_{i=1}^{k} \sum_{n_l \varepsilon C_j} \left\| n_l - w_j \right\|^2$$

where C_j is the j^{th} cluster whose value is a disjoint subset of input patterns.

K means algorithm works iteratively on a given set of *K* clusters. Each iteration consists of two steps:

- Each data item is compared with the K centroids and associated with the closest centroid creating K clusters.
- The new sets of centroids are determined as the mean of the points in the cluster created in the previous step.

The algorithm repeats until the centroids do not change or when the error reaches a threshold value. The computational complexity of algorithm is *O(NKd)*.

Table 1. K–means clustering algorithm

Input: a dataset T of n records and a positive integer k. Output: a partitioning $G = \{G_1, G_2, ..., G_K\}$ of T, where $K = \|G\|$ and $G_i \geq k$ for $i = 1$ to K. 1. Let $K = \text{int}\left\lfloor \frac{n}{k} \right\rfloor$; 2. Form a cluster with all records in T and calculate the information; loss. Obviously the information loss would be 1; 3. Form one more cluster that causes least information loss among all; possible combination of such clusters. Check each cluster satisfy; k–anonymity requirement; 4. Choose clusters that cause least information loss and satisfy the; k-anonymity requirement; 5. Repeat steps 3-4 for up to K clusters and finally select clusters; where least information loss and k–anonymity are guaranteed;

4.2 K–Means Anonymization Technique

Based on the clustering technique and the definition of the microaggregation problem, next we discuss the k–means clustering microaggregation algorithm.

The algorithm first identifies the maximum number of clusters by, $K = \frac{n}{K}$, where k is the anonymity parameter for k-anonymization and round this as integer. Form a cluster with all the n records in the dataset. It will then form two clusters (see step 3 of Table 1) that causes least information loss and satisfy the k-anonymity requirement. The algorithm compares the information loss with the previous step and selects clusters that satisfy both the requirements of data quality and the anonymity parameter (see step 4 of Table 1. The algorithm then continues to build clusters (see step 5 of Table 1) up to K (maximum number of clusters) and finally selects the optimum number of clusters where both the least information loss and the k-anonymity requirements are satisfied.

5 Experimental Results

The objective of our experiment is to investigate the performance of our approach in terms of data quality. We demonstrate the effectiveness of the proposed approach by comparing it against a basket of well-known techniques. The following two datasets [3], which have been used as benchmarks in previous studies to evaluate various microaggregation methods, were adopted in our experiments.

1. "Tarragona" dataset contains 834 records with 13 numerical attributes.
2. The "Census" dataset contains 1,080 records with 13 numerical.

To accurately evaluate our approach, the performance of the proposed *K*-means clustering mcroaggregation algorithm is compared in this section with various microaggregation. Tables 2-3 show the information loss for several values of *k* for the *Census* and for the *Tarragona* datasets respectively.

Table 2. Information loss comparison using Census dataset

Method	k = 3	k = 4	k = 5	k = 10
MDAV-MHM	5.6523	9.0870	14.2239	
MD-MHM	5.69724	8.98594	14.3965	
CBFS-MHM	5.6734	8.8942	13.8925	
NPN-MHM	6.3498	11.3443	18.7335	
M-d	6.1100	8.24	10.3000	17.1700
μ -Approx	6.25	8.47	10.78	17.01
TFRP-1	5.931	7.880	9.357	14.442
TFRP-2	5.803	7.638	8.980	13.959
MDAV-1	5.692186279	7.494699833	9.088435498	14.15593043
MDAV-2	5.656049371	7.409645342	9.012389597	13.94411775
DBA-1	6.144855154	9.127883805	10.84218735	15.78549732
DBA-2	5.581605762	7.591307664	9.046162117	13.52140518
K-C	**3.575**	**3.9561**	**4.532**	**6.8419**

Table 3. Information loss comparison using Tarragona dataset

Method	k = 3	k = 4	k = 5	k = 10
MDAV-MHM	16.9326	22.4617	33.1923	
MD-MHM	16.9829	22.5269	33.1834	
CBFS-MHM	16.9714	22.8227	33.2188	
NPN-MHM	17.3949	27.0213	40.1831	
M-d	16.6300	19.66	24.5000	38.5800
μ -Approx	17.10	20.51	26.04	38.80
TFRP-1	17.228	19.396	22.110	33.186
TFRP-2	16.881	19.181	21.847	33.088
MDAV-1	16.93258762	19.54578612	22.46128236	33.19235838
MDAV-2	16.38261429	**19.01314997**	22.07965363	33.17932950
DBA-1	20.69948803	23.82761456	26.00129826	35.39295837
DBA-2	**16.15265063**	22.67107728	25.45039236	34.80675148
K-C	20.2425	20.2425	**20.2425**	**23.9761**

The information loss is compared with the *K*-means clustering microaggregation algorithm among the latest microaggregation methods listed above. Information loss is measured as $\frac{\text{SSE}}{\text{SST}} \times 100$, where SST is the total sum of the squares of the dataset. Note that the within-groups sum of squares SSE is never greater than SST so that the reported information loss measure takes values in the range [0,100].

Tables 2-3 show the lowest information losses obtained by applying all the microaggregation methods. The information loss of the proposed algorithm (**K-C**) is at the last row of each table. The lowest information loss for each dataset and each k value is shown in bold face. Note that the proposed algorithm has the best performance

among all the techniques for the *Census* dataset. For the *Tarragona* dataset **K-C** has the lowest information for $k = 5$ and $k = 10$, but DBA-2 and MDAV-2 have the lowest values for $k = 3$ and $k = 4$, respectively. The information losses of methods DBA-1, DBA-2, MDAV-1 and MDAV-2 are quoted from [11]; the information losses of methods MDAV-MHM, MD-MHM, CBFS-MHM, NPN-MHM and M-d (for $k = 3, 5, 10$) are quoted from [3]; the information losses of methods μ -Approx and M-d (for $k = 4$) are quoted from [4], and the information losses of methods TFRP-1 and TFRP-2 are quoted from [10]. TFRP is a two-stage method and its two stages are denoted as TRFP-1 and TRFP-2 respectively. The TFRP-2 is similar to the DBA-2 but disallows merging a record to a group of size over $(4k - 1)$. The experimental results illustrate that in all of the test situations, the *K*- means algorithm incurs significantly less information loss than any of the microaggregation methods listed in the table.

6 Conclusion

Microaggregation is an effective method in SDC to protect privacy in microdata and has been extensively used world-wide. This work has presented a new *K*-means clustering mcroaggregation method for numerical attributes that works by partitioning the dataset into as few clusters as possible with the lowest information loss. A comparison has been made of the proposed algorithm with the most widely used microaggregation methods using two benchmark datasets (Census and Tarragona). The experimental results show that the proposed algorithm has a significant dominance over the recent microaggregation methods with respect to information loss. Is very effective microaggregation method in preserving the privacy of data.

References

1. Domingo-Ferrer, J., Mateo-Sanz, J.: Practical data-oriented microaggregation for statistical disclosure control. IEEE Transactions on Knowledge and Data Engineering 14(1), 189–201 (2002)
2. Domingo-Ferrer, J., Torra, V.: Ordinal, continuous and heterogeneous kanonymity through microaggregation. Data Mining and Knowledge Discovery 11(2), 195–212 (2005)
3. Domingo-Ferrer, J., Martinez-Balleste, A., Mateo-Sanz, J.M., Sebe, F.: Efficient multivariate data-oriented microaggregation. The VLDB Journal 15(4), 355–369 (2006)
4. Domingo-Ferrer, J., Sebe, F., Solanas, A.: A polynomial-time approximation to optimal mul tivariate microaggregation. Computer and Mathematics with Applications 55(4), 714–732 (2008)
5. Samarati, P.: Protecting respondent's privacy in microdata release. IEEE Transactions on Knowledge and Data Engineering 13(6), 1010–1027 (2001)
6. Sweeney, L.: k-Anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10(5), 557–570 (2002)
7. Kabir, M.E., Wang, H.: Systematic Clustering-based Microaggregation for Statistical Disclosure Control. In: Proc. IEEE International Conference on Network and System Security, Melbourne, pp. 435–441 (September 2010)
8. Kabir, M.E., Wang, H., Bertino, E., Chi, Y.: Systematic Clustering Method for l-diversity Model. In: Proc. Australasian Database Conference, Brisbane, pp. 93–102 (January 2010)

9. Kabir, M.E., Wang, H.: Microdata Protection Method Through Microaggragation: A Median Based Approach. Information Security Journal: A Global Perspective (in press)
10. Chang, C.-C., Li, Y.-C., Huang, W.-H.: TFRP: An efficient microaggregation algorithm for statistical disclosure control. Journal of Systems and Software 80(11), 1866–1878 (2007)
11. Lin, J.-L., Wen, T.-H., Hsieh, J.-C., Chang, P.-C.: Density-based microaggregation for statistical disclosure control. Expert Systems with Applications 37(4), 3256–3263 (2010)
12. Lloyd, S.: Least squares quantization in PCM. IEEE Transactions on Information Theory 28(2), 129–137 (1982)

Content Based Image Retrieval Using Sketches

M. Narayana and Subhash Kulkarni

Jaya Prakash Narayan College of Engineering, Mahabubnagar, AP, India
sai_15surya@yahoo.co.in, subhashsk@gmail.com

Abstract. This paper aims to introduce the problems and challenges concerned with the design and creation of CBIR systems, which is based on a free hand sketch (Sketched based image retrieval-SBIR). This analysis led us to studying the usability of a method for computing dissimilarity between user-produced pictorial queries and database images according to features extracted from Gray-Level Co-occurrence Matrix (GLCM) automatically.

CBIR is generally characterized by the methods that consumes less time. Hence fast content – based image retrieval is a need of the day especially image mining for shapes, as image database is growing exponentially in size with time. In this paper, texture features extracted from GLCM, tested, and investigated on different standard databases is proposed, it exhibits invariant to rotation. The retrieval performance of the proposed method is showed for both the dinosaurs retrieval efficiency achieved about 95% and precision also 95% where color is not dominant. It is also observed that the proposed method achieved low retrieval performance over these four image features for sketch based and color dominant images. This process can be used as coarse level in hierarchical CBIR that reduces the database size from very large set to a small one. This tiny database can further be scrutinized rigorously using the Edge Histogram Descriptor (EHD) and Color and Color Co-occurrence Matrix (CCM) etc.

1 Introduction

The growing of data storages and revolution of internet had changed the world. The efficiency of searching in information set is a very important point of view. In case of texts we can search flexibly using keywords, but if we use images, we cannot apply dynamic methods. Two questions can come up. The first is who yields the keywords. And the second is an image can be well represented by keywords. In many cases if we want to search efficiently some data have to be recalled. The human is able to recall visual information more easily using for example the shape of an object [9, 12, 13], or arrangement of colors and objects. Our purpose is to develop a content based image retrieval system, which can retrieve using sketches in frequently used databases. The user has a drawing area where he can draw those sketches, which are the base of the retrieval method [4, 8, 11]. Using a sketch based system can be very important and efficient in many areas of the life. In the following paragraph some application possibilities are analyzed. The CBIR systems have a big significance in the criminal investigation. The identification of unsubstantial images, tattoos and graffities can be supported by these systems. Similar applications are implemented in [5, 6, 7].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1117–1123.
springerlink.com

Another possible application area of sketch based information retrieval is the searching of analog circuit graphs from a big database [3].

The Sketch-based image retrieval (SBIR) was introduced in QBIC [2] and Visual SEEK [10] systems. In these systems the user draws color sketches and blobs on the drawing area. The images were divided into grids, and the color and texture features were determined in these grids. The applications of grids were also used in other algorithms, for example in the edge histogram descriptor (EHD) method [1]. The disadvantage of these methods is that they are not invariant opposite rotation, scaling and translation.

2 Proposed Architecture

The objective of the proposed work in this paper is to study the texture features from GLCM as effective features for CBIR. CBIR system retrieves the relevant shapes from the image database for the given query sketch image or original image by computing the features of the query image and comparing with similar feature set of corresponding images in the database. Relevant shapes having minimum distance (or maximum similarity) computed between features of query image and feature set in image database are retrieved.

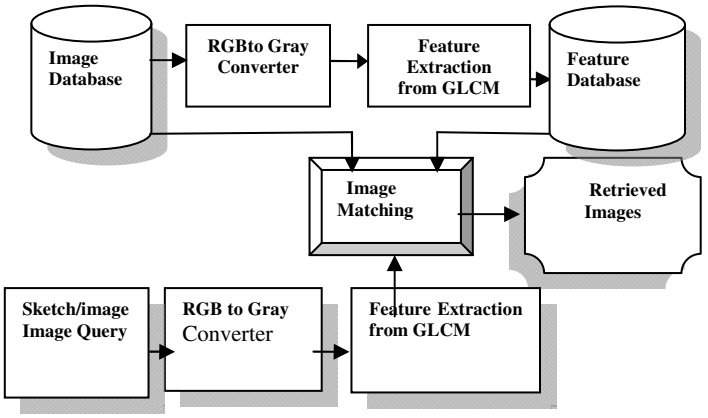


Fig. 1. Architecture of the system

2.1 The Purpose of the System

Even though the measure of research in sketch-based image retrieval increases, there is no widely used SBIR system. Our goal is to develop a content-based associative search engine, which databases are available for anyone looking back to freehand drawing. CBIR is generally characterized by the methods that consumes less time. Hence fast content – based image retrieval is a need of the day especially image mining for shapes, as image database is growing exponentially in size with time. In this paper, texture features extracted from GLCM is proposed and tested on standard

databases, it exhibits invariant to rotation. This process can be used as coarse level in hierarchical CBIR that reduces the database size from very large set to a small one. This tiny database can further be scrutinized rigorously using the Edge Histogram Descriptor (EHD), the histogram of oriented gradients (HOD), and Color and Color Co-occurrence Matrix (CCM) etc. In our system the iteration of the utilization process is possible, by the current results looking again, thus increasing the precision.

Examine the data flow model of the system from the user's point of view. It is shown in Figure 1. First the user draws a sketch or loads an image. When the drawing has been finished or the appropriate representative has been loaded, the retrieval process is started. The content-based retrieval as a process can be divided into two main phases. The first is the database construction phase, in which the texture features are extracted from GLCM and stored in the form of feature vectors – this is the off-line part of the program. This part carries out the computation intensive tasks, which has to be done before the program actual use. The other phase is the retrieval process, which is the on-line unit of the program.

The performance of the proposed CBIR system is tested by retrieving the specified number of shapes from the database. The average retrieval rate and retrieval time are the main performance measures in the proposed CBIR system. The average retrieval rate is known as the percentage average number of shapes belonging to the same image as the test (query) shape in the top 'N' matches. 'N' indicates the number of retrieved shapes.

3 Texture Feature Extraction Based on GLCM

GLCM creates a matrix with the directions and distances between pixels, and then extracts meaningful statistics from the matrix as texture features. GLCM texture features commonly used are shown in the following:

GLCM is composed of the probability value, it is defined by $p(i, j \mid d, \theta)$ which expresses the probability of the couple pixels at θ direction and d interval. When θ and d is determined, $p(i, j \mid d, \theta)$ is showed by $P_{i,j}$. Distinctly GLCM is a symmetry matrix; its level is determined by the image gray-level. Elements in the matrix are computed by the equation showed as follow:

$$P(i, j \mid d, \theta) = \frac{P(i, j \mid d, \theta)}{\sum_i \sum_j P(i, j \mid d, \theta)} \tag{1}$$

GLCM expresses the texture feature according the correlation of the couple pixels gray-level at different positions. It quantificationally describes the texture feature. In this paper, four features is selected, include energy, contrast, entropy, inverse difference.

$$\text{Energy } E = \sum_x \sum_y P(x, y)^2 \tag{2}$$

It is a gray-scale image texture measure of homogeneity changing, reflecting the distribution of image gray-scale uniformity of weight and texture.

$$\text{Contrast } I = \sum\sum (x-y)^2 P(x,y) \tag{3}$$

Contrast is the main diagonal near the moment of inertia, which measure the value of the matrix is distributed and images of local changes in number, reflecting the image clarity and texture of shadow depth. Contrast is large means texture is deeper.

$$\text{Entropy } S = -\sum_x \sum_y P(x,y) \log P(x,y) \tag{4}$$

Entropy measures image texture randomness, when the space co-occurrence matrix for all values is equal, it achieved the minimum value; on the other hand, if the value of co-occurrence matrix is very uneven, its value is greater. Therefore, the maximum entropy implied by the image gray distribution is random.

$$\text{Inverse difference } H = \sum_x \sum_y \frac{1}{1+(x-y)^2} P(x,y) \tag{5}$$

It measures local changes in image texture number. Its value in large is illustrated that image texture between the different regions of the lack of change and partial very evenly. Here $p(x,y)$ is the gray-level value at the coordinate (x,y).

