Lecture Notes in Networks and Systems 40

Simon Fong · Shyam Akashe
Parikshit N. Mahalle *Editors*

Information and Communication Technology for Competitive Strategies

Proceedings of Third International Conference on ICTCS 2017

Springer

Lecture Notes in Networks and Systems

Volume 40

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
e-mail: kacprzyk@ibspan.waw.pl

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

More information about this series at http://www.springer.com/series/15179

Simon Fong · Shyam Akashe
Parikshit N. Mahalle
Editors

Information and Communication Technology for Competitive Strategies

Proceedings of Third International Conference on ICTCS 2017

Springer

Editors
Simon Fong
Department of Computer and Information Science
University of Macau
Macau, People's Republic of China

Parikshit N. Mahalle
Smt. Kashibai Navale College of Engineering
Pune, India

Shyam Akashe
Department of Electronics and Communication Engineering
ITM University
Gwalior, India

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-981-13-0585-6 ISBN 978-981-13-0586-3 (eBook)
https://doi.org/10.1007/978-981-13-0586-3

Library of Congress Control Number: 2018941974

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This LNNS volume contains the papers presented at the ICTCS 2017: Second International Conference on Information and Communication Technology for Competitive Strategies. The conference was held during 16–17 December 2017, Udaipur, India, and organized by Association of Computing Machinery, Udaipur Professional Chapter, in association with The Institution of Engineers (India), Udaipur Local Center and Global Knowledge Research Foundation. It will target state-of-the-art as well as emerging topics pertaining to information and communication technologies (ICTs) and effective strategies for its implementation for engineering and intelligent applications. The objective of this international conference is to provide opportunities for the researchers, academicians, industry persons and students to interact and exchange ideas, experience and expertise in the current trend and strategies for ICT. Besides this, participants will also be enlightened about vast avenues, current and emerging technological developments in the field of ICT in this era, and its applications will be thoroughly explored and discussed. The conference is anticipated to attract a large number of high-quality submissions, stimulate the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, students from all around the world and provide a forum to researcher; propose new technologies, share their experiences and discuss future solutions for design infrastructure for ICT; provide a common platform for academic pioneering researchers, scientists, engineers and students to share their views and achievements; enrich technocrats and academicians by presenting their innovative and constructive ideas; focus on innovative issues at international level by bringing together the experts from different countries. Research submissions in various advanced technology areas were received, and after a rigorous peer review process with the help of programme committee members and external reviewer, 74 papers were accepted with an acceptance rate of 21%. The conference featured many distinguished personalities like Dr. Dharm Singh, Namibia University of Science and Technology, Namibia; Dr. Nilanjan Dey, Techno India College of Engineering, Kolkata, India; Dr. Poonam Dhaka, University of Namibia. Separate invited talks were organized in industrial and academia tracks on both days. We are indebted to all organizing partners for their immense support to make this conference possible in such a grand scale. A total of eight sessions were organized as a part of *ICTCS 2017* including six technical, one plenary and one inaugural session. A total of 61 papers were presented in the six technical sessions with high discussion insights. The total number of accepted submissions was 74 with a focal point on ICT and intelligent systems. Our sincere thanks to all sponsors, press, print and electronic media for their excellent coverage of this conference.

Macau, People's Republic of China — Simon Fong
Gwalior, India — Shyam Akashe
Pune, India — Parikshit N. Mahalle

Contents

Multilayer Visual Cryptography with Soft Computing Approach for Authentication 1
Pallavi Chavan

Design and Implementation of Adders and Multiplier in FPGA Using ChipScope: A Performance Improvement 11
Suhas B. Shirol, S. Ramakrishna and Rajashekar B. Shettar

Signaling Protocol Specification for Signaling Approach-Based Virtual Output Queue Router Architecture 21
Jaya R. Surywanshi and Dinesh V. Padole

Prediction-Based Parallel Clustering Algorithm for M-Commerce 31
Likhesh N. Kolhe, Vaishali Khairnar and Ashok Kumar Jetawat

Effects of Cognitive Ability, Trust and Time-Saving: Predicting Further Amelioration and Successive Usage of E-ticketing with TAM, TPB and Cognitive Frameworks 41
Rifat Hossain, S. M. Hasan Mahmud, Md. Altab Hossin, Touhid Bhuiyan and Yin Xiao Hua

Knowledge Management Systems Leveraging the Competitive Advantage of Top IT Organizations: A Multi-case Study of Benchmarking Practices 53
Geeta Sachdeva and Seema

Performance Analysis of Parallel Implementation of Morphological Operations 65
Ashwin Geet D'Sa and B. G. Prasad

US Dollar's Influence on Indian Gold Price: Assessment Using Artificial Neural Network 81
Deepa Bogale, Aniket Muley, Parag Bhalchandra and Govind Kulkarni

Attribute Based Storage Mechanism with De-duplication Filter: A Review Paper 89
Amarja Hanumant Bhosale and Amrita A. Manjrekar

An Approach for Traffic Congestion Detection and Traffic Control System 99
Tarun Kumar and Dharmender Singh Kushwaha

An Analytical Implementation of CART Using RStudio for Churn Prediction 109
Vani Kapoor Nijhawan, Mamta Madan and Meenu Dave

Traffic Control and Management Over IoT for Clearance of Emergency Vehicle in Smart Cities 121
Biru Rajak and Dharmender Singh Kushwaha

Review and Analysis of Simulation Results for Consumer Attack Model and Grid Sensor Placement Algorithm 131
Yuvaraj S. Patil and Swati V. Sankpal

An Approach to Analyze Behavior of Network Events in NS2 and NS3 Using AWK and Xgraph 137
Radhika Patel, Nehal Patel and Sandip Patel

Hybrid Latent Semantic Analysis and Random Indexing Model for Text Summarization 149
Niladri Chatterjee and Nidhika Yadav

State-of-the-Art Energy-Efficient Thermal-Aware Scheduling in Cloud 157
Ritu Garg and Rama Rani

Eye Gaze Tracking-Based Adaptive E-learning for Enhancing Teaching and Learning in Virtual Classrooms 165
I. Joe Louis Paul, S. Sasirekha, S. Uma Maheswari, K. A. M. Ajith, S. M. Arjun and S. Athesh Kumar

A Novel Text-Based User Authentication Scheme Using Pseudo-dynamic Password 177
Ramsha Fatima, Nadia Siddiqui, M. Sarosh Umar and M. H. Khan

Concurrency Control Algorithms for Translation Lookaside Buffer 187
Manisha Agarwal and Manisha Jailia

Digitization of Disaster Management: A Multimedia Ontological Approach 197
Vishal Lama and Sarika Jain

Prevalent Study of Cooperative Biometric Identification System Using Multimodal Biometrics 205
Shaleen Bhatnagar and Sushmita Kumari

Most Persistent Feature Selection Method for Opinion Mining of Social Media Reviews 213
Savita Sangam and Subhash Shinde

Physical Layer Security with Energy Harvesting Techniques: A Study 223
Poonam Jindal and Rahul Kumar

One-Tip Secure: Next-Gen of Text-Based Password 235
Zainab Zaheer, Aysha Khan, M. Sarosh Umar and Muneeb Hasan Khan

A Technique to Envisage Investigative Codes for Enduring Maladies Using Data Mining Techniques 245
Vinitha Dominic, Shaleen Bhatnagar and Garima Goyal

Review of Machine Learning Techniques for Optimizing Energy of Home Appliances 255
Jasmeet Kaur and Anju Bala

Performance Analysis of "En-SPDP: Enhanced Secure Pool Delivery Protocol" for Food Delivery Network 265
Havan Somaiya, Radhakrishna Kamath, Vijay Godhani, Yash Ahuja and Nupur Giri

Survey on Fault-Tolerance-Aware Scheduling in Cloud Computing 275
Chesta Kathpal and Ritu Garg

Video Redundancy Investigation and Cooperative Cache Model in Wireless Network 285
E. M. Malathy and N. Radha

Efficient Energy Sensor Arrays with Uneven Transport Interval in WSN 293
Chinmya Kumar Nayak and Satyabrata Das

Stress and Modal Analysis of Six-Axis Articulated Robot Using ANSYS 303
Supriya Sahu and B. B. Choudhury

New Text-Based User Authentication Scheme Using CAPTCHA 313
Atiya Usmani, Amrah Maryam, M. Sarosh Umar and Muneeb Hasan Khan

Modeling and Control of a Six-Wheeled Mobile Robot 323
Jayprakash Rout, Subham Agrawal and B. B. Choudhury

FPGA Implementation of Parallel Transformative Approach in AES Algorithm 333
Padma Prasada, Sathisha, Ajay Prinston Pinto and H. D. Ranjith

An Approach to Wireless Sensor Networks for Healthcare Scrutiny in Hospitals 341
K. Sujatha, K. SenthilKumar, V. Balaji, R. KrishnaKumar and Nallamilli P. G. Bhavani

Foetal Heartbeat and Volume Detection Using Spherical Harmonics Shape Model for Level Set Segmentation 351
K. Sujatha, V. Balaji, Nallamilli P. G. Bhavani and S. Jayalakshmi

An Efficient Approach for Clustering US Census Data Based on Cluster Similarity Using Rough Entropy on Categorical Data 359
G. Sreenivasulu, S. Viswanadha Raju and N. Sambasiva Rao

An Algorithm for Load Balancing in Computational Grid Using Activity-Based and CPU Usage Approach 377
Ramesh T. Prajapati, Tejas Kadiya, Bhavesh Jain and RajeshKumar

ICT's Role in Building and Understanding Indian Telemedicine Environment: A Study 391
S. Syed Thouheed Ahmed, M. Sandhya and S. Shankar

Age Classification with LPCC Features Using SVM and ANN 399
Gaurav Aggarwal and Latika Singh

Analyzing Challenges to Transportation for Successful Sustainable Food Supply Chain Management Implementation in Indian Dairy Industry 409
Yogesh Kumar Sharma, Sachin Kumar Mangla and Pravin P. Patil

Smart Focused Web Crawler for Hidden Web 419
Sawroop Kaur and G. Geetha

Current Methodologies for Energy Efficient Cloud Data Centers 429
Sanjay Patel and Krutika Vyas

An Enhanced Scheme for PHR on Cloud Servers Using CP-ABE 439
Nishant Doshi, Mayur Oza and Nikita Gorasia

Decision Tree Algorithms for Prediction of Heart Disease 447
Srabanti Maji and Srishti Arora

Green Communication for Wireless Cooperative Networks: A Survey 455
Sonika Pahuja and Poonam Jindal

Self-training with Neighborhood Information for the Classification of Remote Sensing Images 465
Prem Shankar Singh Aydav and Sonajharia Minz

Instance Selection Using Multi-objective CHC Evolutionary Algorithm 475
Seema Rathee, Saroj Ratnoo and Jyoti Ahuja

Inclusive Device to Connect People with the Surrounding World 485
Vishal B. Pattanashetty, Nalini C. Iyer, H. L. Viswanath and Shweta Kore

Profiling-Based Effective Resource Utilization in Cloud Environment Using Divide and Conquer Method 495
Namra Bhadreshkumar Shah, Neel Dhananjaybhai Shah, Jitendra Bhatia and Harshal Trivedi

Algorithmic Approach for Error-Correcting Capability and Decoding of Linear Codes Arising from Algebraic Geometry 509
Arunkumar Patil and Nitin Darkunde

Interoperability Among Internet of Things (IoT) Components Using Model-Driven Architecture Approach 519
Kiranpreet Kaur and Anil Sharma

Hand Phone Open-Air Face Image Retrieval for Varying Focal Length and Occlusion from Cloud Storage 535
V. Elanangai

Survey of Different Approaches Used for Food Recognition 551
Sandhya Arora, Gauri Chaware, Devangi Chinchankar, Eesha Dixit and Shevi Jain

A Novel Deep Learning Framework Approach for Natural Calamities Detection 561
Rahul Nijhawan, Megha Rishi, Amit Tiwari and Rajat Dua

Comprehensive Analysis of Nonlocal Means Algorithm on Wide Set of ECG Signals 571
Supraja Kulkarni and Layak Ali

Periocular Region-Based Biometric Identification Using Local Binary Pattern and Its Variants 581
K. Kishore Kumar and Movva Pavani

Hand Geometry- and Palmprint-Based Biometric System with Image Deblurring 591
Hiren J. Galiyawala and Rupal Chaudhari

A Low Power, High Swing and Robust Folded Cascode Amplifier at Deep Submicron Technology 605
Varsha Bendre, Abdul Kadir Kureshi and Saurabh Waykole

Syllable-Based Concatenative Speech Synthesis for Marathi Language 615
Pravin M. Ghate and Suresh D. Shirbahadurkar

Inverted U-shape Monopole Antenna with Notch-Band Characteristics for UWB Communication Applications 625
Priyanka Gawhale, Bharat Patil, Archana Wasule and Vijay K. Sambhe

A Distance-Based Outlier Detection Using Particle Swarm Optimization Technique 633
Abdul Wahid and Annavarapu Chandra Sekhara Rao

Fast and Memory Leak-Free Verification for Bridge Transitions in VLSI Designs 645
Gaurav Sharma and Lava Bhargava

Privacy-Preserving K-NN Classification Using Vector Operations 655
Hanumantharao Jalla and P. N. Girija

A Comparative Study on Segmentation Techniques for Brain Tumor MRI 665
Imayanmosha Wahlang, Pallabi Sharma, Syeda Musfia Nasreen, Arnab Kumar Maji and Goutam Saha

Integrating Concentrating Solar Plant-Based System in Multi-area AGC Using LabVIEW 675
Dillip K. Mishra, Tapas Kumar Panigrahi, Asit Mohanty and Prakash K. Ray