3.1 Distance Metric for Similarity Measure

In conventional image retrieval technique, Euclidean distance is used to find the similarity between the query image and image database. Similarity score is used to find the best match of query image from the database image. The distance metric gives minimum distance between the query shape and its nearest shape in the database is the best metric. For better classification, the maximum intra-class distance should be less than the minimum of the inter-class distances. We assume P and Q represent the feature vectors for database image and query image respectively in each distance metric. The present work evaluates and compares the CBIR performance for computing distance d (P, Q) using the following distance metrics:

3.1.1 Euclidean L_2 Distance

Euclid stated that the shortest distance between two points on a plane is a straight line and is known as Euclidean distance. Euclidean distance metric as in equation (6) was often called Pythagorean metric since it is derived from Pythagorean Theorem. Euclidean distance metric is defined for p=2. In Euclidean distance metric difference of each feature of query and database image is squared which increases the divergence between the query and database image.

$$d_{Euc}(P,Q) = \sqrt{\sum_{j=1}^{N} | P_j - Q_j |^2} \tag{6}$$

4 Experimental Results and Analysis

In this paper, the system was tested with more than one sample database to obtain a more extensive description of its positive and negative properties. The first test was conducted by selecting sketch image as query images from the Database, the system was tested for top 20-retrieved images; the database consists of 20 images with 5 versions of rotation of each image, and the results have been shown for two different dinosaur sketch mages in Figure (2) and (3). For both the dinosaurs Retrieval efficiency achieved about 95% and Precision also 95%.

The second test was conducted on dataset contains 1000 images from Wang Database of images, divided into 10 categories, each category has 100 images. By selecting query images from the Database, the system was tested for top 20-retrieved images; and the results have been shown for bus and flower images in Figure (4) and (5). Low retrieval efficiency has been achieved; hence, this process can be used as coarse level in hierarchical CBIR that reduces the database size from very large set to a small one. The third test was conducted on dataset contains 15 images from Scene categories Database and 10 images from Wang Database with 5 versions of rotation of each image, and the results have been shown for two different sketch images in Figure (6) and (7). This process also can be used as coarse level in hierarchical CBIR that reduces the database size from very large set to a small one.

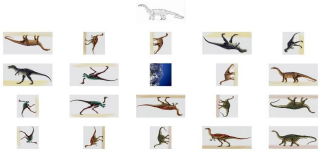


Fig. 2. Shows Top 20 retrieved images based on Dinosaur Sketch image as query image

Fig. 3. Shows Top 20 retrieved images based on another Dinosaur Sketch image as query image

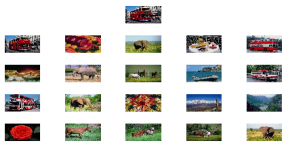


Fig. 4. Shows Top 20 retrieved images based on bus image as query image

Fig. 5. Shows Top 20 retrieved images based on flower image as query image

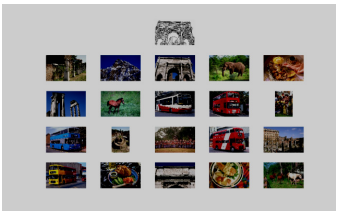


Fig. 6. ShowsTop20 retrieved images based on sketch image as query image

Fig. 7. Shows Top 20 retrieved images based on another sketch image as query image

5 Conclusion

We have proposed system architecture for the Content Based Image Retrieval by Gray level Co-occurrence Matrix (GLCM) derived four image features. The retrieval performance of the proposed method is showed in Figures (2) and (3). For both the dinosaurs Retrieval efficiency achieved about 95% and Precision also 95%. In this work, it is observed that the proposed method achieved low retrieval performance over these four image features for sketch based and color dominant images showed in Figure (4-7). The future work will focus on improved retrieval performance of sketch and color dominant images by exploring additional image features. Further, a research is in progress to improve the method aiming to increase the retrieval rate.

References

[1] Eitz, M., Hildebrand, K., Boubekeur, T., Alexa, M.: An evaluation of descriptors for large-scale image retrieval from sketched feature lines. Computers and Graphics 34, 482–498 (2010)

[2] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Hiang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by image and video content: the QBIC system. IEEE Computer 28, 23–32 (2002)

[3] Györök, G.: Embedded hybrid controller with programmable analog circuit. In: IEEE 14th International Conference on Intelligent Systems, pp. 59.1–59.4 (May 2010)

[4] Hu, R., Barnard, M., Collomosse, J.: Gradient _eld descriptor for sketch based image retrieval and localization. In: International Conference on Image Processing, pp. 1–4 (2010)

[5] Jain, A.K., Lee, J.E., Jin, R.: Sketch to photo matching: a feature-based approach. In: Proc. SPIE, Biometric Technology for Human Identification VII, vol. 7667, pp. 766702–766702 (2010)

[6] Jain, A.K., Lee, J.E., Jin, R., Gregg, N.: Graffiti-ID: matching retrieval of graffiti images. In: ACM MM, MiFor 2009, pp. 1–6 (2009)

[7] Jain, A.K., Lee, J.E., Jin, R., Gregg, N.: Content based image retrieval: an application to tattoo images. In: IEEE International Conference on Image Processing, pp. 2745–2748 (November 2009)

[8] Liu, Y., Dellaert, F.: A classification based similarity metric for 3D image retrieval. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 800–805 (June 1998)

[9] Lowe, D.G.: Object Recognition from Local Scale-Invariant Features. In: IEEE International Conference on Computer Vision, vol. 2, p. 1150 (1999)

[10] Smith, J.R., Chang, S.F.: VisualSEEK: a fully automated content based image query system. In: ACM Multimedia 1996, pp. 97–98 (1996)
[11] Szántó, B., et al.: Sketch4Match–Content-based Image Retrieval System Using Sketches. In: 9th IEEE International Symposium on Applied Machine Intelligence and Informatics, January 27-29 (2011)
[12] Narayana, M., Sandeep, V.M., Kulkarni, S.: Skeleton based Signatures for Content based Image Retrieval. International Journal of Computer Applications (IJCA) 23(7), 29–34 (2011)
[13] Narayana, M., Sandeep, V.M., Kulkarni, S.: Skeletal Distance Mapped Functional Features for Improved CBIR. International Congress for Global Science and Technology (ICGST)-GVIP Journal 11(3), 47–56 (2011)

Component Based Software Development Using Component Oriented Programming

Ruchi Shukla[1] and T. Marwala[2]

[1] Department of Electrical and Electronic Engineering Science,
University of Johannesburg,
Johannesburg,
South Africa
ruchishuklamtech@gmail.com
[2] Fac. of Engineering and Built Environment,
University of Johannesburg
tmarwala@uj.ac.za

Abstract. Software industries today are striving for techniques to improve the software developer's productivity, software quality and flexibility within the constraint of minimum time and cost. Component based software development (CBSD) is proving more suitable for the evolving environment of software industry. This paper demonstrates a sample application of component-oriented programming concepts for CBSD. Some of the potential risks and challenges in CBSD are also presented.

1 Introduction

The rapid growth in software and IT industry is leading to new software development paradigms, demanding faster delivery of software. Component based software (CBS) has recently started receiving attention among vendors, developers and IT organizations. There is a definite shift from structured programming written software for mainframe systems to object-oriented UML designed Smalltalk/C++ written client server software, to component based ADL designed C++/Java written N-tier distributed systems.

Reusability is the key issue today for building software systems quickly and reliably. The software system should be able to cope with complexity and adaptable to changing requirements by adding, removing and replacing the software components. Due to this component oriented programming has become the most popular programming technique [10]. With the emergence of component based software engineering (CBSE), component based software development (CBSD) has become an inevitable paradigm leading to less manpower/cost, increased quality, productivity and flexibility, reduced time to market, better usability and standardization.

According to Heineman and Councill: "A software component is a software element that conforms to a component model and can be independently deployed and

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1125–1133.
springerlink.com

composed without modification according to a component standard". A more widely accepted definition by Szyperski is: "A software component is a unit of composition with contractually specified interfaces and explicit context dependencies only. It can be deployed independently and is subject to composition by third parties" [8].

The interface of a component consists of the specifications of its provided and required services. To specify these dependencies precisely, it is necessary to match the required services to the corresponding provided services. Component model defines how components can be constructed, assembled and deployed [12]. The current state of component models usage justifying the need for a component model selection framework was presented in [2]. However, the question arises as to how these CBS are developed for reusability and what are the metrics taken into consideration. Further, the success of CBSD using third-party components mainly depends on the selection of a suitable component for the intended application [3].

The rest of the paper is structured as follows: Section 2 and 3 present a brief overview of the CBSD process and the Component frameworks respectively. Section 4 demonstrates a basic example of CBSD employing component-oriented programming concept. Section 5 presents some critical risks and key challenges in CBSD while Section 6 concludes the work.

2 Component Based Software Development Process

The CBSD approach uses various similar components identified in various software systems from Commercial-off-the-shelf (COTS) components for large-scale software reuse. CBSD consists of the following major activities [5]:

(1) requirements analysis,
(2) software architecture selection and creation,
(3) component selection,
(4) integration, and
(5) component-based system testing.

The development cycle of a component-based system is different from the traditional (waterfall, spiral, iterative and prototype based) models. Figure 1 shows a comparison between the traditional waterfall model and the modern CBD process. The requirements gathering and design in the waterfall process model corresponds to finding and selecting components. Similarly, the implementation, test and release in the waterfall model are equivalent to creating, adapting and deploying the components, and maintenance corresponds to replacing the components [5].

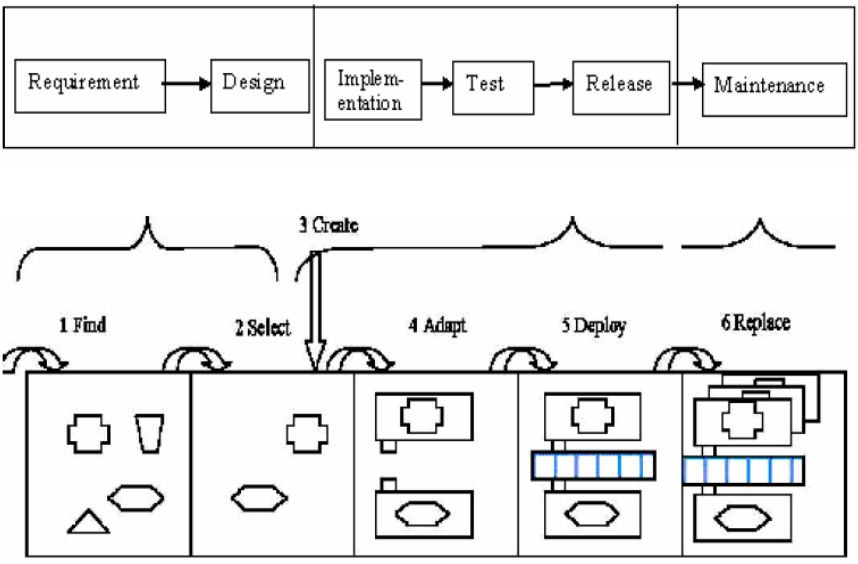


Fig. 1. Comparison of Waterfall Model Cycle With Component Based Development [5]

Most of the research so far has focused on the development and use of components within the following two development paradigms

1. Rapid application driven (RAD) paradigm where visual tools are used to create user interface and associate components with elements identified in interface, suitable for small to medium sized systems [13].

2. Object-oriented analysis and design (OOAD) paradigm where development takes place based on a conventional software life cycle and is oriented to the tasks and relevant objects, including their interactions, generalization and composition.

Here, we represent the OOAD approach from three aspects, i.e. functional, behavioural and structural, corresponding to use case, communication diagram and class diagram respectively. A component includes several use cases. A communication diagram is used to obtain the object usages and depict the dynamic behaviour of each use case. A set of participated classes are also specified by the communication diagram. We can extract the structural relationships among the objects from the class diagram. The relationship between classes and components is shown in Figure 2.

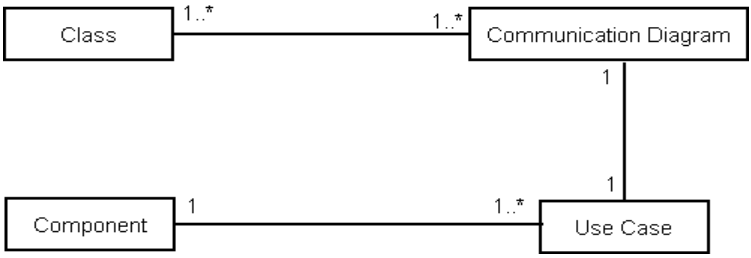


Fig. 2. Relationship Between Classes And Components

The process of component identification begins by clustering related objects and making them reusable. Each clustered object identified from clustering object approach are relating to their corresponding classes for allocating candidate components. Finally, in the last phase, we can identify reusable components from refining the candidate component [7, 9, 14].

3 Component Frameworks

A framework is a set of constraints on components and their interactions. Three standardized component frameworks are: CORBA, COM/DCOM, JavaBeans/EJB. The distributed version of COM is the DCOM. In the year 2002 .NET was released, which presents a platform-independent target for software development. It relies mainly on software component and the COP paradigm. EJB architecture is another component based architecture for developing and deploying component objects.

Any component can exhibit varying degree of distribution, modularity and independence of platform or language. Mapping of components is done on the following 3-dimensional space [4]:

1. Monolithic systems (0,0,0) – non-distributed, non modular, language dependent and platform dependent.
2. VB components (0,1,0) - neither distributed, nor language independent.
3. CORBA – distributed and language independent but the underlying components often remain platform dependent.

Java components are cross platform in scope. Wrapping a platform independent language such as Java with language independent middleware such as CORBA would yield a component worthy of (1,1,1) status [4].

4 Component Oriented Programming

Component-oriented programming (COP) enables programs to be constructed from reusable software components, following certain predefined standards including interface, versioning, deployment and connections [15]. COP as against OOP includes Polymorphism + Real late binding + Real and Enforced encapsulation + Interface inheritance + Binary reuse. COP allows various kinds of reuse including white-box reuse and black-box reuse. White-box reuse means that the source of a software component is made available and can be studied, reused, adapted, or modified. Black-box reuse is based on the principle of information hiding.

Example-1 demonstrates how to write a .NET serviced component that implements the IMessage interface and displays a message with "Hello Component" in it when the interface's ShowMessage( ) method is called [15].

Example-1: A simple .Net component

```
using System;
using System.EnterpriseServices;
namespace MyNamespace
{
public interface IMessage
{
void ShowMessage( );
}
public class MyComponent:ServicedComponent,IMessage
{
public MyComponent( ) //Default Constructor
{
}
public void ShowMessage( )
{
Console.WriteLine("Hello Component! ");
}
}
}
```

4.1 Registering Assemblies

Before adding the serviced components to a COM+ application, we need to register their assembly with COM+ [11]. This can be done using the RegSvcs.exe command line utility. The code then has to be signed with a cryptographic key. Open the *AssemblyInfo.cs* file, and at the bottom, insert the full path to the key against the AssemblyKeyFile entry [6]. The step is shown below:

```
C:\MyNamespace\MyNamespace\bin\Debug>regsvcs MyNamespace.dll
Installed Assembly:
Assembly: C:\MyNamespace\MyNamespace\bin\Debug\MyNamespace.dll
Application: MyNamespace
TypeLib: C:\MyNamespace\MyNamespace\bin\Debug\MyNamespace.tlb
C:\MyNamespace\MyNamespace\bin\Debug>
```

Now, the component is installed. If we open Component Services (Start - Settings - Control Panel - Administrative Tools - Component Services) and navigate down through the tree to COM+ Applications, we can see our newly installed application Figure 3.

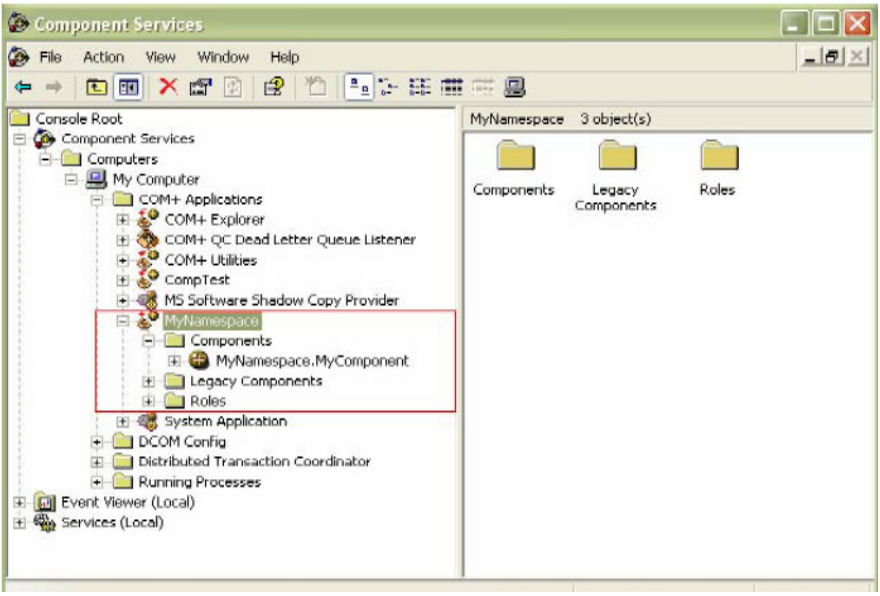


Fig. 3. Component Services

4.2 The Client Application

After building the serviced component library, we can create a client application. Then we can write the code to instantiate a new MyClient instance, and invoke the method ShowMessage ( ) [15]. The client code is shown below:

Example-2: A simple client application

```
using System;
using System.EnterpriseServices;
using MyNamespace;
namespace MyClientApplication
{
class MyClient
{
static void Main(string[ ] args)
{
MyComponent mycom = new MyComponent( );
mycom.ShowMessage( );
Console.ReadLine( );
}
}
}
```

The output is as shown in Figure 4.

Fig. 4. Client Application

5 Risks and Challenges in CBSD

Risks

- Changing nature of modern day software products and robustness in CBD.
- Reusable components are not designed and not coded from scratch.
- Development based on in-house, multi-origin reusable, 3rd party COTS software or open source software components.
- Few empirical studies that investigate how to use and customize COTS-based development processes for different project contexts [3].
- Component interfaces are defined by models with less information for functional testing.
- Instability in CBSD standards.
- Reliability of CBS.
- A usage of web based COTS leads to system security threats.
- Architectural risks of CBS and its agility with respect to software architecture, quality and maintenance.
- CB risk analysis should adopt similar principles of encapsulation and modularity as CBD methods (ISO, 2009a, b).
- Generalization of components for future reuse.

Challenges

- Challenges in component configuration during integration.
- Many complicated CBSD process models have been proposed, but no step-by-step guidance for implementing them is available [2].
- Limited knowledge about current industrial OTS selection practices.
- Effort estimation and fault identification.
- Relevant or new suite of software metrics and effort estimation and costing models for component assemblies are still not available.
- Challenges in multi-language component selection and integration.
- Lack of formal component selection methods and non availability of documentation impacting the component integration.
- Challenges of using semi-formal techniques like UML and formal techniques like VDM, Z and B in the early stages of component development.
- Challenges to check the presence of virus due to non availability of source code.
- Prediction of system behavior from component behavior.
- The maintenance process of CBS from system to product to component level.

Complexities involved in the development of an effort estimation tool for CBSD

- Relatively new concept, hence limited published metrics and available tools [1].
- Information about system artefacts, relationships and dependencies can be obscure, missing, or incorrect as a result of continued changes to the system.
- Changes must conform or be compatible with an existing architecture, design and code constraints.
- Multi-language, multi-platform software.
- Real time (dynamic/probabilistic/adaptive) components and cost drivers.

6 Conclusions

This paper demonstrates by means of a simple example, the use of component concepts and COP based software development method. The basic aim was to orient the programmers towards using an innovative software development approach for real life projects. This approach is a beginning of seamless support and better integration of the development tools, runtime, component services, and the component administration environment. However, the use of COTS components comes with its own challenges and risks which need to be analysed before arriving at a decision to use them for CBSD.

Acknowledgments. The financial support offered to RS by the FEBE of the University of Johannesburg, Johannesburg is gratefully acknowledged.

References

1. Aris, H., Salim, S.S.: Issues on the application of component oriented software development: Formulation of research areas. Inform. Tech. J., 1–7 (2008)
2. Aris, H., Salim, S.: State of component models usage: Justifying the need for a component model selection framework. I. Arab J. Inform. Tech. 8(3), 310–317 (2011)
3. Ayala, C., Hauge, O., Conradi, R., Francha, X., Li, J.: Selection of third party software in Off-the-shelf-based software development-An interview study with industrial practitioners. J. Syst. Softw. 84, 620–637 (2011)
4. Brereton, P., Budgen, D.: Component-based systems: A classification of issues. Comput. 33(11), 54–62 (2000)
5. Crnkovic, I.: Component-based software engineering – New challenges in software development. In: 25th International Information Technology Interfaces (ITI) Conference, Cavtat, Croatia (2003)
6. `http://www.codeproject.com/Articles/6736/A-Very-Simple-Persistent-Cache-in-a-COM-Component` (accessed March 31, 2012)
7. Kim, S.D., Chang, S.H.: A systematic method to identify software components. In: 11th Asia-Pacific Software Engineering Conference, Seoul, South Korea, pp. 538–545 (2004)
8. Lau, K.K., Wang, Z.: Software component models. Trans. Softw. Engg. 33(10), 709–724 (2007)

9. Lee, S.D., Yang, Y.J., Cho, E.S., Kim, S.D., Rhew, S.Y.: COMO: A UML based component development methodology. In: 6th Asia Pacific Software Engineering Conference, Takamatsu, Japan, pp. 54–61 (1999)
10. Liu, Y., Cunningham, H.C.: BoxScript: A component-oriented language for teaching. In: 43rd ACM Southeast Conference, Kennesaw, USA (2005)
11. Lowy, J.: Component services, http://ondotnet.com/pub/a/dotnet/excerpt/com_dotnet_ch10/index.html?page=3 (accessed March 31, 2012)
12. Mahmood, S., Lai, R., Kim, Y.S.: Survey of component-based software development. IET Softw. 1(2), 57–66 (2007)
13. Panfilis, S.D., Berre, A.J.: Open issues and concerns on Component Based Software Engineering. In: 9th International Workshop on Component-Oriented Programming, Oslo, Norway (2004)
14. Sook, M., Cho, E.S.: A component identification technique from object-oriented model. Springer, Heidelberg (2005)
15. Wang, A.J.A., Qian, K.: Component-oriented programming. John Wiley & Sons (2005)

Transformation of Artistic Form Text to Linear Form Text for OCR Systems

Vishwanath C. Kagawade, Vijayashree C.S., and Vasudev T.

P.E.T Research Centre, PES College of Engineering, Mandya, Karnataka, India-571401
Vishwanath.1312@gmail.com, {cs.vijayashree,banglivasu}@yahoo.com

Abstract. The existing Optical Character Readers (OCRs) are capable of reading linear form text and have limitations to read artistic and non-linear form text. This paper presents a technique to transform printed English artistic form text to linear form text in order to make an OCR to read the text. The technique starts with artistic form text as input and transforms the same to linear form. First, the characters in artistic text are segmented and extracted using Connected Component Analysis technique. Due to the intrinsic nature of artistic text, the extracted characters exhibit skew. In the next stage, such implicit skew in extracted characters is detected using Hough Transform and corrected. Further, skew corrected characters are concatenated to put in linear form. Experimental results of the proposed method show an average 80% of readability by OCR as efficiency.

Keywords: artistic text, linear text, OCR, Connected Component Analysis, skew detection, Hough Transform.

1 Introduction

A significant area in the field of Digital Image Processing is Document Image Analysis(DIA). DIA is very important in applications like document identification/ recognition, language identification, automatic reading from document etc. Many researchers are working on different problems on document images starting from image acquisition to image understanding (Nagabhushan, P., 2001; O'Gorman et.al., 1998). Processing activities in DIA can be divided into Pre-processing, Segmentation, Script Identification, Page Layout Analysis (PLA) and Classification, Character Recognition etc (Vasudev T. et.al, 2005), and these have lead into many vibrant research problems (O'Gorman et al, 1998). The results of the research on the above problems are gradually converging towards the generic solutions to major issues in DIA.

In spite of considerable research work in the area of DIA, a major issue which is not sufficiently addressed is, reading or extracting the contents of the text which appear in artistic form in a document (Vasudev T. et.al, 2007). Many documents, especially Certificates, Marks Cards, Sign boards, Logos, etc., have artistic text. In addition, many official seals/stampings used for document authentication purpose are also artistic in nature. The contents of such artistic text definitely have some valuable information that has to be processed. If such texts have to be processed by an Optical Character Reader (OCR), then proper pre-processing is required to make that text

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1135–1143.
springerlink.com

readable by OCR as the available OCRs are capable of reading only linear-form-text. Few such artistic texts in documents are, text appearing in triangular-form, arc-form, circular-form, wave-form, shadow-form, telescopic-form etc. Fig.1 shows few such examples of artistic form text. The contents of such text normally conveys the identity information like Company's Name, type of document, Brand Name, Event Name, Type of document etc., which is the main source for classification of the document. Therefore, text conversion from artistic form to linear form should be done before document processed with OCR. Hence, it is necessary to transform artistic-form-text into linear-form-text and make it suitable for reading by an OCR. This motivated us to make an attempt to transform artistic-form-text into linear-from-text.

In this research work, we propose a methodology to transform artistic-form-text to linear-form-text which is suitable for OCR processing. The model uses Connected Component Analysis Technique for character extraction from document image having artistic-form-text. The extracted characters poses implicit skew (Vasudev T. et.al, 2007) due to this the OCRs fail considerably to read such skewed characters. A modified Hough-Transform is used for detection of skew in characters. Finally, all the skew corrected characters are concatenated to present the text in linear form. The input image is assumed to be free from noise and contains only artistic text.

Fig. 1. Samples of artistic form texts

The rest of the paper is organized as follows: The sequence of stages in the proposed work along with character extraction using connected component analysis is discussed in section 2. The theory of Hough-Transform along with detection and correction of skew in characters is explained in section 3. Experimental results are illustrated in section 4. Conclusion is given in section 5.

2 Stages in the Proposed Work

The block diagram shown in Fig.2 indicates the sequence of different processing stages performed in the proposed system.

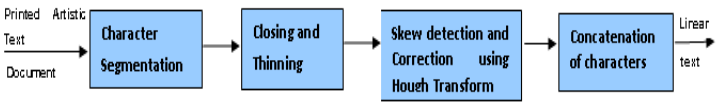


Fig. 2. Sequence of stages in the proposed work

The printed artistic text document image obtained at 300 dpi resolution is taken as input to the system. Character segmentation from artistic text is made through a series of image processing operations. To begin with, the image is normalized to 340 X 340

pixels and image is binarized to black and white, in order to make invariant to image size and colour. Then the characters are extracted from an image through connected component analysis. Labelling of characters are made using 8-connectivity logic (C.M.Velu et.al, 2010). Remove all objects containing less than the threshold pixels value which are treated as noise. Using the labels, the connected components (characters) are extracted. The extracted characters are then resized to 4*5 X 45* pixels.

The steps for Text Extraction from document image are given below in the form of algorithm:

Input: *Scanned Image Document containing artistic text*
Output: *A set of character images extracted from input image*

1. *Read the image*
2. *Convert input image to Gray scale and Gray scale to Black and White image*
3. *Scan the image document from left to right to label characters*
4. *Label and Count connected components which form the Candidate region*
5. *Remove all object containing less than threshold pixels value*
6. *Crop the character using Candidate region*
7. *Repeat step 6 until all characters are extracted from the input document image.*

Once all the labelled components are extracted, it is noticed that some additional blocks of residues are produced due to the limitation in cropping procedure. Such residues are addressed as noises in this work and are to be removed. It is evident from the experimental results that the connected components regions of such noises have less than 20 pixels (width to height ratio) i.e., connected component region Area < connected components regions MaxArea / 20. Considering 20 pixels as threshold, the regions which are less than 20 pixels are removed from the resultant image.

The result of the implemented methodology is shown through an artistic-form-text shown in Fig.3. Fig.4 shows extracted characters with additional blocks as noise. Fig.5. shows the output obtained after removing noise through defined threshold.

Fig. 3. Sample input image for Connected Component Analysis Technique

Fig. 4. Extracted Characters using Connected Component

Fig. 5. Extracted Characters after noise is removed

It is noticed that the extracted characters are not free from skew. The skew in characters are detected using modified Hough Transform and is explained in next section.

3 Hough-Transform

Hough-Transform is widely used in image analysis, computer vision and digital image processing. Hough-Transform technique is an approach preferred when the objective is to find lines or curves formed by groups of individual points on an image plane (Rajiv Kapoor et.al, 2004; H.K.Chethan, 2010). The method involves a transformation from an image plane to a parameter space. Consider the case in which line is the object of interest. Mathematically, the line is expressed as (Rafael C Gonzales & Richard E Woods, 2002)

$$\rho = X \cos\theta + Y \sin\theta. \quad (1)$$

There are two line parameters namely, the distance (ρ) and the angle (θ) which defines transformation space. Each coordinate (x, y) of ON pixel in the image plane is mapped onto the locations in the transformed plane for all possible straight lines. For all possible values of ρ and θ the transformations intersect at the same point on the transformed plane when multiple points are collinear. Therefore, the point (ρ, θ), which has the greatest accumulation of mapped points, indicates lines with these parameters. In practice, due to discrimination error and noise, points mapped will not be exactly collinear. Thus the points do not map on to exactly the same location on the transformed plane. For connected lines or positions of lines, computations can be reduced greatly by considering not all (ρ, θ) points but only those (ρ, θ) points that are in one orientation as indicated by the angle.

Before applying Hough Transform on the skewed characters, morphological 'closing' and 'thinning' operations are performed for better results and Hough Transform is applied on the thinned character image. The image is divided into two parts, in order to get better resolution in angle estimation for each half of the image. Perform the Hough Transformation to determine the position of the high pixel concentrations from both the parts, which indicates the skew angle. The Hough-Transform maps the individual pixels from the image domain into the parameter domain i.e. each coordinate (x, y) of the image plane is mapped onto the locations in the transformed plane for all possible straight lines. For all possible values ρ and θ the transformations intersect at the same point on the transformed plane when multiple points are collinear. Therefore, the points (ρ, θ), which has the greatest pixel concentrations of mapped points, indicates lines with parameters. The position of high pixel concentrations of the both the parts of image indicate the skew angle. Use skew angle to rotate the character and align it in the horizontal manner. Finally all the skew corrected characters are concatenated to form the linear form text.

The output of skew detection and correction on the characters extracted in previous section is shown in Fig.6.