Maximal Stable Extremal Region Extraction of MRI Tumor Images Using Successive Otsu Algorithm 687
Priya Jyotiyana and Saurabh Maheshwari

Synthesis of Reversible Array Divider Circuit 701
Somnath Rakshit, Sraboni Mondal, Aruna Chakraborty, Apurba Sarkar and Dipak K. Kole

Finding Right Doctors and Hospitals: A Personalized Health Recommender 709
M. Sridevi and R. Rajeshwara Rao

An Improved Hybrid Offset-Based Min-Sum Decoding Algorithm 721
Michaelraj Kingston Roberts and Elizabeth Sunny

Simulation of Phasor Measurement Unit for Research and Teaching Applications 727
Hannan Ahmad Khan, Mohd. Rihan and Mohammad Zaid

Maximum Network Flow Aware Multi-lightpath Survivable Routing in WDM Networks 737
Dinesh Kumar Tyagi and V. K. Chaubey

Impact of Feature Selection on EEG Based Motor Imagery 749
Mridu Sahu and Sneha Shukla

Various Machine Learning Algorithms for Twitter Sentiment Analysis 763
Rishija Singh and Vikas Goel

Author Index 773

Editors and Contributors

About the Editors

Simon Fong graduated from La Trobe University, Australia, with a first-class honours B.E. in Computer Systems and a Ph.D. in Computer Science in 1993 and 1998, respectively. He is now working as an associate professor at the Department of Computer and Information Science, University of Macau, China. He is also one of the founding members of the Data Analytics and Collaborative Computing Research Group at the Faculty of Science and Technology. Prior to his academic career, he held various managerial and technical positions, such as systems engineer, IT consultant and e-commerce director in Australia and Asia. He has published over 300 international conference and peer-reviewed journal papers, mostly in the areas of data mining and optimization algorithms.

Shyam Akashe is a professor at ITM University, Gwalior, Madhya Pradesh, India. He completed his Ph.D. at Thapar University, Punjab, India, and his M.Tech. in Electronics and Communication Engineering at the Institute of Technology and Management, Gwalior. His more than 190 publications include over 50 papers in SCI-indexed journals. The main focus of his research is low-power system-on-chip (SoC) applications in which static random access memories (SRAMs) are omnipresent. He has authored two books entitled *Moore's Law Alive: Gate-All-Around (GAA) Next Generation Transistor* published by Lambert Academic Publishing, Germany, and *Low Power High Speed CMOS Multiplexer Design* published by Nova Science Publishers, Inc., New York, USA. He has also published over 120 papers in refereed journals of national and international repute. He has participated in numerous national and international conferences and presented over 100 papers.

Parikshit N. Mahalle obtained his B.E. in Computer Science and Engineering from Sant Gadge Baba Amravati University, Amravati, India, and his M.E. in Computer Engineering from Savitribai Phule Pune University, Pune, India. He completed his Ph.D. in Computer Science and Engineering with specialization in Wireless Communication at Aalborg University, Aalborg, Denmark. He has more than 16 years of teaching and research experience and has authored eight books: *Identity Management for Internet of Things*, River Publishers; *Identity Management Framework for Internet of Things*, Aalborg University Press; *Data Structures and Algorithms*, Cengage Publications; *Theory of Computations*, Gigatech Publications; *Fundamentals of Programming Languages*—I, Gigatech Publications; *Fundamentals of Programming Languages*—II, Gigatech Publications; *Design and Analysis of Algorithms: A Problem Solving Approach* (In Process), Cambridge University Press. He is the recipient of "Best Faculty Award" from Sinhgad Technical Education Society (STES) and Cognizant Technology Solutions. He has also delivered an invited talk on "Identity Management in IoT" to Symantec Research Lab, Mountain View, California. Currently, he is working as a professor and head of the Department of Computer Engineering at STES's Smt. Kashibai Navale College of Engineering, Pune, India. His recent research interests include algorithms, Internet of things, identity management and security.

Contributors

Gaurav Aggarwal Manipal University Jaipur, VPO Dehmikalan, Jaipur, India

Manisha Agarwal Banasthali Vidyapith, Vanasthali, Rajasthan, India

Subham Agrawal Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India

Jyoti Ahuja Government Post Graduate College for Women, Rohtak, India

Yash Ahuja Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

K. A. M. Ajith Department of Information Technology, SSN College of Engineering, Chennai, India

Layak Ali School of Engineering, Central University of Karnataka, Gulbarga, India

S. M. Arjun Department of Information Technology, SSN College of Engineering, Chennai, India

Sandhya Arora Cummins College of Engineering, Pune, India

Srishti Arora Dehradun Institute of Technology, Dehradun, Uttrakhand, India

S. Athesh Kumar Department of Information Technology, SSN College of Engineering, Chennai, India

Prem Shankar Singh Aydav DTE, Lucknow, UP, India

Anju Bala Thapar University, Patiala, PB, India

V. Balaji Faculty of Electrical Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia

Varsha Bendre Rajarshi Shahu College of Engineering, Savitribai Phule Pune University, Pune, India

Parag Bhalchandra School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

Lava Bhargava Department of Electronics and Communication, Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Jitendra Bhatia Department of Computer Engineering, Vishwakarma Government Engineering College, Chandkheda, Ahmedabad, Gujarat, India

Shaleen Bhatnagar Presidency University, Bengaluru, India

Nallamilli P. G. Bhavani Department of EEE, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. M.G.R. Educational & Research Institute, Vels University, Chennai, India

Amarja Hanumant Bhosale Department of Technology, M. Tech. Computer Science and Technology, Shivaji University, Kolhapur, Maharashtra, India

Touhid Bhuiyan Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

Deepa Bogale School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

Aruna Chakraborty Department of Computer Science and Engineering, St. Thomas College of Engineering and Technology, Kolkata, West Bengal, India

Niladri Chatterjee Indian Institute of Technology Delhi, Hauz Khas, New Delhi, Delhi, India

V. K. Chaubey Department of EEE, BITS Pilani, Pilani, India

Rupal Chaudhari Department of Electronics & Communication Engineering, CGPIT, Uka Tarsadia University, Surat, India

Pallavi Chavan Department of Information Technology, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, MH, India

Gauri Chaware Cummins College of Engineering, Pune, India

Devangi Chinchankar Cummins College of Engineering, Pune, India

B. B. Choudhury Department of Mechanical Engineering, Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India

Ashwin Geet D'Sa Department of Computer Science, BMS College of Engineering, Bengaluru, India

Nitin Darkunde Swami Ramanand Teerth Marathwada University, Nanded, India

Satyabrata Das Department of CSE & IT, Veer Surendra Sai University of Technology (VSSUT), Burla, Sambalpur, Odisha, India

Meenu Dave Faculty of Engineering and Technology, Jagan Nath University, Jaipur, Rajasthan, India