Fig. 6. Skew corrected characters

The complete steps for transformation of artistic-from-text to linear-form-text, is given below in the form of algorithm:

Input: *Printed Document Image containing artistic form text*
Output: *Image containing linear form text*

1. *Read the image*
2. *Extract the character from the document image using Connected Component Technique*
3. *Perform the following steps for each extracted character*
   a. *Apply region based Heuristic Filtering to remove noise in extracted character*
   b. *Perform morphological 'closing' and 'thinning' operations*
   c. *Divide the image into two parts and perform the Hough Transformation on both the parts to determine skew angle*
   d. *Using skew angle rotate the character and align it in the horizontal manner*
4. *Repeat the step 2 and 3 until all the characters are extracted*
5. *Concatenate all the output characters of step 3 to get resultant linear form text.*

The Hough-Transform under/over estimates the skew angle when character has large variation in intensity value across the border and high density concentrations at one side of the character. The Hough-Transform parameter spaces peak detection is a problem for cluster detection and there is no skew estimation when characters have uniform spatial neighbourhood correlation and similar intensity distribution.

4 Experimental Results

The proposed methods have been implemented in the MATLAB R2009a. The different documents from Newspapers, Journals, Text Books, Magazines, Advertisement Articles, Pamphlets, and the artistic text document created from paint/word are considered. The experiments have been conducted on more than 150 image samples. Figs.7-15 shows some sample input texts and corresponding results. Within each figure first row indicates the input artistic form text, second row indicates extracted characters using Connected Component, and third row indicates linear form text.

Experimental results illustrated from Figs.7-15 indicate that the proposed approach considerably transforms artistic form text to linear form text which are better suitable for OCRs. The impact of implementation of Transformation model on characters is to increase the readability of OCR. Analysis of readability by an OCR before transformation, after character extraction and after skew correction is performed with respect to English text using the OCR "Readiris Pro 9" (http://www.irislink.com). Few instances from the worst case to the best case are considered and analysis of readability by OCR at three stages is tabulated in Table 1. The Table 1 indicates, an artistic form text, given as input to OCR, it recognizes the text as picture and the readability of text by OCR is obviously 0%. This demands that most of OCRs do require linear text for reading.

Fig. 7.

Fig. 8.

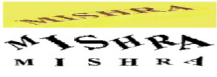


Fig. 9.

Fig. 10.

Fig. 11.

Fig. 12.

Fig. 13.

Fig. 14.

Fig. 15.

Table 1. OCR's readability of artistic form text

Input artistic form text	OCR Recognition	Readability
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture

The extracted characters using Connected Component analysis technique, introduces readability for OCR to some extent as show in Table 2, but still many input samples recognized as picture and broken characters because of skew in characters. Table 3 shows that OCR readability considerably increases after character skew is corrected. Table 4 indicates an average readability of OCR is 0-55% in case of extracted characters and also indicates the skew correction introduces readability for OCR upto 80%. The experimental results are illustrated in Fig.16.

Table 2. OCR's readability after Character Extraction use CC (Prior to skew correction)

Extracted Text	OCR Recognition	Readability
		0.0 % Recognized as picture
		0 .0 % Recognized as picture
	COODU't-'\	57.14 % (4/7) Recognition
		0.0 % Recognized as picture
	ICMAT2ItI"1MBAIWlCl\	46.67 %(7/15) Recognition
		0.0 % Recognized as picture
	C1qSS~CPOLG	72.7% (8/11) Recognition
		0.0 % Recognized as picture
		0.0 % Recognized as picture
		0.0 % Recognized as picture

Table 4 indicates that average readability of OCR of artistic form text is 0% indicating that OCRs are not capable of reading text other than linear form. After extracting the characters and placing the characters linearly without skew correction shows some improvement in readability. Once the skew is corrected in the extracted characters, OCR performance enhances to read at an average readability of 80%.

Table 3. OCR's readability after Character skew correction

Skew corrected text	OCR Recognition	Readability
	.t'UCCES.t'	5 / 7 → 71.4 %
	nR"IFTF"lIR	6/ 9 → 66.6 %
	GOODD~~	5 / 7 → 71.4 %
	AAlpCilteeroll	8 / 11 → 72.7 %
	KMI\T10"11MBJIIMCI\	11 / 15 → 73.3 %
	SITESSAlE	8 / 9 → 88.9 %
	Cl~.sS.ICPlO	10 /11 → 90.9 %
	SANGEETHt>,	8 / 9 → 88.9 %
	AIVNI~lERSf\R'	7 / 11 → 63.6 %
	MISHR<1	5 / 6 → 83.3 %

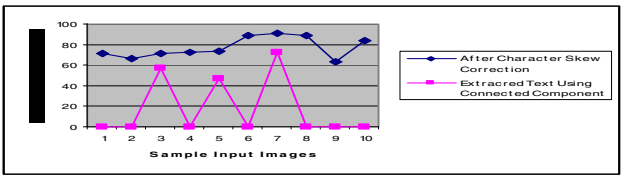


Fig. 16. OCR Readability Analysis for different Image Samples

Table 4. Readability analysis

Input text	Readability range	Average readability
Artistic form Text	0.00%	0.00%
Extracted Characters text with skew	00%-57%	5.70%
After Character skew correction	65% - 91%	80%

5 Conclusion

In this paper, a method to transform artistic form text to linear form text of English Printed Text is presented. The experimental results exhibit an average readability of 80% by OCR. However, the approach has certain limitations in estimating the skew angle exactly in some cases and also changes the size of some of the character after skew correction. The proposed method is not suitable when document text appears in circle, circle with multiple lines of linear text, semi-arc with linear text, triangular form, and text with complex background. There is much scope for further research in this work especially to investigate a better skew estimation model without changing the size of the character.

References

1. Twogood, R.E., Graham Sommer, F.: Digital Image Processing. IEEE Transactions on Nuclear Science NS-29(3) (June 1982)
2. Vasudev, T., Hemanthakumar, G.H., Nagabhushan, P.: Transformation of arc-form-text linear-form-text- suitable for OCR. Science Direct, Pattern Recognition Letters 28, 2343–2351 (2007)
3. O'Gorman, L., Rangachar, K.: Executive briefing: Document image analysis. IEEE Computer Society Press (1998)
4. Pal, U., Mitra, M., Choudari, B.B.: Multi-Skew Detection of Indian Script Documents. In: Proc. Int. Conf. on Document Analysis and Recognition, ICDAR 2001 (2001)
5. Saragiotis, P., Papamarkos, N.: Local Skew Correction in Documents. International Journal of Pattern Recognition and Artificial Intelligence 22(4), 691–710 (2008)
6. Nagabhushan, P., Anagadi, S.A., et al.: Geometric Model and Projection Based Algorithms for Tilt Correction and Extraction of Ascenders/Descenders for Cursive Word Recognition. In: IEE-ICSCN, pp. 488–491 (February 2007)
7. Vasudev, T., Hemanthakumar, G.H., Nagabhushan, P.: Segmentation of characters in an arc-form text. In: Proc. 7th Int. Conf. on Cognitive Systems (ICCS 2005), CD version, India (2005)
8. Chethan, H.K., Hemantha Kumar, G.: Graphics Separation and Skew Correction of Mobile Captured Documents and Comparative analysis with Existing Methods. International Journal of Computer Science and Applications 7(3) (September 2010)
9. Yamaguchi, T., Nakano, Y., et al.: Digit Classification on Signboards for Telephone Number Recognition. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003). IEEE (2003), 0-7695-1960-1/03
10. Jipeng, T., Hemantha Kumar, G., Chethan, H.K.: Skew Correction for Chinese Character using Hough Transform. International Journal of Computer Applications (0975–8887) 22(2) (May 2011)
11. Nandini, N., Srikanta Murthy, K., Hemantha Kumar, G.: Estimation of Skew Angle in Binary Document Images Using Hough Transform. World Academy of Science, Engineering and Technology 42 (2008)
12. Kapoor, R., Bagai, D., Kamal, T.S.: A new algorithm for skew detection and correction. Science Direct, Pattern Recognition Letters 25, 1215–1229 (2004)
13. Velu, C.M., Vivekandan, P.: Automated letter sorting for Indian Postal Address Recognition System based on PIN codes. Journal of Internet and Information System 1(1), 6–15 (2010)
14. Gonzales, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Pearson Education Publication (2002)

A Fuzzy Sectional Real-Time Scheduling Algorithm Based on System Load

Annappa B.

annappa@nitk.ac.in

Abstract. Earliest Deadline First (EDF) Algorithm is one of the most widely known dynamic real-time task scheduling algorithms. However, when a real-time system is overloaded, experiments and analysis have proved that EDF algorithm is ineffective. Considering the algorithm's instability during the practical task executing environment in an overloaded state, it is necessary to apply a few decision making techniques to ensure a good overall performance. In this paper, we propose a dynamic sectional real-time scheduling algorithm called Fuzzy Sectional Scheduling (FSS), which identifies the system load and employs suitable scheduling techniques to improve overall performance. The simulation results show that the Fuzzy Sectional Scheduling Algorithm could improve the real-time system performance to a considerably greater extent compared to the classical algorithms such as EDF, HVF (Highest Value First) and HDF (Highest Density First) algorithms; under all workload conditions.

Keywords: Real-time system, Dynamic priority scheduling, Fuzzy logic, Deadline Missing Ratio.

1 Introduction

In a real-time system, all tasks are characterized by their deadlines. A deadline may be defined as the latest time by which the task needs to be completed. In hard real-time systems [14, 15], it is necessary for the tasks to complete before the specified deadline; otherwise the tasks are worthless and can cause catastrophic results. Whereas, a soft-real time system [15] can have more relaxed set of constraints. In such systems, meeting all the deadlines is not the only consideration. An occasional missed deadline can be considered tolerable; and more importance is given to the throughput of the system. Our discussions in this paper deal with preemptive dynamic scheduling algorithms applicable to hard real time systems.

There are a number of optimal scheduling algorithms that can guarantee the feasibility of a schedule. Most of these classical algorithms, such as EDF, fail to adapt to dynamically varying systems, specifically if there are considerable changes in the workload, which is the case with complex real-time systems. Hence, it may be safe to assume that these classical algorithms [7] work ideally in relatively smaller systems [2]. According to the work carried out in [8], the real-time systems need to adapt to the changes dynamically and quickly. Hence, we make use of a Fuzzy Inference system to adapt to the various changes in system.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1145–1153.
springerlink.com

When a real-time system is overloaded, it is practically impossible for a feasible schedule [9,15], i.e. , all tasks cannot be completed before their deadline. Some tasks need to be rejected [2]. The objective of the algorithm we present in this paper is to schedule the most important tasks in overloaded situations [8]. In order to successful implement this in our algorithm, we need to add Importance Value [1,2] to the set of parameters of each task. Considering the Importance value of a task in an overloaded system will also help avoid the so called Domino effect [1]. This effect usually occurs when a task that missed the deadline causes subsequent and more important tasks to miss their deadlines resulting in a catastrophic state. Locke proved that EDF is prone to the domino effect in overloaded conditions in [4]. To avoid the domino effect, our algorithm needs to plan the task schedules at runtime to adapt to the changes in overload [2]. Few of these dynamic scheduling algorithms include EDF (Earliest Deadline First)[9], HVF(Highest Value First) [1,2] and HDF(Highest Density First) [1,3] . These algorithms have been defined and discussed in [1,2]. There are many parameters that could be considered for a particular real-time system [16]. The challenge is to choose the right parameters to determine and analyse various system properties.

Over the last decade, Fuzzy logic has been widely applied to the concepts of operating systems [12,13] as well as those of real-time systems, especially in the scenario of scheduling algorithms. In this paper, we propose a Fuzzy Sectional Scheduling algorithm to tackle the problem of the domino effect and to increase efficiency in system overload conditions. The algorithm uses Fuzzy Logic principles [10, 11] to compute the Current Load State of the real-time system. Xian-Bo He in [5] discusses the steps involved in a fuzzy inference mechanism in the real-time scheduling scenario. Section 2 describes the task model of the FSS algorithm. The scheduling policy of the proposed algorithm is mentioned in Section 3. Section 4 describes the properties of the proposed Value-Density algorithm which is a part of the FSS algorithm. The experimentation and observations have been presented in Section 5. We conclude the paper in Section 6 and identify future scope of this research project.

2 Fuzzy Sectional Scheduling Algorithm

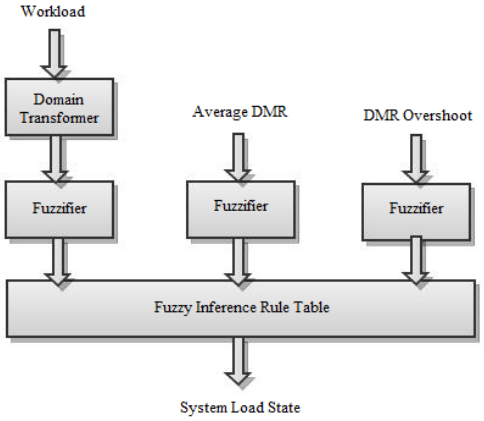


Fig. 1. The process of identifying Load State of the System using Fuzzy Inference

A. Fuzzification [10,11] and the discrete fuzzy set domain

In our FSS algorithm, we set the fuzzy set domain, U as follows [5, 6]:

$$U=\{0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}.$$

This is based on the fuzzy inference and the simulation experiments carried out.

The three real-time system metrics [16] we fuzzify to make a decision are as follows:

1) **Workload of the system**, ***W***
2) **The Average deadline miss-ratio,** M_a, a traditional metric for real-time systems, is defined as the total number of deadline misses divided by the total number of task instances throughout the run-time
3) **Deadline Miss Ratio Overshoot,** M_o, represents the worst-case transient performance of a system in response to the load profile [16].

The fuzzy partition sets of Workload of the system, W= {Normal, High, Very High), and figure 1 is its membership function. The fuzzy partition sets of the Average Deadline Miss ratio, M_a= {Low, Medium, High}, and figure 2 is its membership function. The fuzzy partition sets of the Deadline Miss ratio Overshoot, M_o= {Low, Medium, High}, and figure 3 is its membership function. The values for membership were selected after extensive experimentation with various test cases.

The fuzzy partition sets of final System Load, L = {Normal, Overload}.

The reason why we need to consider the Average Deadline Miss Ratio (M_a) and Deadline Miss Ratio Overshoot (M_o) to decide the Load State of the system is that these metrics help us analyse the current situation in the system and hence allow us to decide if the system is ending up what the Domino Effect. For example, a very high DMR overshoot may indicate that a considerable number of recent jobs have missed deadlines, or in other words the dominoes have started to fall.

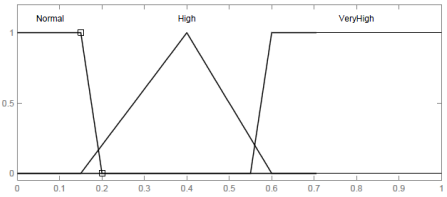


Fig. 2. Membership Function of Workload ,W

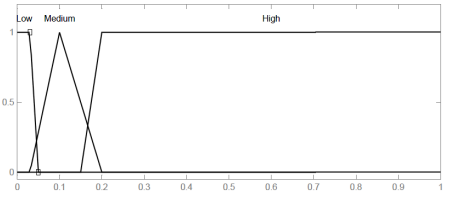


Fig. 3. Membership Function of Average Deadline Miss Ratio, M_a

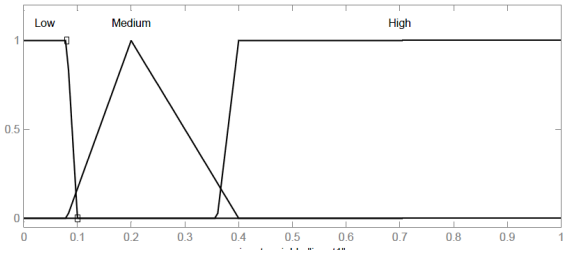


Fig. 4. Membership function of Deadline Miss Ratio Overshoot

To fuzzify the above mentioned metrics, we transform them into the fuzzy sets according to the following steps:

1) Domain transformation:

$$Relative\_workload,\ W_r = W / Max\_workload$$

where, Max_workload is the Maximum Workload that is allowed to be processed by the system.

The other metrics, namely, Average deadline Miss ratio (M_a) and Deadline Miss ratio Overshoot (M_o) already belong to the fuzzy set domain U.

2) Fuzzification:

We employ a linear proportion method as in [6] to fuzzify the metrics described above.

B. Fuzzy Inference

Table 1 has the fuzzy inference rules to identify the final system load state. These rules were formulated based on experimentation.

It can be observed from Table 1 that, if the fuzzy inference rules were not applied to the system and considering the situation when there is a quick increase the no of deadlines missed, the algorithm would ignore a possible domino effect and consider the system to be in a normal state. The Fuzzy Inference rule no 5 and 6 counter this situation and avoid the possible domino effect.

C. Defuzzification

To defuzzify a certain metric, we adopt the Similarity Nearness Degree (SND) [5] to decide upon the fuzzy set in the fuzzy partition corresponding to a special fuzzy set input.

After obtaining the fuzzy sets of the above mentioned metrics, we match them with the items of fuzzy set standard pattern by computing the Similarity Nearness Degree (SND) to get the corresponding inputs to the fuzzy inference rule table. Hence, by looking into the table for the corresponding fuzzy values, we can identify the current system load state.

3 Scheduling Policy of the FSS Algorithm

Once the Load State of the system is determined by Fuzzy inference, the FSS algorithm employs the respective algorithm corresponding to a Load State:

3.1 Normal Load

During Normal load conditions we employ the traditional EDF algorithm since its theoretically always produces a feasible schedule.

3.2 Overload

We employ the Value-Density Algorithm, which is as described in the following section.

4 The Value-Density Scheduling Algorithm

We propose a Value-Density scheduling algorithm, where we combine the properties of the two traditional scheduling algorithms namely, Highest Value First (HVF) and Highest Density First (HDF). The algorithm is characterized by one parameter, α which is used to compute the priority, calculated to schedule the tasks in the execution queue. The priority P_{ri} of the algorithm is defined below.

$$P_{ri} = (1- \alpha)*V_n + \alpha*D_n$$

Where,
V_n – Normalised Importance Value,V of the task (used as priority in HVF algorithm)
D_n – Normalised Density,D of the task (used as priority in the HDF algorithm), essentially the ratio between the importance value and the remaining execution time left for that particular task [2].
α – The parameter used to determine the priority in the proposed algorithm

It is important to note that we have used the normalised values of V and D in the above equation to determine the priority; reason being the ranges of these values differ by a considerable margin and could give erroneous results.

The parameter α can decide the balance factor between the 2 algorithms. It can be observed that when α is 0, the algorithm is equivalent to HVF and when α is 1, the algorithm is equivalent to HDF. The parameter, α, can play a very important role to decide the priority, especially since this algorithm is employed in overloaded situations. It becomes crucial to choose the tasks which need to be scheduled first, and essentially rejecting the tasks that may not contribute to the overall throughput of the system. Hence, we need to critically evaluate the process of computing α. We apply fuzzy approximation techniques to compute the value of α.

We propose a Fuzzification function, F which is defined on the parameters Relative_Workload(W_r) and average deadlines missed ratio(M_a):

$$F(\beta,\gamma) = \gamma^{1-\beta}$$

Where,
$\beta = M_a$, Average deadlines missed ratio
$\gamma = W_r$, Relative_Workload

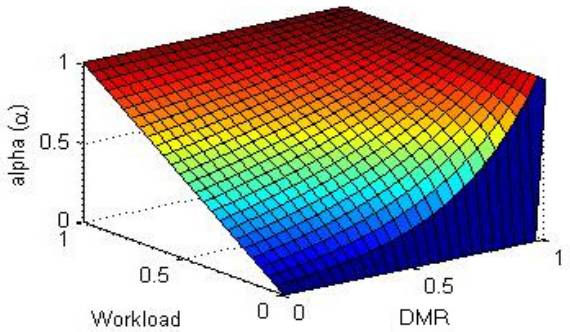


Fig. 5. The fuzzification function of α

We simply assign the value of the function F, computed using the above method, to α which in turn is used to compute the priority of the Value-Density algorithm which schedules the next important task for execution. We have not taken into consideration DMR Overshoot, M_o for the fuzzification of α, since the value of M_a is sufficient and it takes into consideration the amount of deadlines missed and M_o would not add any additional and useful properties to the equation for this particular task logically.

5 Performance Evaluation

Considering the implementation point of view, the value α for the Value-Density algorithm proposed was computed at frequent intervals such that system state was appropriately identified. And a quick sort algorithm would be ideal for the sorting process of ready tasks. The reason being, once α is computed, since the task queue would already be sorted for the old α, the quick-sort process would have much lesser comparisons and the time taken to sort the tasks would be considerably small.

In this paper, the performances of the scheduling algorithms are determined from the parameter, HVR (Hit Value Ratio) which is defined later in the section.

The following are the rules that were followed while simulating the algorithm [2]:

(1) In all the simulations, a task set contains 1000 tasks, J_i where i=1,2,3,…,1000.
(2) For each task, the worst case execution time C_i, was chosen as a random variable with a uniform distribution between 5 and 200 time units.
(3) Importance Value of a task is modelled as a uniform distribution from 1-100 which are divided into 5 criticalness levels from 1 to 5.

The results portrayed here are the average of 200 different simulations. In the model used for simulation, the workload, ρ changes from 0.5 to 4.5. In other words, the maximum workload allowed for the system to work is 4.5.

1) Hit Value Ratio (HVR)

It is defined as the ratio between the sum of all the significance values collected during the task set execution and the total value of all the tasks submitted to the system.

$$HVR = \frac{TA}{\sum_{i=1}^{n} Vi}$$

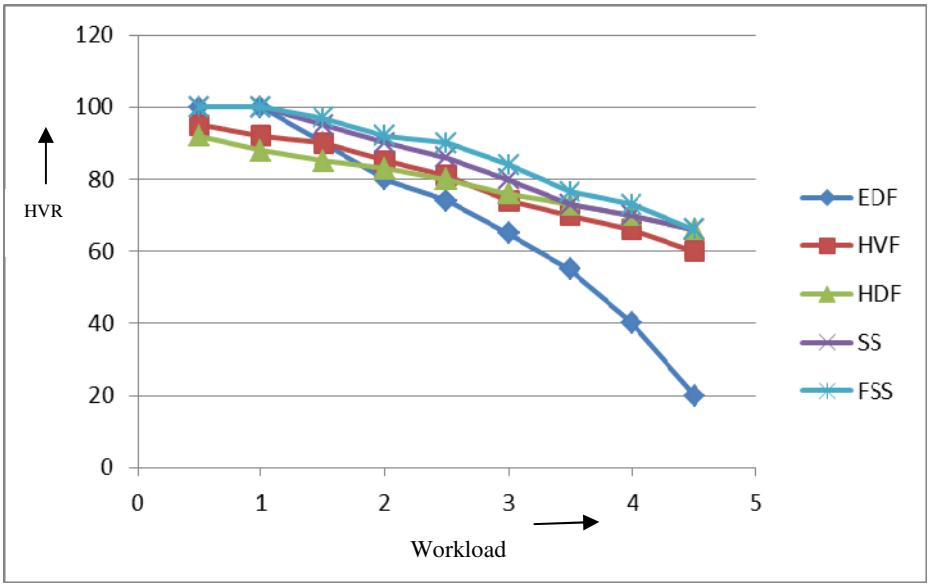


Fig. 5. Hit Value Ratio of the classical Algorithms

As it can be observed, EDF algorithm works ideally under Normal load conditions. But as the workload increases EDF degrades rapidly. The SS Algorithm proposed in [2] is compared to the FSS algorithm proposed in this paper and clearly the FSS scheduling performs better in all overload conditions. Since the SS algorithm discretely changes the scheduling algorithm, it works well in extreme conditions, i.e, when the system just enters a state of overload and when the system is in a serious overload condition. The properties of HVF and HDF algorithms are merely replicated by the SS algorithm during different workloads. The FSS algorithm dynamically computes the priority in all workload conditions after choosing the best properties of both the algorithms using fuzzy logic and hence provides better results in terms of Value. In extreme cases of workload, FSS algorithm will tend to the SS algorithm, as it can be gathered from the graph.

The FSS algorithm avoids the domino effect successfully and the proposed Value-Density algorithm, which is employed in overloaded conditions by the FSS algorithm, degrades gracefully and the hit value ratio is better than the traditional HVF and HDF algorithms since it incorporates properties of both these algorithms depending upon the system state.

6 Conclusions

We have presented a sectional real-time scheduling algorithm called the FSS which uses Fuzzy logic to determine the load on the system and employ suitable task scheduling algorithm for each case. We use the performance of these classical scheduling algorithms, namely EDF, HVF and HDF, as the basis to compare the performance of the FSS algorithm. Simulation results proved that the FSS algorithm performs better in an unstable system, especially under overload situations.

The FSS algorithm successfully encounters the Domino effect, which was the objective of the paper, and also significantly improves the value of the overall system's performance in overload conditions. In other words, the FSS algorithm executes the most important tasks in overloaded conditions and rejects the less important ones.

As a future scope for this particular research project, we can analyse other parameters that could suit better for the model and improvise on the inference system.

References

[1] Buttazzo, G., Spuri, M., Sensini, F.: Value vs. deadline scheduling in overload conditions. In: Proc. of the 16th IEEE Real-Time Systems Symp., vol. 12(8), pp. 90–99. IEEE Computer Society Press, Pisa (1995)

[2] Ding, W., Guo, R.: Design and Evaluation of Sectional Real-Time Scheduling Algorithms Based on System Load. In: Proc. of The 9th International Conference for Young Computer Scientists (2008), doi:10.1109/ICYCS.2008.208

[3] Jensen, E.D., Locke, C.D., Toduda, H.: A time-driven scheduling model for Real-Time operating systems. In: Proc. of the 6th IEEE Real-Time Systems Symp., vol. 15(3), pp. 112–122. IEEE Computer Society Press, San Diego (1985)

[4] Locke, C.D.: Best effort Decision Making for Real-time Scheduling. PhD thesis, Computer Science Department, Carnegie-Mellon University (1986)

[5] He, X.-B.: An Improved LLF Scheduling Algorithm Based on Fuzzy Inference in the Uncertain Environments. In: Proc. of the 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT (2010), doi:10.1109/ICCSIT.2010.5563720

[6] Litoiu, M., Tadei, R.: Real-Time Task Scheduling with Fuzzy Deadlines and and processing times. Fuzzy Set and Systems 117(1), 35–45 (2001)

[7] Buttazzo, G., Stankovic, J.: RED: A Robust Earliest Deadline Scheduling Algorithm. In: Proc. of 3rd International Workshop on Responsive Computing Systems, Austin (1993)

[8] Terrier, F., Rioux, L., Chen, Z.: Real time scheduling under uncertainty. In: Nakanishi, S. (ed.) Proc. of the 4th IEEE Int'l Conf. on Fuzzy Systems, vol. 3, pp. 1177–1184. IEEE Computer Society, Piscataway (1995)

[9] Haritsa, J.R., Livny, M., Carey, M.J.: Earliest Deadline Scheduling for Real-Time Database Systems. In: Proceedings of Real- Time Systems Symposium (December 1991)

[10] Litoiu, M., Tadei, R.: Real-Time task scheduling with fuzzy deadlines and processing times. Fuzzy Set and Systems 117(1), 35–45 (2001)

[11] Sabeghi, M., Naghibzadeh, M.: A Fuzzy Algorithm for Real-Time Scheduling of Soft Periodic Tasks. IJCSNS International Journal of Computer Science and Network Security 6(2A), 227–235 (2006)

[12] Kandel, A., Zhang, Y.-Q., Henne, M.: On use of fuzzy logic technology in operating systems. Fuzzy Sets and Systems 99, 241–251 (1996)
[13] Gupta, N., Abhinav, K.R., Annappa: Fuzzy File Management. In: Proc. of 3rd International Conference on Electronics Computer Technology (2011), doi:10.1109/ICECTECH.2011.5941594
[14] Biyabani, S.R., Stankovic, J.A., Ramamritham, K.: The integration of deadline and criticalness in hard real-time Scheduling. In: Proc. of the 9th IEEE Real-Time Systems Symp., vol. 18(7), pp. 152–160. IEEE Computer Society Press, Huntsville (1988)
[15] Jane, W.S., Liu, P.B.: Real-Time Systems, 8th edn. Pearson Education, Inc. (2009)
[16] Lu, C., Stankovic, J.A., Abdelzaher, T.F., Tao, G., Son, S.H., Marley, M.: Performance Specifications and Metrics for Adaptive Real-Time Systems. In: Proc. of the 21st IEEE Real-Time Systems Symposium (2000), doi:10.1109/REAL.2000.895992

An Intelligent and Robust Single Input Interval Type-2 Fuzzy Logic Controller for Ball and Beam System

Sumanta Kundu and M.J. Nigam

Indian Institute of Technology-Roorkee, Roorkee-247667, India
sumanta.kundu.ec@gmail.com
mkndnfec@ iitr.ernet.in

Abstract. The Ball and Beam system (BBS) is a nonlinear and unstable system which resembles with many real-time complicated systems. Providing an appropriate beam angle to give the stability to the ball on the beam in a specific position, is a challenging task for the control system researchers. In this paper a robust interval type-2 fuzzy logic controller (IT2FLC) is designed. The dimension of the rule base is reduced by using signed distance method. This signed distance method makes the IT2FLC to a Single input Interval Type-2 Fuzzy logic Controller (SIIT2FLC). The type-2 fuzzy sets resolve the problem of determining membership functions in type-1 fuzzy systems. The membership function of a type-2 fuzzy set is three dimensional, where third dimension is the value of membership function at each point on its two dimensional domain that is called its footprint of uncertainty (FOU). The ability of FOU to represent more uncertainties enables one to cover the input and output domains with less number of fuzzy sets. This SIIT2FLC gives the smooth 2-D control surface and robustness to the system. The simulation work is carried out in simulink environment of MATLAB (7.8.0) software. The simulation results are also validated in the real-time implementation of the BBS, designed by Googol Technology. Experimental results show that the performance of the proposed controller is better than the single input type-1 fuzzy logic controller in terms of transient and steady state response. The ability of handling uncertainty (robustness) of proposed controller has been checked by applying a disturbance signal to the position sensor's output and also by parameter variation of the BBS.

Keywords: Ball and Beam system, single input interval type-2 fuzzy logic controller, single input fuzzy logic controller, signed distance method, footprint of uncertainty, robustness.

1 Introduction

The BBS is one of the most enduringly popular and important laboratory models for testing different control techniques, as its open loop operation shows instability. The mechanical plant [2] in Fig. 1 consists of a base, a beam, a ball, a lever arm, a gear box, a support block, a motor and an embedded electrical power supply.

The ball can roll freely along the whole length of the beam. The beam is connected to the fixed support block at one end and to the movable lever arm at other end. The motion of the lever arm is controlled by the DC brush motor through the gear.

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1155–1162.
springerlink.com

The motor has built-in rotary optical incremental encoder that provides feedback information about current actual position of the motor shaft. There is a linear potentiometer sensor that senses current linear actual position of the ball on the beam. This measured position is fed back to the comparator to compute the error between the actual position and desired position. The main control job is to automatically regulate the position of the ball on the beam by changing the tilt angle of the beam. This is difficult control task because the ball does not stay in one place on the beam but moves along the beam with an acceleration that is proportional to the tilt angle of the beam. There are many research work has been done to control the ball in desired position on the beam. Some conventional techniques have been developed for BBS, such as Proportional plus Derivative (PD) controller [16]. There are number of advanced control techniques have been designed , such as state observer with state feedback [9], linear quadratic regulator [11], robust stabilization using time scaling and Lyapunov redesign [10], sliding mode controller [7], fuzzy controller [6], variable universe fuzzy controller [8], single input fuzzy logic controller[1], interval type-2 fuzzy neural network [15].