Eesha Dixit Cummins College of Engineering, Pune, India

Vinitha Dominic Presidency University, Bengaluru, India

Nishant Doshi Pandit Deendayal Petroleum University, Gandhinagar, India

Rajat Dua Graphic Era University, Dehradun, India

V. Elanangai Sathyabama Institute of Science and Technology, Chennai, India

Ramsha Fatima Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Hiren J. Galiyawala Department of Electronics & Communication Engineering, CGPIT, Uka Tarsadia University, Surat, India

Ritu Garg Department of Computer Engineering, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Priyanka Gawhale Usha Mittal Institute of Technology, Mumbai, India

G. Geetha Lovely Professional University, Phagwara, Punjab, India

Pravin M. Ghate JSPM's Rajarshi Shahu College of Engineering, Pune, Maharashtra, India

Nupur Giri Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

P. N. Girija SCIS, UOH, Hyderabad, TS, India

Vijay Godhani Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Vikas Goel Department of CS&E, Ajay Kumar Garg Engineering College, Ghaziabad, India

Nikita Gorasia Marwadi Education Foundation, Rajkot, India

Garima Goyal Presidency University, Bengaluru, India

S. M. Hasan Mahmud Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

Rifat Hossain Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh

Md. Altab Hossin University of Electronic Science and Technology of China, Chengdu, China

Yin Xiao Hua University of Electronic Science and Technology of China, Chengdu, China

Nalini C. Iyer B.V. Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India

Manisha Jailia Banasthali Vidyapith, Vanasthali, Rajasthan, India

Bhavesh Jain Indrashil Institute of Science and Technology, Rajpur, Kadi, Gujarat, India

Sarika Jain National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Shevi Jain Cummins College of Engineering, Pune, India

Hanumantharao Jalla CBIT, Hyderabad, TS, India

S. Jayalakshmi Deptartment of EEE, Ramachandra College of Engineering, Eluru, Andhrapradesh, India

Ashok Kumar Jetawat Pacific Academy of Higher Education and Research University, Udaipur, India

Poonam Jindal Department of Electronics & Communication, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

I. Joe Louis Paul Department of Information Technology, SSN College of Engineering, Chennai, India

Priya Jyotiyana Department of Computer Science and Engineering, Government Women Engineering College, Ajmer, India

Tejas Kadiya Indrashil Institute of Science and Technology, Rajpur, Kadi, Gujarat, India

Radhakrishna Kamath Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

Chesta Kathpal Department of Computer Engineering, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Jasmeet Kaur Thapar University, Patiala, PB, India

Kiranpreet Kaur School of Computer Applications, Lovely Professional University, Phagwara, Kapurthala (Jalandhar), India

Sawroop Kaur Lovely Professional University, Phagwara, Punjab, India

Vaishali Khairnar Terna Engineering College, Navi Mumbai, Maharashtra, India

Aysha Khan Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Hannan Ahmad Khan Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India

M. H. Khan Department of Computer Engineering, Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Muneeb Hasan Khan Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

K. Kishore Kumar Faculty of Science & Technology, Department of ECE, IFHE, Hyderabad, India

Dipak K. Kole Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India

Likhesh N. Kolhe Pacific Academy of Higher Education and Research University, Udaipur, India

Shweta Kore J.S.S College, Dharwad, Karnataka, India

R. KrishnaKumar School of Electrical Engineering, Vel's University, Chennai, India

Govind Kulkarni School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

Supraja Kulkarni School of Engineering, Central University of Karnataka, Gulbarga, India

Rahul Kumar Department of Electronics & Communication, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Tarun Kumar Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, UP, India

Sushmita Kumari Presidency University, Bengaluru, India

Abdul Kadir Kureshi Vishwabharati Academy's College of Engineering, Ahmednagar, India

Dharmender Singh Kushwaha Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, UP, India

Vishal Lama National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Mamta Madan Information Technology, Vivekananda Institute of Professional Studies, Pitam Pura, Delhi, India; Information Technology, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

Saurabh Maheshwari Department of Computer Science and Engineering, Government Women Engineering College, Ajmer, India

Arnab Kumar Maji North-Eastern Hill University, Shillong, Meghalaya, India

Srabanti Maji Dehradun Institute of Technology, Dehradun, Uttrakhand, India

E. M. Malathy Depatment of Information Technology, SSN College of Engineering, Chennai, Tamilnadu, India

Sachin Kumar Mangla Knowledge Management and Business Decision Making, Plymouth Business School (PBS), University of Plymouth, Plymouth, UK

Amrita A. Manjrekar Department of Technology, M. Tech. Computer Science and Technology, Shivaji University, Kolhapur, Maharashtra, India

Amrah Maryam Department of Computer Engineering, Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Sonajharia Minz JNU, New Delhi, India

Dillip K. Mishra International Institute of Information Technology, Bhubaneswar, India

Asit Mohanty College of Engineering & Technology, Bhubaneswar, India

Sraboni Mondal Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India

Aniket Muley School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Maharashtra, India

Syeda Musfia Nasreen North-Eastern Hill University, Shillong, Meghalaya, India

Chinmya Kumar Nayak Department of CSE & IT, Veer Surendra Sai University of Technology (VSSUT), Burla, Sambalpur, Odisha, India

Rahul Nijhawan Indian Institutes of Technology Roorkee, Roorkee, India

Vani Kapoor Nijhawan Information Technology, Vivekananda Institute of Professional Studies, Pitam Pura, Delhi, India; Information Technology, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

Mayur Oza Marwadi Education Foundation, Rajkot, India

Dinesh V. Padole G.H. Raisoni College of Engineering, Nagpur, India

Sonika Pahuja Department of Electronics and Communication, HMRITM, Hamidpur, New Delhi, India

Tapas Kumar Panigrahi International Institute of Information Technology, Bhubaneswar, India

Nehal Patel Charusat University, Anand, Gujarat, India

Radhika Patel Charusat University, Anand, Gujarat, India

Sandip Patel Charusat University, Anand, Gujarat, India

Sanjay Patel Department of Computer Engineering, LDRP-ITR, Gandhinagar, Gujarat, India

Arunkumar Patil Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded, India

Bharat Patil Usha Mittal Institute of Technology, Mumbai, India

Pravin P. Patil Department of Mechanical Engineering, Graphic Era (Deemed To be University), Dehradun, Uttarakhand, India

Yuvaraj S. Patil D. Y. Patil College of Engineering and Technology, Affiliated to Shivaji University, Kolhapur, Maharashtra, India

Vishal B. Pattanashetty B.V. Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India

Movva Pavani Faculty of Science & Technology, Department of ECE, IFHE, Hyderabad, India

Ajay Prinston Pinto MITE, Mangalore, India

Ramesh T. Prajapati Indrashil Institute of Science and Technology, Rajpur, Kadi, Gujarat, India

B. G. Prasad Department of Computer Science, BMS College of Engineering, Bengaluru, India

Padma Prasada MITE, Mangalore, India

N. Radha Depatment of Information Technology, SSN College of Engineering, Chennai, Tamilnadu, India

Biru Rajak Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad, UP, India

RajeshKumar Indrashil Institute of Science and Technology, Rajpur, Kadi, Gujarat, India