In this paper a single input interval type-2 fuzzy logic controller is designed for BBS and real time experiment has been carried out. As the type-1 fuzzy logic controller is designed based on human experience about the system input output, the type-2 fuzzy logic controller is also designed based on human experience. The type-2 fuzzy sets [13] resolve the problem of finding the membership functions in type-1 fuzzy system. The footprint of uncertainties (FOU) [3], [14] of the type-2 fuzzy membership function is capable of handling the uncertainties related to the position sensing and parameter variation. In this paper signed distance method is used to reduce the number of input to the type-2 fuzzy controller as well as the number of scaling factor.

2 Mathematical Model of Ball and Beam System

2.1 Physical Structure of BBS

Physical structure of BBS [2] is shown in Fig.1. The beam is supported by a support block at one end and by a lever arm at the other end. The lever arm is attached with a servo gear. This servo gear can make positive and negative angle by rotation in both direction. The tilt angle of the beam is controlled by the gear. Depending on the tilt angle and gravity, ball can freely roll along the beam. The actual position of the ball is measured by a linear potentiometer sensor which is attached with the beam.

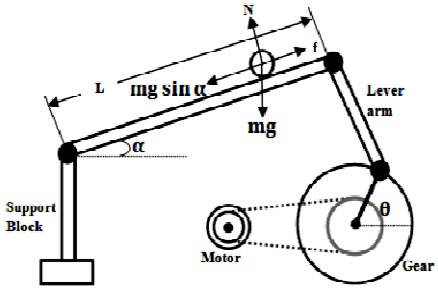


Fig. 1. Physical structure of BBS

2.1 Mathematical Model of BBS

Let the angle between the lines that connects the joint of the lever arm with the center of the gear and the horizontal line be; the distance between the center of the gear and joint be **d** and the length of the beam be **L**. Then the beam angle α can be expressed in terms of the rotation angle of the gear θ according to the following equation (1)

$$\alpha = \frac{d}{L}\theta \tag{1}$$

The angle θ is connected with the rotational angle of motor shaft trough the reduction gear ratio n=4.28. The ball on the beam is subjected to the gravity, inertial and centrifugal forces. The dynamic equation of the ball on the beam can be described by using Lagrange method:

$$(\frac{J}{R^2} + m)\ddot{r} + mg \sin\alpha - mr(\dot{\alpha})^2 = 0 \tag{2}$$

Where g is the gravitational acceleration, m is the mass of the ball; J is the ball moment of inertia; r is the position of the ball along the beam; R is the radius of the ball. In this paper we have assumed that the ball rolls without slipping and friction between the beam and ball is negligible. Our main interest is to keep the angle α close to zero. We can linearize the dynamic equation (2) with respect to α in the neighborhood of zero. Then we get the linear approximation of the system.

$$\ddot{r} = \frac{-mgd}{L(\frac{J}{R^2} + m)}\theta \tag{3}$$

This (3) is used to model the dynamic of the BBS. The values of the parameters are listed in the Table 1.

Table 1. Value Of The BBS Parameters

Name of Parameters	Value
m (kg)	0.028
R (m)	0.01
g (m/s^2)	-9.8
L (m)	0.4
d	0.04
J	1.12*10^{-6}

3 Design of SIIT2FLC

3.1 Interval Type-2 Fuzzy Logic System

Uncertainty affects the decision making. The concept of information is inherently associated with the concept of uncertainty [14]. The type-1 fuzzy sets are able to handle the uncertainty by the numbers in range [0,1]. However it is not reasonable to use an accurate membership function for something uncertain, so in this case we need

another type of fuzzy sets, those which are able to handle these uncertainties, the so called type-2 fuzzy sets. The amount of uncertainty related to system can be reduced by using type-2 fuzzy sets because these sets have better capability to handle linguistic uncertainty. The type-2 fuzzy set [13] can be described by Ã, is characterized by a type-2 membership function $\mu_{\tilde{A}}(x,u) \in [0,1]$ as follows

$$\int_{x\in X}\int_{\mu\in J_x} \mu_{\tilde{A}}(x,u)/(x,u) \tag{4}$$

Where $x \in X$ and $u \in J_x \subseteq [0,1]$. An interval fuzzy set $\tilde{A}_I$ and can be defined as:

$$\tilde{A}_I = \int_{x\in X}\int_{\mu\in J_x\subseteq[0,1]} 1 /(x,u) \tag{5}$$

The main characteristic of type-2 fuzzy set, which makes it better than type-1 fuzzy set, is its uncertainty. The footprint of uncertainty (FOU) [3] covers a bounded region that can be taken as a measure of scattering of the system input. The FOU is bounded by a lower membership function (LMF) and upper membership function (UMF). The FOU is shown in Fig. 2

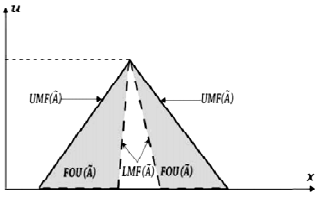


Fig. 2. FOU of a triangular MF

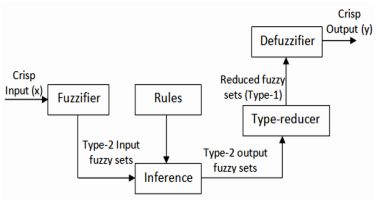


Fig. 3. Type-2 fuzzy logic controller

An interval type-2 fuzzy logic system [14] shown in Fig.4 contains fuzzifier, rules, inference engine, type reducer and defuzzifier. Up to the inference engine the functionality is same as type-1 fuzzy logic system. The inference engine combines the inferred rules and produces a type-2 fuzzy set output. The type reducer reduces the type-2 fuzzy set to a type-1 fuzzy set which is then converted to a crisp output by using defuzzification method. So there is a mapping existing between crisp inputs and crisp outputs of the IT2FLC. This relation can be expressed by y=f(x). Among many reduction techniques, centroid method is used here to convert the type-2 fuzzy sets to a type-1 fuzzy set and such sets are completely characterized by their left and right end points; hence computing the centroid of a type-2 fuzzy set only requires computing those two end points.

3.2 Signed Distance Method

For general type-2 fuzzy PD controller, we need two inputs, error (e) and rate of change of error ($\dot{e}$). Here e is the difference between desired position and actual position of ball. If e and $\dot{e}$ are consist of seven membership functions each, there will be 7^2 rules. So the numbers of rules are not too small and it requires large computational time as well as high performance processor. If we summarize the two input variables (e, $\dot{e}$) as a single input variable, then the number of rules will be 7 and controller will be single input type-2 fuzzy logic controller.

A powerful method is used to summarize the two inputs (e, $\dot{e}$), it is signed distance method [15], [16]. It is applicable, when the rule table of two input type-2 fuzzy logic controller is in skew-symmetric form, shown in Fig. 4. It is common for the rule table to have the same output membership in a diagonal direction. Each point on the particular diagonal line has a magnitude that is proportional to the distance from its main diagonal line (L_Z). For any combination of (e, $\dot{e}$), the output membership function will lie in any one of the diagonal line (L_{NL}, L_{NM}, L_{NS}, L_Z, L_{PS}, L_{PM}, L_{PL}). The main diagonal line (L_Z) can be represented by

$$\dot{e} + \lambda e = 0 \tag{6}$$

Here variable λ is the slope magnitude of the main diagonal line L_Z. The distance from any point (e. $\dot{e}$) to the main diagonal line can be written as:

$$d = \frac{\dot{e} + \lambda e}{\sqrt{1 + \lambda^2}} \tag{7}$$

Depending on the distance d, the new rule table can be constructed and given in Table 3. Rule table is one dimensional and contains only seven rules and confirms smooth 2-D control surface. Rule table for SIIT2FLC is given in Fig. 5.

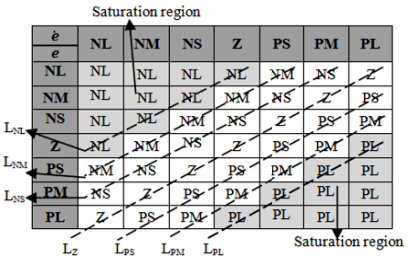


Fig. 4. IT2FLC RULE BASE

d	L_{NL}	L_{NM}	L_{NS}	L_Z	L_{PS}	L_{PM}	L_{PL}
u	NL	NM	NS	Z	PS	PM	PL

Fig. 5. SIIT2FLC RULE BASE

4 Design and Implementation of SIIT2FLC for BBS

Block diagram of complete BBS is given in Fig.6.

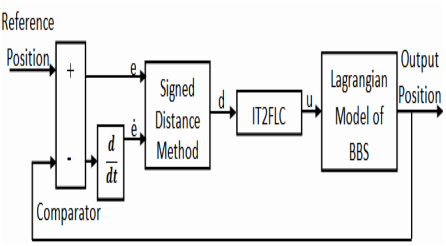


Fig. 6. Complete Block Diagram of BBS

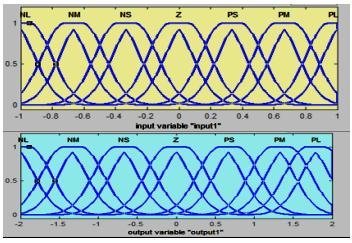


Fig. 7. Membership functions for input and output respectively

The input variable (d) and output variable (u) of the SIIT2FLC is divided into seven Gaussian membership functions, shown in Fig. 7. The universe of discourse of input and output sets are [-1, 1] and [-2, 2] respectively. Here Mamdani inference method is used to infer the appropriate rules and centroid meth od is used for defuzzification. In our experiment reference position is 0.3 m. The single input type-1 fuzzy logic controller is also implemented to show the efficiency of SIIT2FLC. In this experiment only two tuning parameters (λ, Gu) are needed. Here λ=1.10 and Gu=3 are chosen. Output responses of both controllers are given in Fig. 8, 9 for ball position and ball velocity.

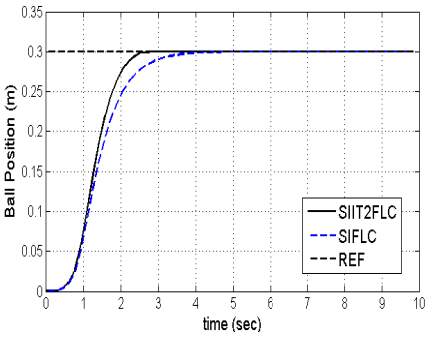


Fig. 8. Output response for ball position

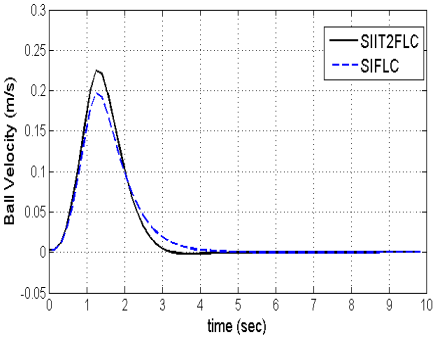


Fig. 9. Output response for ball velocity

To proof the robustness of the proposed controller, we have introduced a disturbance signal to the output of the position sensor of both controllers and we have changed the system parameters also to check the capability of handling uncertainty for both controllers. Here a sinusoidal signal of 0.1v amplitude and 1Hz frequency is used as a disturbance signal. The value of ball mass and radius are changed to 0.05m and 0.015m respectively.

The responses of BBS, after applying the disturbance signal and parameter variation, are given in Fig. 10 and Fig. 11.

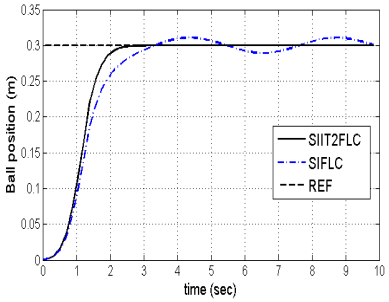


Fig. 10. Output response for ball position with parameter uncertainty and disturbance

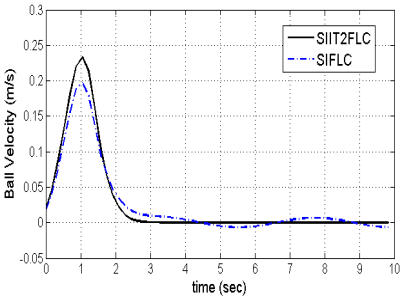


Fig. 11. Output response for ball velocity with parameter uncertainty and disturbance

5 Comparison of Experimental Results

From the real-time implementation of the SIIT2FLC and SIFLC, it has been found out that both controllers are able to give the stability to the BBS. There are no overshoots in position response for both controllers. In the case of SIIT2FLC, the ball reaches to the steady state position in 2.50 sec and it is 3.60 sec for the case of SIFLC. After applying the disturbance and parameter variation, the output position response is almost same for SIIT2FLC, but the position response is not stable (oscillatory) for SIFLC.

6 Conclusions

This paper has described a robust single input interval type-2 fuzzy logic controller for ball and beam system. Here the signed distance method makes the rule base smaller. The FOU gives the power of handling the uncertainty related to the system. The transient response is faster for proposed controller than single input fuzzy controller. In the presence of disturbance and parameter uncertainty, the proposed controller shows stable operation of the ball and beam system. But the single input fuzzy logic controller is not able to give stable operation of ball and beam system with disturbance and parameter uncertainty. The experimental results depict the robustness of the proposed controller.

References

[1] Amjad, M., Kashif, M.I., Abdullah, S.S., Shareef, Z.: A simplified intelligent controller for ball and beam system. In: 2nd International Conference on Education Technology and Computer (ICETC), vol. 3, pp. 494–498 (2010)

[2] Ball & Beam GBB1004 User's Guide & Experimental Manual by Googol Technology, 2nd edn (March 2006)

[3] Bartolomeo, C., Mose, G.: Type-2 fuzzy control of a bioreactor. In: IEEE International Conference on Intelligent Computing and Intelligent Systems, vol. 2, pp. 700–704 (2009)
[4] Choi, B.-J., Kwak, S.-W., Kim, B.K.: Design of a single-input fuzzy logic controller and its properties. Fuzzy Sets Syst. 106(8), 299–308 (1999)
[5] Choi, B.J., Kwak, S.W., Kim, B.K.: Design and Stability Analysis of Single-Input Fuzzy Logic Controller. IEEE Transaction on Systems, Man and Cybernetics-Part B: Cybernetics. 30(2), 303–309 (2000)
[6] Nguyen, D.-H., Huynh, T.-H.: A SFLA based fuzzy controller for balancing a ball and beam system. In: 10th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 1948–1953 (2008)
[7] Hirschorn, R.M.: Incremental sliding mode control of ball and beam. IEEE Transactions on Automatic Control 47, 1696–1700 (2002)
[8] Beibei, H., Yan, G.: Variable Universe Fuzzy Controller with Correction Factors for Ball and Beam System. In: 3rd International Workshop on Intelligent Systems and Applications (ISA), pp. 1–4 (2011)
[9] Jo, N.H., Seo, J.H.: A state observer for nonlinear systems and its application to ball and beam system. IEEE Transactions on Automatic Control 45, 968–973 (2000)
[10] Maruthi, T.R., Mahindrakar, A.D.: Robust stabilization using time scaling and Lyapunov redesign: The ball and beam system. In: 11th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 1661–1666 (2010)
[11] Keshmiri, M., Jahromi, A.F., Mohebbi, A., Amoozgar, M.H., Xie, W.-F.: Modeling and control of ball and beam system using model based and non-model based control approaches. International Journal on Smart Sensing and Intelligent Systems 5(1), 14–35 (2012)
[12] Karnik, N.N., Mendel, J.M.: Type-2 Fuzzy Logic Systems. IEEE Transactions on Fuzzy Systems 7(6), 643–658 (1999)
[13] Liang, Q., Mendel, J.M.: Interval type-2 fuzzy logic systems: Theory and design. IEEE Trans. Fuzzy Syst. 8(5), 535–550 (2000)
[14] Sepulveda, R., Lin, P.M., Rodriguez, A., Mancilla, A., Montiel, O.: Analyzing the effects of the Footprint of Uncertainty in Type-2 Fuzzy Logic Controllers. Engineering Letters, EL_13_2_12 (Advance online publication) 13(2) (August 4, 2006)
[15] Chan, W.-S., Lee, C.-Y., Chang, C.-W., Chang, Y.-H.: Interval type-2 fuzzy neural network for ball and beam systems. In: International Conference on System Sci-ence and Engineering (ICSSE), pp. 315–320 (2010)
[16] Yu, W., Ortiz, F.: Stability analysis of PD regulation for ball and beam system. In: IEEE Conference on Control Applications, pp. 517–522 (2005)

Adaptive Neuro Fuzzy Inference Structure Controller for Rotary Inverted Pendulum

Rahul Agrawal and R. Mitra

Indian Institute of Technology, Roorkee
rahulaggr@gmail.com, rmtrafec@iitr.ernet.in

Abstract. This paper presents the adaptive neuro fuzzy inference (ANFIS) controller for the Rotary inverted pendulum to balance it at it's the up-right position. The steps for implementation of four input controller is presented and shown that designing of this controller is very simple and at the same time it reduces the time and space complexity of the controller. The controller and the inverted pendulum are simulated in the Matlab Simulink environment with the help of ANFIS editor GUI. Simulation result shows that ANFIS controller is much better in comparison to conventional PID and Fuzzy logic controller in terms of settling time, overshoot and parameter variation.