R. Rajeshwara Rao JNTU, Kakinada, Kakinada, Andhra Pradesh, India

Somnath Rakshit Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India

S. Ramakrishna Department of Electronics & Communication Engineering, B V Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India

Rama Rani Department of Computer Engineering, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

H. D. Ranjith MITE, Mangalore, India

Annavarapu Chandra Sekhara Rao Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Dhanbad, Jharkhand, India

Seema Rathee Guru Jambheshwar University of Science and Technology, Hisar, India

Saroj Ratnoo Guru Jambheshwar University of Science and Technology, Hisar, India

Prakash K. Ray Nanyang Technical University, Singapore, Singapore

Mohd. Rihan Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India

Megha Rishi College of Engineering Roorkee, Roorkee, India

Michaelraj Kingston Roberts Department of Electronics and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

Jayprakash Rout Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India

Geeta Sachdeva National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

Goutam Saha North-Eastern Hill University, Shillong, Meghalaya, India

Mridu Sahu Department of IT, National Institute of Technology, Raipur, Raipur, Chhattisgarh, India

Supriya Sahu Department of Mechanical Engineering, Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India

N. Sambasiva Rao VCE, Hyderabad, India

Vijay K. Sambhe Veermata Jijabai Technological Institute, Mumbai, India

M. Sandhya School of Computer, Information and Mathematical Sciences, BSA Crescent University, Chennai, India

Savita Sangam Shivajirao S. Jondhale College of Engineering, Dombivli(E), Thane, Maharashtra, India

Swati V. Sankpal D. Y. Patil College of Engineering and Technology, Affiliated to Shivaji University, Kolhapur, Maharashtra, India

Apurba Sarkar Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Shibpur, West Bengal, India

M. Sarosh Umar Department of Computer Engineering, Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

S. Sasirekha Department of Information Technology, SSN College of Engineering, Chennai, India

Sathisha MITE, Mangalore, India

Seema National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India

K. SenthilKumar Faculty of Electrical Engineering, Department of ECE, Dr. M.G.R. Educational & Research Institute, Chennai, India

Namra Bhadreshkumar Shah Department of Computer Engineering, Vishwakarma Government Engineering College, Chandkheda, Ahmedabad, Gujarat, India

Neel Dhananjaybhai Shah Department of Computer Engineering, Vishwakarma Government Engineering College, Chandkheda, Ahmedabad, Gujarat, India

S. Shankar School of Computer, Information and Mathematical Sciences, BSA Crescent University, Chennai, India

Anil Sharma School of Computer Applications, Lovely Professional University, Phagwara, Kapurthala (Jalandhar), India

Gaurav Sharma Department of Electronics and Communication, Malaviya National Institute of Technology, Jaipur, Rajasthan, India

Pallabi Sharma North-Eastern Hill University, Shillong, Meghalaya, India

Yogesh Kumar Sharma Department of Mechanical Engineering, Graphic Era (Deemed To be University), Dehradun, Uttarakhand, India

Rajashekar B. Shettar Department of Electronics & Communication Engineering, B V Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India

Subhash Shinde Lokmanya Tilak College of Engineering, Kopar Khairane, Navi Mumbai, Maharashtra, India

Suresh D. Shirbahadurkar Zeal College of Engineering & Research, Pune, Maharashra, India

Suhas B. Shirol Department of Electronics & Communication Engineering, B V Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India

Sneha Shukla Department of IT, National Institute of Technology, Raipur, Raipur, Chhattisgarh, India

Nadia Siddiqui Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Latika Singh The NorthCap University, Gurugram, India

Rishija Singh Department of CS&E, Ajay Kumar Garg Engineering College, Ghaziabad, India

Havan Somaiya Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India

G. Sreenivasulu Department of CSE, ACEEC, Hyderabad, India

M. Sridevi Anurag Group of Institutions, Hyderabad, Telangana, India

K. Sujatha Faculty of Electrical Engineering, Department of EEE, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. M.G.R. Educational & Research Institute, Vels University, Chennai, India

Elizabeth Sunny Department of Electronics and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India

Jaya R. Surywanshi G.H. Raisoni College of Engineering, Nagpur, India

S. Syed Thouheed Ahmed School of Computer, Information and Mathematical Sciences, BSA Crescent University, Chennai, India

Amit Tiwari College of Engineering Roorkee, Roorkee, India

Harshal Trivedi SoftVan, Ahmedabad, India

Dinesh Kumar Tyagi Department of CSE, MNIT Jaipur, Jaipur, India

S. Uma Maheswari Department of Information Technology, SSN College of Engineering, Chennai, India

Atiya Usmani Department of Computer Engineering, Zakir Hussain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

S. Viswanadha Raju JNTU Jagityala, Karimnagar, India

H. L. Viswanath Dayanand Sagar Engineering College, Bengaluru, Karnataka, India

Krutika Vyas Department of Computer Engineering, LDRP-ITR, Gandhinagar, Gujarat, India

Abdul Wahid Department of Computer Science and Engineering, Indian Institute of Technology (Indian School of Mines), Dhanbad, Dhanbad, Jharkhand, India

Imayanmosha Wahlang North-Eastern Hill University, Shillong, Meghalaya, India

Archana Wasule K. J. Somaiya Polytechnic, Mumbai, India

Saurabh Waykole Pimpri Chinchwad College of Engineering, Pune, India

Nidhika Yadav Indian Institute of Technology Delhi, Hauz Khas, New Delhi, Delhi, India

Zainab Zaheer Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Mohammad Zaid Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India

Multilayer Visual Cryptography with Soft Computing Approach for Authentication

Pallavi Chavan

Abstract In this paper, a novel approach of visual cryptography is presented as a paradigm shift in domain of visual cryptographic schemes. General concepts of visual cryptography and its strategies are also discussed in this paper. Visual cryptography is a variant of secret sharing schemes which encrypt the secret information in such a way that mathematical operations are not necessary to decode the secret. Human visual system is necessary and sufficient condition for secret decryption. Multilayer visual cryptography scheme encrypts the secret information through number of levels. Prior knowledge of the depth of the encryption is available. The scheme encrypts randomly generated secrets in binary form, handwritten text as well as biometric secrets. The paper also states the methodology of key share to reduce the superimposition overhead while decrypting the secrets. The scheme is absolutely expansionless and eliminates the graying effect from the decoded secret. Multilayer encryption technique is suitable for authentication and authorization. Soft computing approach of authentication stated in this paper is based on feed-forward perceptron neural network.

Keywords Soft computing · Visual cryptography · Authentication

1 Introduction

Multilayer visual cryptography is capable of performing variety of cryptographic computations. In multilayer visual cryptography context, multiple layers indicate the depth of encryption. Multilayer visual cryptography evolved as a multilevel scheme of visual cryptography. The design of multilayer visual cryptography is fundamentally determined by how the secret is handled in multiple levels without expansion. Multilayer scheme aims to experiment and demonstrate the secret

P. Chavan (✉)
Department of Information Technology, Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, MH, India
e-mail: pallavichavan11@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_1

sharing without expansion through number of levels. The basic model of visual cryptography used as a platform in this paper is visual cryptography, which can be considered as secret sharing scheme [1]. However, in contrast to the conventional cryptographic schemes, most visual cryptography algorithms encrypt the visual information only. It has a great impact on reduction of mathematical computations while revealing the secret. Any visual cryptography technique argues that the original secret can be revealed directly through necked eyes without mathematical computations. The most important advantage of visual cryptography is the simple reconstruction mechanism [2]. The proposed research work attempts to integrate the concept of layered encryption with visual cryptography. Multilayer approach is proposed in two different modes: First mode is proposed as plain mode of multilayer visual cryptography. Second mode is proposed as threshold-based approach of multilayer visual cryptography. It aims to encrypt biometric secrets in gray-level representation. In this paper, authors follow XOR method for reconstruction. All the shares produced by multilayer visual encryption are in digital form requires no transparencies for decoding. The results of multilayer decoding are absolutely expansionless, eliminating graying effect and reduces delay in reconstruction. Multilayer visual cryptography is very different from image steganography and watermarking. With steganography, in order to hide secret message, cover images are essentially used. Such cover images are not used with multilayer visual cryptography. Watermarking is another issue seems similar to visual cryptography, but actually it is not. Multilayer visual cryptography does not emboss the secret pattern in another cover image [1, 2]. Visual cryptography is the practice that encrypts visual information. Handwritten text, signatures, and the images are visual information of spatial domain. The input information for encryption is known as secret. Encryption process yields the cipher; unreadable information presented by cipher is called as share. The part of the secret in scrambled form is known as a share. The fundamental concept of visual cryptography states the sharing of secret information among a group of n participants. In order to share particular secret, it is decomposed into n number of shares. These shares are distributed among the participants. In order to reconstruct the original secret, each participant has to provide his/her own share. In some of the VC schemes, complete knowledge of $n - 1$ shares is unable to decrypt the secret. Visual cryptography was independently introduced by Shamir [3] in the form of secret sharing. Shamir divided secret information D, into n number of pieces (shares) such that D is reconstructable from any k number of shares [3].

1.1 Visual Cryptography Schemes

In (2, 2) scheme, the secret image is divided into exactly two shares. Division of the pixels into shares is random. To reconstruct the original secret, these two shares are superimposed together [4]. Improved version of visual cryptography is (k, n) scheme. It is a generalized scheme of visual cryptography which allows dividing a secret into n number of shares. The secret can be revealed from any k number of shares among n. Knowledge of any (k − 1) shares do not reveal any information. The major problem associated with this scheme is that the user needs to maintain many shares. Maintaining the multiple shares increases memory consumption. Increased number of shares involves more participants in the scheme [4]. In (n, n) scheme, secret is divided into n number of shares such that reconstruction of secret depends upon all n shares. Different types of shares appear with visual cryptography [4]. *Horizontal Shares*: In this type of shares, black and white pixels appear horizontally as rows. *Diagonal Shares*: In case of diagonal shares, the combination of black and white pixels is presented diagonally. *Vertical Shares*: In case of vertical shares, a single pixel is represented as a column of pixels in corresponding shares. *Random Shares*: In case of random shares, black and white pixels appear as random collection. Despite of all the share patterns, visual cryptography reflects some important properties. Some of the properties are listed here: *Divide and Reconstruct, Decoding Computations, Graying Effect*, *Use of Transparencies.* Youmaran R. et al., in 2006, proposed new concept of hiding the secret information into multiple color images. These color images are acting as covers to hide the secret. Encoding of secret message is done in such a way that the decoding is possible by superimposition of all covers generated out the scheme. The human visual system is the sufficient condition to reveal the hidden information. No computations are involved in the decryption process. N number of images participates in encoding and decoding process. This method achieves recovery of secret without any loss [5, 6]. Sagar Kumar Nerella et al., in 2012, proposed a novel method of secret encryption using wavelet and visual cryptography. The authors stated that different visual cryptography schemes shall be defined as per the color model of the secret image. Wavelet transform is used to convert a secret of any color model to its equivalent gray-level representation. Antoni 9/7 filter is used to transform color secret to its gray-level representation [7]. Kaur K. et al., in 2013, presented a secure method encrypting a secret information or image by dividing it into shares. They focus on distinct properties of visual cryptography of decoding the information visually. RSA is a public key encryption method used to protect the shares. Shares are not available in standard form. This approach adds extra layer of security over visual cryptography. The scheme is secured against denial-of-service attack (DoS) and dictionary attack [8–10]. The proposed approach of multilayer visual cryptography reduces all associated factors of VC, including mean square error, graying effect, and pixel expansion ratio. As the levels of encryption increase, secrecy of information tends to increase. Varieties of methods have been implemented with visual cryptography. Almost all the method introduces graying effect during encryption due to which quality of decrypted secret is affected badly. These methods and existence of graying effect are represented in Table 1.

Table 1 Standard methods with visual cryptography

Sr. no.	VC scheme	Method used	Graying effect
1	Flexible visual cryptography [11]	Multiple shares	Present
2	Multisecret sharing [12]	(k, n)	Present
3	Novel VC [13]	Secret sharing	Present
4	Thresholding [14]	(n, t, n)	Present
5	Contrast-enhanced VC [15]	Extra pixel patterns	Present
6	Secured VC [16]	Biometric encryption	Present
7	Color visual cryptography [17]	Intelligent approach	Present