1 Introduction

A mechanical system which has greater number of joints than the number of actuator present in the system such system is called the underactuated system [1]. Because of this, the strategies developed for fully actuated system may not be directly applied to underactuated system. The control study of underactuated system has drawn a great interest in last few decades as most of the physical systems have underactuated dynamics as those in robotics, aerospace engineering and marine engineering including the example of flexible-link robots, walking robots acrobatic robots, helicopter, satellite, space robot, spacecrafts etc.

The Rotary Inverted Pendulum is a widely investigated nonlinear system due to its property of unstable, higher order, multi-variable and highly coupled which can be treated as highly non-linear control system. Rotary system provides an excellent experimental platform for examining specific control theories or typical solution and thus promoting the development of new theories. This system can be taken as the problem of balancing the pendulum at up-right position which is the most common issue in robotics. This explains the fact that many investigations have been carried out on the rotary inverted pendulum problem [1]-[5].

For control the balancing act of the rotary pendulum, a control system is needed. As known the ANFIS can be used as controller as it can model the human decision making based on the IF-THEN rules and become a very popular tool for the approximation and inexpensive tool to implement and shows the adaptive and robust behavior in comparison to more commonly used conventional controller like PID and compensator like lead-lag and fuzzy controller. As known the conventional controller completely relies on the mathematical model of underlying system while efficient fuzzy controller, designed with the help of LQR parameters can be easily implemented to linear and nonlinear systems [3].

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp. 1163–1170.
springerlink.com

In ANFIS, fuzzy inference system is blended with the neural networks and uses the human intelligence to design the controller. This paper presents the schematic design of ANFIS model of the controller for the rotary inverted pendulum. Here the rotary inverted pendulum and controller is model before putting them into the simulation and controller is train, validate and check the performance with the noise and varying parameter. Then the controller and system model is implemented in Simulink environment of MATLAB and the performance of the controller is measured and run in real-time workshop. This is followed by the implementation and comparison of the PID and fuzzy controller with ANFIS controllers through simulations.

2 Rotary Inverted Pendulum Model

The Rotary system, as shown in Fig. 1, consists of a vertical pendulum, a horizontal arm, a gear chain, and a servomotor which drives the pendulum through the gear transmission system. The rotating arm is mounted on the output gear of the gear chain. An encoder is attached to the arm shaft to measure the rotating angle of the arm. At the end of the rotating arm there is a hinge instrumented with an encoder. The pendulum is attached to the hinge.

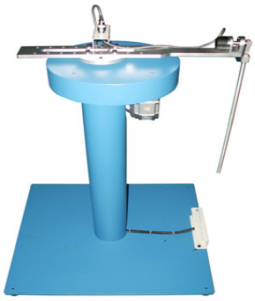


Fig. 1. Rotary Inverted Pendulum System

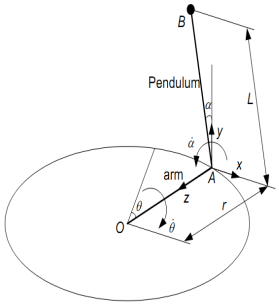


Fig. 2. Simplified model of the rotary inverted pendulum system

The inverted pendulum (mechanical part only) is sketched in Fig 2, α and θ are employed as the generalized coordinates to describe the inverted pendulum system. The pendulum is displaced with a given α while the arm rotates an angle of θ. We assume the pendulum to be a lump mass at point B which is located at the geometric center of the pendulum. The xyz frame is fixed to the arm at point A.

Nonlinear dynamic equations

$$a\ddot{\theta} - b\cos(\alpha)\ddot{\alpha} + b\sin(\alpha)\dot{\alpha}^2 + e\dot{\theta} = fV_m \tag{1}$$

$$-b\cos(\alpha)\ddot{\theta} + c\ddot{\alpha} - d\sin(\alpha) = 0 \tag{2}$$

Where $a = J_{eq} + mr^2 + \eta_g K_g^2 J_m, b = mLr, c = \frac{4}{3}ml^2, d = mgl$

$e = B_{eq} + \frac{\eta_m \eta_g K_t K_g^2 K_m}{R_m}, f = \frac{\eta_m \eta_g K_t K_g}{R_m}, J_{eq}$ is moment of inertia of pendulum and arm about the axis θ and η_m and η_g are the motor and gear efficiency respectively. K_g, K_m, K_t are the servo system gear ratio, back-emf constant and motor torque constant respectively.

Linearzing (1, 2) under the assumption that $\alpha \approx 0$ and $\dot{\alpha} \approx 0$, we get the linearized model as follows:

$$a\ddot{\theta} - b\ddot{\alpha} + e\dot{\theta} = fV_m \tag{3}$$

$$-b\ddot{\theta} + c\ddot{\alpha} - d\alpha = 0 \tag{4}$$

The overall block diagram of a rotary pendulum system with a feedback ANFIS control block is shown in Fig. 3. The output of the plant $(\theta, \dot{\theta}, \alpha, \dot{\alpha})$ is fed back to the controller to produce subsequent amount of voltage to balance the pendulum to its up-right position and at the same time maintaining the arm at the initial position.

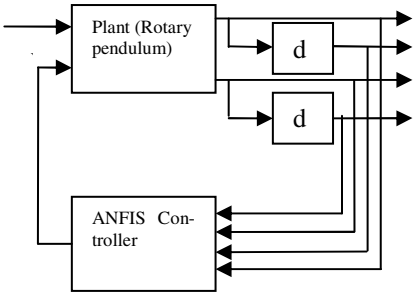


Fig. 3. Block Diagram of Inverted Pendulum System with feedback ANFIS controller

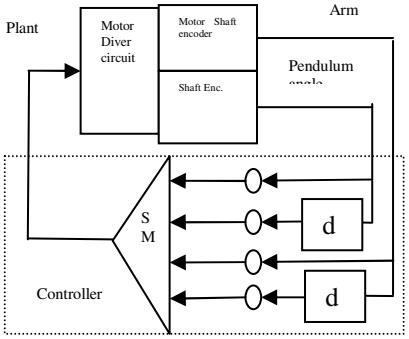


Fig. 4. Plant and controller block diagram

The model of the inverted pendulum and the controller is created using Simulink. As a whole the algorithm for the controller to balance the pendulum at up-right position is to calculate the voltage which needs to give to the servomotor. Fig. 4 shows how the voltage is calculated from the pendulum angle, pendulum angle acceleration, arm angle, arm angle acceleration measured from their respected sensor. The Fig. 4 shows that motor shaft encoder gives the arm angle while another shaft encoder placed at the end of the arm gives the pendulum angle and then the angle accelerations are derived from arm and pendulum angle.

The four circles (Fig. 4) K1, K2, K3, k4 are four "knobs" used to provide the gain to the four feedback signals. They are summed together and feed-back to the system as to give the voltage to the motor to rotate the arm. This can be expressed as

$$V_m = (K1 * \theta) + (K2 * \dot{\theta}) + (K3 * \alpha) + (K4 * \dot{\alpha}) \tag{5}$$

The controller input gains K1, K2, K3 and K4 are determined using the Linear-Quadratic Regulator (LQR) method described by Friedland [6]. This method finds the optimal K based on the state feedback law and the state-space equation derived earlier. For finding out the closed loop stability analysis of inverted pendulum we find out the root locus analysis, frequency analysis and many techniques.

3 ANFIS Controller

ANFIS adaptive Neuro-Fuzzy system was first introduced by J. Jang in 1993 [8]. ANFIS constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a backpropagation algorithm alone or in combination with a least squares type of method. This uses a network-type structure similar to that of a neural network, which maps inputs through input membership functions and associated parameters, and then through output membership functions and associated parameters to outputs, can be used to interpret the input/output map. The parameters associated with the membership functions changes through the learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector. This gradient vector provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. When the gradient vector is obtained, any of the learning algorithms is applied in order to adjust the parameters to reduce the error (squared difference between actual and desired outputs).

Here the structure of the fuzzy inference system (FIS) is Takagi-Sugeno type and four input variables arm angle, arm angular velocity, pendulum angle and pendulum angular velocity are considered and all input variables are having two membership functions. The parameter values of these membership functions are trained by ANFIS to provide the appropriate value of the voltage applied to motor which achieves the goal of balancing the pendulum. Fig. 5 shows the structure of the ANFIS controller.

For generating the FIS structure ANFIS editor GUI, already available in MATLAB is used. In editor grid portion type structure is selected and hybrid learning is chosen. Training data is available from the above mentioned LQR method which is randomly divided into training, testing and checking data. After training checking and testing of the ANFIS controller, above shown (Fig.5) structure is obtained. This structure is considered as five layer feed-forward neural network.

A. The first layer- This layer is a basic input Fuzzification layer where the crisp inputs are allocated relative fuzzy values.

B. The second layer- This layer of the nodes labels defines the specified membership functions for each input created in the layer one. Gaussian shaped fuzzy memberships are utilized here.

C. The third layer- The nodes in this layer represent the rules generated for different combinations and instances of inputs. This layer will give the information regarding which rules are to be fired for different possibilities of inputs.

D. The fourth layer- This layer produces the defuzzified Takagi-Sugeno-type output for each previous ith output. Here a particular defuzzified value is getting generated for each and every rule fired.

E. The fifth layer- The single node in this layer computes the overall outputs as the summation of all incoming signals. That gives the overall output that is generated from all the rules fired for particular set of input values.

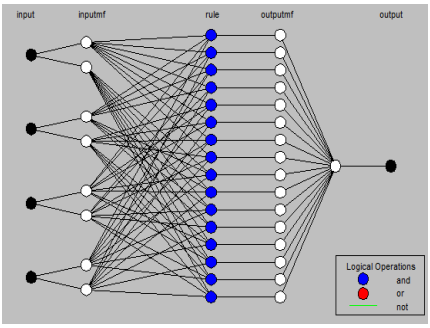


Fig. 5. Structure of ANFIS controller

Table 1. Values used in simulations

Parameter	Values	Parameter	Values
B_{eq}	0.004	K_m	0.00767
J_{eq}	0.0035842	K_t	0.00767
J_m	3.87 e-7	L	0.1675
K_g	70	r	0.215
R_m	2.6	η_g	0.9
η_m	0.69	g	9.8
m	0.125		

Therefore the output of the ANFIS is clearly is a linear function of all the inputs. This can be seen as the Rule-Base of this controller is given by

Rule Base: If θ is A1 and $\dot{\theta}$ is B1 and α is C1 and $\dot{\alpha}$ D1 then

$$V_m = (K1 * \theta) + (K2 * \dot{\theta}) + (K3 * \alpha) + (K4 * \dot{\alpha})$$

4 Simulations Result and Discussion

The rotary inverted pendulum and controller are implemented in Matlab Simulink environment. For the controller firstly FIS file is generated from the ANFIS editor GUI and used in a fuzzy logic controller block in the Simulink. The non-linear model of rotary pendulum is designed in the Matlab Simulink. The experiment is tested in real-time also.

4.1 Simulation Results

The Simulink model is simulated with ode5 solver and 0.001s sampling time. In Fig. 6 the falling angle of the pendulum and voltage applied to the servomotor is shown.

The applied voltage is calculated by the ANFIS controller which has a maximum and minimum limit of $\pm 6V$. In figure it can be seen that the pendulum is get stable in 1.3s.

The response of the arm angle versus desired angle is plotted in the Fig. 7. This shows that the desired arm angle which is 30 degree in this case is achieved in the nominal time about 1sec.

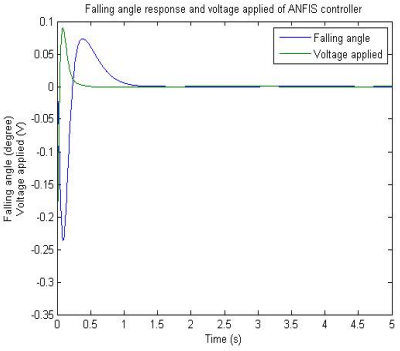


Fig. 6. Falling angle and voltage applied plot

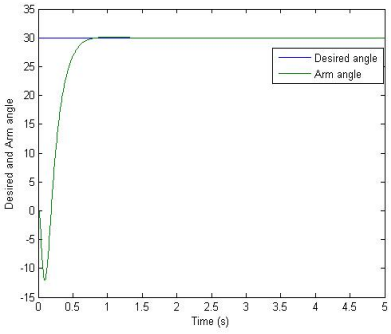
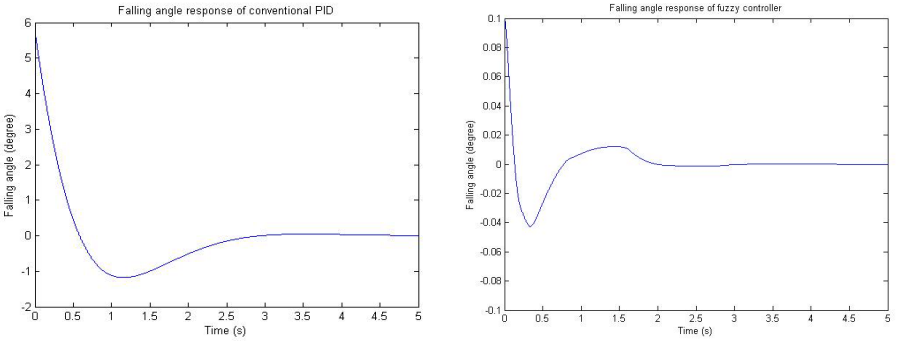


Fig. 7. Desired arm position and arm response of ANFIS controller

4.2 Comparison of ANFIS and Conventional PID and Fuzzy Control

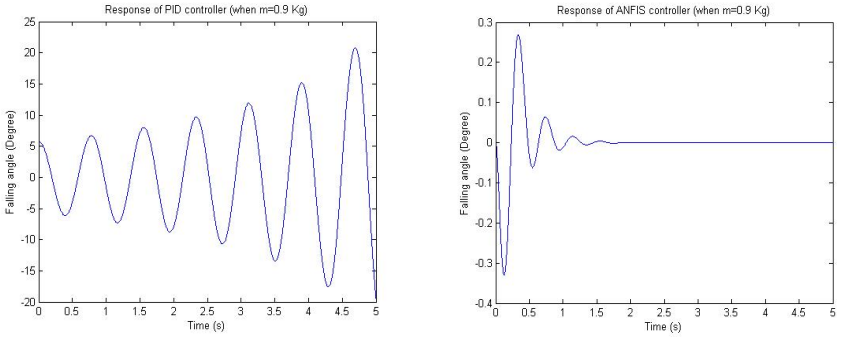
The conventional PID controller and fuzzy controller are designed for the same rotary inverted pendulum problem to compare the result with the proposed ANFIS controller. For the same plant parameter a PID controller is designed with proportional gain KP, derivative gain KD, integral gain KI 5, 11, and 0.02 respectively and efficient fuzzy controller [3] based on LQR. The falling angle of the pendulum in case of ANFIS, PID and efficient fuzzy is plotted in Matlab shown in Fig. 8. The graph shows that the settling time and overshoot of the ANFIS controller are much less than PID and efficient fuzzy controller.

In another set of experiment the robustness of the two ANFIS and PID has been checked by changing the mass of the Pendulum is changed from 0.125 to 0.85 kg without changing the parameters of the controllers. Fig. 9 shows that the conventional controller gives the un-damped oscillation and unable to stabilize the pendulum anymore while the ANFIS gives the reasonable response however poor than the previous one when no mass has been changed and stabilizes the pendulum in 1.7 sec. This proves that proposed ANFIS controller is more robust and does not rely on mathematical description of the plant.



(a) (b)

Fig. 8. Falling pendulum angle of (a) PID (b) efficient fuzzy controller



(a) (b)

Fig. 9. Falling pendulum angle response of (a) PID controller (b) ANFIS controller when mass is changed to 0.85 Kg

5 Conclusions

In this paper, ANFIS controller is designed for rotary inverted pendulum in Matlab Simulink with the help of ANFIS editor GUI. The designing of this controller has the advantages of both the intelligent technique Fuzzy and Neural networks together. In comparison of the modern control design technique, ANFIS is simpler to implement as it eliminate the complicated mathematical process and use the soft computing techniques. In the simulation result it is shown that ANFIS controller is more robust to system parameter variation in comparison to conventional PID and fuzzy controller.

References

1. Saber, R.O.: Nonlinear control o f underactuated mechanical systems with application to robotics and aero space vehicles, PhD Thesis. MIT (2001)
2. Akhtaruzzaman, M., Shafie, A.A.: Modeling and control of a rotary inverted pendulum using various methods, comparative assessment and result analysis. In: International Conference on Mechatronics and Automation (ICMA), August 4-7, pp. 1342–1347 (2010)

3. Krishen, J., Becerra, V.M.: Efficient fuzzy control of a rotary inverted pendulum based on LQR mapping. In: IEEE International Symposium on Intelligent Control, October 4-6, pp. 2701–2706 (2006)
4. Ozbek, N.S., Efe, M.O.: Swing up and stabilization control experiments for a rotary inverted pendulum an educational comparison. Systems Man and Cybernetics (SMC), 2226–2231 (October 2010)
5. Khanesar, M.A., Teshnehlab, M., Shoorehdeli, M.A.: Sliding mode control of Rotary Inverted pendulm. In: Mediterranean Conference on Control & Automation, MED 2007, June 27-29, pp. 1–6 (2007)
6. Friedland, B.: Control System Design, 2nd edn. McGraw Hill (1986)
7. Tatikonda, R.C., Battula, V.P., Kumar, V.: Control of inverted pendulum using adaptive neuro fuzzy inference structure (ANFIS). In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), May 30-June 2, pp. 1348–1351 (2010)
8. Jang, J.-S.R.: ANFIS: Adaptive-Network-based Fuzzy Inference Systems. IEEE Transactions on Systems, Man, and Cybernetics 23(3), 665–685 (1993)

Erratum: A Fuzzy Sectional Real-Time Scheduling Algorithm Based on System Load

Annappa B.

annappa@nitk.ac.in

Aswatha Kumar M. et al. (Eds.): Proceedings of ICAdC, AISC 174, pp.1145–1153.
springerlink.com

DOI 10.1007/978-81-322-0740-5_142

In this paper, the second author's name and his affiliation were missing in the published version. Please find below the second author's name and his affiliation:

Abhinav KR

National Institute of Technology, Karnataka
abhinav.kr.90@gmail.com

The original online version for this chapter can be found at
http://dx.doi.org/10.1007/978-81-322-0740-5_139

Author Index

Accamma, I.V. 997
Agarwal, Shivam 95
Aghav, Jagannath 163
Aghav, Sushila 1045
Agrawal, Rahul 1163
Ahlawat, Priyanka 347
Ahmadi, Samaneh 75
Aithal, Shridhar 441
Ajitha, S. 41
Alevoor, Praveen 137
Algur, Siddu P. 773
Ameen, Mohd Noorul 645
Amrutha, K.M. 69
Angadi, S.A. 209
Annappa, B. 1145
Ansari, Nazneen 801
Apparao, A. 575
Arulheethayadharthani, S. 475
Ashok Kumar, A.R. 739
Ashwini, P. 69
Aswatha Kumar, M. 753
Ayachit, N.H. 773
Azath, M. 371

Bakthavachalam, R. 247
Bala Tripura Sundari, B. 539
Balgurgi, Pooja P. 699
Banakar, R.M. 339, 461
Bandyopadhyay, Susmita 387
Banerjee, Anwesha 435
Banerjee, Sreeparna 1083
Baskaran, R. 873
Bellikatti, Vinodkumar I. 353
Bhagyashala, J. 707
Bhat, Naagesh S. 361
Bhat, Parvaiz Ahmad 319
Bhatnagar, Charul 859
Bhoi, Sourav Kumar 415, 865
Biswas, Bhaskar 297
Blumenstein, Michael 905

Chadha, Shraddha 1037
Chakraborty, Saptarshi 667
Chakraborty, Sudipta 651
Chandrika, T. 747
Chatterjee, Navonil 651
Chayadevi, M.L. 1091
Chetan, B.V. 911
Chetan, S. 353
Chouhan, Dharamendra 1
Chouhan, Vikas 511
Chowdhury, Pinaki Roy 229

Dalal, U.D. 399
Das, Arnab 387
Das, Pratyusha 435
Das, Soumendu 1083
Datta, Shounak 435
Dawn, Rose 1101
Deb, Suash 291
Desai, Vaishali 801
Devaraju, J.T. 491
Dhanni, Yogesh Kr. 257
Dhavachelvan, P. 467, 873
Dhivya Prabha, G. 919
Dilip Kumar, S.M. 1
Dinesh, M.S. 943
Dora, Syamala Kumari 107
D'Souza, Andrea 787

Elango, C. 247
Eswara Reddy, B. 927
Eswar Reddy, B. 1027
Evangelin Geetha, D. 41

Ganga, K. 919
Gangolia, Hemant 95
Garg, Deepak 113
Garg, Nidhi 807
Garg, Urvashi 131
Geetha, D.E. 31
George Philip, C. 429
Ghate, P.M. 1037
Ghodasara, Yogesh 935
Goel, Shivani 113
Gopularam, Bhanu Prakash 407
Gosavi, Vihang 95
Gosh, D. 171, 611
Goswami, Diganta 739
Govardhan, A. 679
Govindarajan, Kannan 561
Gupta, Rahul 807
Gupta, Shalini 611

Hanumanthappa, M. 63, 721, 795
Hareesh, K.S. 673
Hariharan, S. 841
Harikrishnan, V. 841
Hegadi, Ravindra S. 963
Hegde, Kavana 639
Hegde, Soumya 639
Hiremath, Manjunath 977
Hiremath, P.S. 977
Hudgi, Suvarna 329

Jadhav, Nitin P. 773
Jadhav, Omsai 667
Jagannatha, S. 31
Jagtap, Sonal K. 699, 1075
Jain, Vishal 155
Jalal, Anand Singh 859
Jalan, Saket 229
Jalin Gladis, D. 919
Janarthanan, R. 435
Jayanna, H.S. 893
Jayaram, M.A. 899
Jindal, Ankur 621
Jindal, Shaivya 593
John, Snoeji Varghese 1069
Jose Moses, G. 423

Kabir, Md. Enamul 1109
Kagawade, Vishwanath C. 1135
Kaliangra, Fibinse 1069
Kambale, Ankita 1037
Kanhar, Debananda 107
Kantharaj, Chethana 303
Kapoor, Kalpesh 137
Karegowda, Asha Gowda 899
Karthikeya Sharma, T. 101
Kavitha Rani, G. 69
Keshava Reddy, E. 927
Khilar, Pabitra Mohan 415, 865
Kiran, Ravi 113
Kodabagi, M.M. 209
Kokare, Manesh B. 885
Kolhe, Roshan 163
Konar, Amit 435
Koteswara Rao, G. 501
Kotha, Dileep Kumar 527
Krishnan, Varsha 539
Kulkarni, Subhash 1117
Kumar, Abhishek 379
Kumar, Anupam 833
Kumar, Manish 721
Kumar, Preetham 807
Kumar, Sanjeev 183
Kumar, T.N.R. 9
Kumar, Vijay 833
Kumari, Rupu 859
Kundu, Sumanta 1155
Kurian, M.Z. 501
Kushal, K.S. 353

Lavanya, G. 1061

Madhu, T. 519
Madhusudhana Rao, S. 561
Mahadevan, G. 191
Mahanand, B.S. 753
Mahmood, Abdun Naser 1109
Mamatha, Y.N. 101
Manjunath, A.S. 899
Marndi, Raj N. 629
Marwala, T. 1125
Mattihalli, Channamallikarjuna 371
Meena, Yogesh Kumar 131
Megha 787
Minavathi 943
Mir, Roohie Naaz 319

Mishra, Ashirbad 267
Mitra, Anirban 121
Mitra, R. 379, 1163
Mohan, Arvind 297
Mohan Murthy, M.K. 645
Mohanty, Rakesh 267
Mohapatra, Durga Prasad 277
Mukhopadhyay, Sajal 171, 611
Mund, G.B. 601
Mundada, Monica R. 69
Munegowda, Keshava 691
Murali, S. 9, 943
Murthy, Jayanthi K. 483
Murthy, K.N.B. 201
Musa Mohinuddin, K.S.Md. 1021
Mustafa, Abdul K. 1109
Muthusamy, C. 629

Nadaf, Mahmad M. 339
Nagabhushana, B.S. 729
Nagaraja, B.G. 893
Nageswara Rao, G. 575
Naik, Anima 49
Nalini, N. 407
Narayan, V.S. 461
Narayana, M. 1117
Narayana, V.N. 879
Narayanan, N.K. 285
Nargund, V.B. 715
Navalgund, Siddalingesh 639
Navaneethakrishnan, S. 247
Navitha, M.V. 501
Nayak, Chinmohan 585
Nayak, Ramanuja 49
Nelson Kennnedy Babu, C. 455
Nigam, M.J. 257, 955, 1155

Obula Konda Reddy, R. 927

Pai, Anusha R. 221
Pal, Arup Kumar 95
Pal, Srikanta 905
Pal, Umapada 905
Panda, B.S. 49
Panda, Manoj Kumar 833
Panda, Rajanikant 1101
Panda, Sanjaya Kumar 415, 865
Pandeeswari, S. 1061
Pandey, Parul 237
Pandya, Siddharth M. 747
Pani, Santosh 601
Panthi, Vikas 277
Panwar, Adesh 393
Parameshachari, B.D. 1005
Parameswarappa, S. 879
Parhi, Manoranjan 585
Parvathi, K. 49
Patel, Dhara J. 1069
Patel, Minal 303
Patel, Nidhi 955
Patel, Z.M. 399
Patil, Kiran Kumari 729
Patil, Pushpa B. 885
Patil, Siddram R. 329
Pavanje, Nirupama 763
Paygude, Shilpa 1045
Peddoju, Sateesh Kumar 511
Prabhakar, M. 191
Pradeep, K.R. 429
Prakash, B.R. 795
Pramod, Mane 171
Prasad, R.S. 155
Prasantha, H.S. 201
Pratheesh, R. 455
Pratima, S.M. 461
Priyanshu 593
Pujeri, Ramachandra V. 147
Pushpa, M.K. 85

Ragupathy, R. 549
Raj, Divyashree K. 69
Rajagopalan, M.R. 561
Rajani Kanth, K. 31, 41
Rajendiran, M. 455
Rajeswaran, N. 519
Rajpurohit, Vijay S. 715
Raju, G.T. 691, 983, 1091
Raju, Veera Manikandan 691
Rama Krishna Rao, T.K. 575
Ramanaiah, O.B.V. 991
Ramaswamy, Srini 75
Ramkumar, T. 841
Ram Mohana Reddy, G. 787
Rao, Aakarsh 303
Rao, S.V. 739
Rath, Santanu 527
Ravi, Aarthi 787
Ravi Kumar, C.N. 847
Ravi Kumar, G. 629

Reddy, Mallamma V. 63
Roy, Sanjiban Sekhar 667

Sadale, Rohan 163
Sahukari, Ganesh 739
Sambasiva Rao, V. 483
Samuel, Cherian 1011, 1051
Sandhya, M. 1101
Sanjay, H.A. 645
Sannakki, Sanjeev S. 715
Saraswati, Vivek 1011
Sarda, Pratik 137
Saroja, V.S. 461
Sarvesh Babu, N.S. 101
Satapathy, Shashank Mouli 601
Satapathy, Suresh Chandra 49
Sateesh Kumar, P. 621
Saurav, Swapnil 667
Seetharaman, K. 549
Selvamuthukumaran, S. 841
Seshu Kumar, A.N. 813
Sethu Selvi, S. 85
Setia, Amit 347
Shama 899
Shanmugapriya, R.K. 1061
Sharada, G. 991
Sharma, Pankaj 593
Sharma, Tushar 593
Shashidhara, H.L. 201
Shefali, S. 707
Shet, Milan S. 303
Shirdavani, Shiva 75
Shivaputra 353
Shrinivasacharya, Purohit 969
Shukla, K.K. 229
Shukla, Ruchi 1125
Shwetha, D. 491
Siddamal, Saroja V. 339
Singh, J.N. 191
Singh, Manoj Kumar 823
Singh, Vinay Pratap 1011
Singhal, Gaurav 297
Sohid, Moudud 651
Somasundaram, Thamarai Selvi 561
Sowmyarani, C.N. 57
Soyjaudah, K.M.S. 1005
Sreedevi, Y. 1027
Sridhar, Rajeswari 475, 919
Srikumar, Abhishek 747
Srinath, S. 847
Srinivasan, G.N. 57
Srinivasarao, P. 575
Srivatsa, S.K. 9
Subbaiah, P. 1021
Subramanya Bhat, M. 491
Sudhakara, G. 441
Sudhamani, M.V. 969, 983
Sukanya, K. 57
Sultana, Shabana 659
Suma, H.N. 997
Suma, K.V. 303
Sumana, M. 673
Sundar, Aparna 1037
Sunil Kumar, D. 423
Supriya, N. 423
Suresh, R.M. 679
Suresh Kumar, T.V. 31, 41, 721
Suresh Varma, P. 423
Suryakalavathi, M. 519
Suvarna, Ananya 787
Syed Ibrahim, B. 455

Talreja, Maahi 801
Tanna, Paresh 935
Taralabenchi, Jayashree 639
Thasleema, T.M. 285
Thimmappa, P. 483
Thontadharya, H.J. 491
Tibarewala, D.N. 435
Tipu Rahaman, S. 1021
Tiwari, Anunay 1051
Tiwari, Atul Kumar 1011, 1051
Tiwari, Mayank 593
Tripathi, Maheshwari 237
Tripathy, Abinash 121

Udayagiri, Sandeep 163
Uma, R. 467
Umamaheswari, A. 1061
Umashankar, B. 501
Uplane, M.D. 1075

Vasavi, S. 813
Vasudev, T. 1135
Veena, K.N. 19
Veena, R.S. 147
Victer Paul, P. 873
Vidya, T. 899

Vidya Raj, C. 659
Vijaya, P.A. 911
Vijaya Kumar, B.P. 1, 19, 729
Vijayashree, C.S. 1135
Vinay, S. 441
Vinaya Babu, A. 629
Vinoth Kumar, S. 679
Viswanatham, Madhu 667

Waghmare, Kiran 221
Warade, Saket 163
Wathore, Sachin 163

Yang, Xin-She 291
Yasser, Patel Mohammed 645
Yedke, Tejas 155
Yogeesha, C.B. 147