2 Multilayer Visual Cryptography-Plain Mode

Basically, multilayer visual cryptography is defined with two different modes: plain mode and threshold mode. This segregation is done on the basis of the kind of the secret they encrypt. Plain mode is absolutely encrypting the binary secrets in the form of random passwords while threshold mode encrypts biometric secrets. Multilevel encryption (MLE) is the basic building block of proposed approach. It allows secret encryption in multiple levels. MLE encrypts the secrets of spatial domain. It signifies that the encryption process is expansionless. This expansionless encryption constitutes the basic property of plain multilayer visual encryption. In the proposed architecture, two independent levels of encryption are identified. Third level of composition is the additional level in this model. In this mode, secret S is encrypted in the first level of multilayer encryption yielding two shares, namely share1 and share2. These two shares are independently encrypted resulting four different shares, viz. share11, share12, share21, and share22. In third level known as composition level, three shares are combined to form a new share named as key share. The remaining share is known as user share. Thus, plain mode yields two resultant shares at the end of encryption. They are denoted by share11 and key share. Share11 is the remaining share of level 2 of encryption, while key share is the composed share out of three shares of the second level. In this figure, decoding of the secret is performed by logical XOR operation. Decoding is performed by the logical XOR operation. The requirement of the proposed method is that the secrets are assumed in binary representation, i.e., black and white combination of pixels. Secret variety includes passwords, signatures, handwritten text before encryption; the original secret is mapped into the new size which is multiple of 4. After resizing the secret to the 4 multiple size, starting with top-left corner of the secret, every 2×2 pixel block in a sequence is selected for encoding independently. Fifteen random combinations are suggested for encoding of 2×2 pixel block. If all pixels in 2×2 block of secret are black, then the combination is presented as a vector: [0 0; 0 0]. This block is encrypted using four possibilities, viz. [1 0; 0 1] [0 1; 1 0], [0 0; 1 1] [1 1; 0 0], [0 0; 1 1] [1 1; 0 0], and [0 1; 1 0] [1 0; 0 1] randomly. Similarly, if 2×2 block of secret is entirely white indicating the vector [1 1; 1 1], this block is encrypted with random four combinations presented by Eq. 2. There are many possibilities of the 2 × 2 pixel blocks for encoding. Initially, secret is normalized to the size multiple of 4. The window of 2×2 pixel block is navigated over the secret. Unlike traditional visual cryptography, multilayer approach does not rely solely on human visual system for decryption. Figure 1 indicates the secret encrypted and associated shares using plain mode. In this figure, (a) represents original secret S. It is the random password, and (b) and (c) indicate share1 and share2 generated by MLE1. (d) through (g) is collection of all the shares generated by MLE2, viz. share11, share12, share21, share22. All these shares reflect some information by visual inspection.

Many shares are generated through plain mode of VC at second level. To reduce the number of shares used in reconstruction, they need to get combine to form new share. Key share is the novel terminology introduced in the encryption process of multilayer encryption. Key share is defined to reduce overhead of reconstruction process with many shares. Expansionless method of multilayer encryption generates four resultant shares, namely share11, share12, share21, and share22. Key share is formed by combining three shares: share12, share21, and share22. Equation (1) represents the theoretical foundation for key share generation. In this

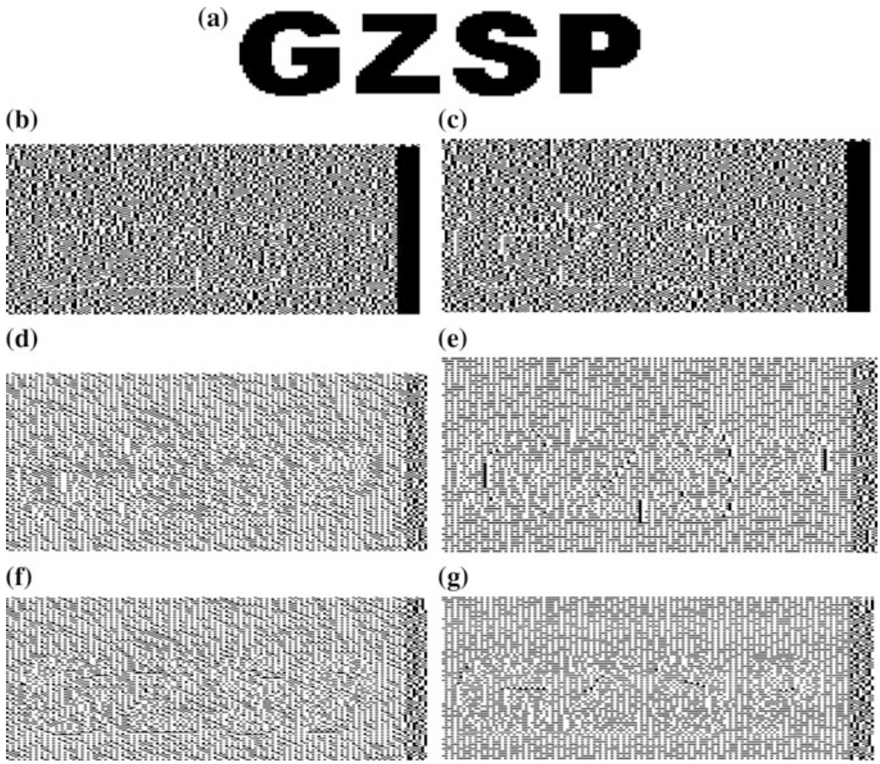


Fig. 1 Original secret and shares with plain mode of multilayer visual cryptography

formulation, S_i is the share with binary values. The nature of key share pixel value is also binary. Three shares are combined by summing method. Hence, the range of summation is 1…3. Maximum value of summation expected as 3 in a scenario when all three shares represent white pixel. Similarly, the minimum possible value of the summation is 0 considering that pixel in all the shares is black.

$$Key\,Share\,pixel\,Value = \begin{cases} 0 & if \sum_{i=1}^{3} Si \leq 1 \\ 1 & if \sum_{i=1}^{3} Si > 1 \end{cases} \tag{1}$$

Theorem 1 *If Share12, Share21, and Share22 are the inputs to key share generator, then the pixel value in key is determined by* Eq. 1.

Proof By referring Eq. 1, it is observed that the value for key share is dependent on individual values of all the shares. For any given value of all the three shares, key share value is determined. Hypothesis is given below: Considering that Share12 is 0, Share21 is 0, and Share22 is 1, then sum of all the values yields (0 + 0 + 1) = 1. As sum is equal to 1, key share value is 0 by referring to Eq. 1. In another case, if Share12 is 1, Share21 is 0, and Share22 is 1, sum of all the shares yields (1 + 0 + 1) = 2. As sum is greater than 1, key share value is 1. Key share also reveals some information by visual inspection. Graying effect is the collection of additional black pixels appearing in the background of the revealed secret. Almost traditional visual cryptographic schemes reflect the graying effect. It is observed that revealed secret is not retaining additional black pixels in the background. Absence of black pixels in background causes graying effect reduction. This reduction is responsible to degrade the revealed secret quality. From this figure, revealed secret is a discontinuous pattern. For alphanumeric secrets, it is feasible to undergo expansionless scheme. However, signature patterns and biometric secrets if encrypted by this approach, cause loss of information about the secret. Shares generated by plain mode are referred as a digital share as revealing is performed without introducing transparencies. As graying effect is reduced, concentration of white pixels in revealed secret increases at the same time concentration of black pixel decreases. Point of differentiation between concentrations of black pixel in original secret and reconstructed secret is that former one is more concentrated as compared to latter one. The revealed secret is a discontinuous pattern reflecting no graying effect while all the shares are random patterns.

3 Multilayer Visual Cryptography—Threshold Mode

This section describes the design and implementation of novel thresholding mode of multilayer visual cryptography. The spatial domain secrets particularly biometric secrets are considered for encryption (iris). In this model, secret is encrypted in three successive levels. Working principle of thresholding mode is equivalent to plain mode. At the first level of threshold mode, biometric secret is encrypted among two shares. At second level of hierarchy, each share is encrypted independently resulting four shares. Finally at third level, key share formation takes place. Key share is generated by combining any three shares of previous level. At the end of THVC encoding, two resultant shares are formed, viz. key share and remaining share out of second level. While encrypting gray-level secrets, dithering aspect is applied prior to encryption. Dithering converts all gray-level secrets into binary representation. Multilayer encoding enables the architecture to encrypt a secret in three different levels. It utilizes the capability of key share to reduce total participations involved in reconstruction of the secret. Theorem 2 describes the thresholding methodology of multilayer architecture.

Theorem 2 *Let p is a fraction such that* $p = \alpha > \lambda$*, for every white pixel of secret, share1, and share2 are assigned p while for every black pixel of secret, share1 is assigned p and share2 is assigned ~p.*

Proof This theorem deals with the distribution of black and white pixel information among the shares. Here, λ denotes threshold value for randomly generated number α. λ lies in the interval [0 1]. As share1 is assigned p, inversion of p is assigned to share2. It is true for every black pixel while share1 and share2 are assigned the same value of p for white pixel encryption. Using Theorem 2, the shares generated are robust, expansionless, and reflect no graying effect.

Experimental analysis of threshold mode is performed based on various parameters. These parameters are: graying effect, pixel expansion ratio, number of shares, share type, mean square error, and peak signal-to-noise ratio. During experimentation of threshold mode, major achievement gained is successful removal of the graying effect in revealed secret. In earlier work related to visual cryptography, graying effect was observed. Computational time for encryption is also reduced in here. Previous plain visual cryptographic technique was losing some information from original secret. This recovery of loss is performed by threshold mode. The graying effect is removed by threshold mode by reducing mean square error θ to zero. Mean square error is defined as the difference between original information and the reconstructed information. Analytically, mean square error is given by:

$$\theta = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \tag{2}$$

It is evaluated in experimentation that $\theta = 0$. In this technique, mean square error is computed between original secret I and the revealed secret K. Peak signal-to-noise ratio β is computed based on θ and found to be ∞. For least mean square error (equal to zero), peak signal-to-noise ratio is expected as ∞.

$$\beta = 10log_{10}(MAX_I^2/\theta) \tag{3}$$

In above formula, MAX denotes maximum possible value of pixel in secret *I*.

4 FFPNN Model Using Thresholding

With the results of threshold-based approach, novel authentication mechanism is proposed in this paper. Visual cryptography is found to be a robust approach for online authentication. Feed-forward perceptron neural network model is a soft computing approach proposed for authentication using multilayer visual cryptography. The architecture of this model indicates forward direction of information processing. Perceptron plays vital role in FFPNN. It is an artificial neuron yielding fixed output either zero or one. This is comfortable nature of neuron for shares generated by thresholding approach. FFPNN is a two-layer network. First layer is input layer directly accessing the input. This layer also acts as hidden layer as its output is not directly accessible. Second layer is the output layer. The number of neurons in both the layers is same and depends upon the number of pixel values in the input. A pixel matrix of secret image X is defined on finite field Si where i = 2 as secret is in binary form. For gray-level secrets, the finite field value appears to be 2^8. Set of shares is Y = $\{y_1, y_2, y_{11}, y_{12}, y_{21}, y_{22}\}$. Level 1 shares: y_1, y_2, level 2 shares: y_{11}, y_{12}, y_{21}, y_{22}, where $y_{ij} \cap y_{ji} = \phi$ for i, j = 1, 2. FFPNN is trained for given set of user share and associated secret with user share. At other side, revealed secret acts as targets. Algorithm 1 is defined for training and simulation of FFPNN network. In this algorithm, user share (from thresholding approach) is the input. As input is submitted to the input layer of FFPNN, network undergoes training. During training phase, error and actual outputs are computed. As the network is linear, activation function used for perceptron is logsig for layer 1 and purelin for layer 2 (well defined with neural network toolbox). Authentication mechanism is defined as: For given user input share, FFPNN model recalls associated revealed secret. In this model, recall is a process of retrieving output for given input. Best validation performance of FFPNN is shown in Fig. 2.

Algorithm 1: FFPNN Algorithm

1. *Input user share.*
2. *Target revealed secret.*
3. *Normalize inputs and targets.*
4. *Create network object in the neural network environment.*
5. *Initialize input size {n}.*
6. *Initialize number of layers {2}.*
7. *Assign transfer function to layer 1 {logsig}*
8. *Assign transfer function to layer 2 {purelin}*

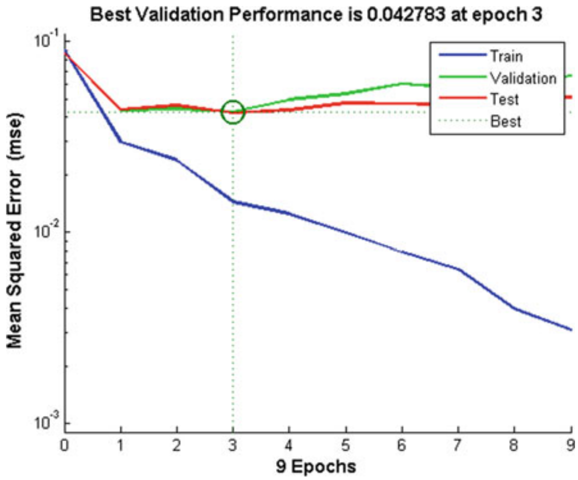


Fig. 2 Best validation performance of FFPNN

9. *Initialize network object.*
10. *Train and simulate the network.*
11. *Output network errors and network_outputs.*

5 Conclusion

In this paper, author presented multilayer approach of visual cryptography. Plain mode and threshold mode of encryption are discussed with the experimental results. Shares generated are perfectly expansionless. The scheme stated is perfectly secured removing the graying effect and the traditional use of transparencies. Digital shares are also introduced in this paper. Mean square error is achieved as zero with PSNR of infinity. Time complexities of plain mode and threshold mode are $O(n^2)$. FFPNN is identified as a soft computing approach of authentication using artificial neural network. In case of multilayer visual cryptography, encryption of color secrets is suggested. Scope of research lies with thresholding approach to propose a novel thresholding with fuzzy inference system.

References

1. Kandar, S., Maiti, A., Dhara, B.C.: Visual cryptography scheme for color image using random number with enveloping by digital watermarking. Int. J. Comput. Sci. **8**(3), 543–549 (2011)
2. Bidgar, P., Shahare, N.: Key based visual cryptography scheme using novel secret sharing technique with steganography. J. Electron. Commun. Eng. (IOSR-JECE) **8**(2), 11–18 (2013)
3. Shamir, A.: How to share a secret. Commun. ACM (Springer) 612–613 (1979)
4. Naor, M., Shamir, A.: Visual cryptography. Int. J. (Springer) 1–11 (1998)

5. Youmaran, R., Adler A., Miri A., An improved visual cryptography for secret hiding. In: Proceedings of IEEE 23rd Biennial Symposium on Communications, pp. 240–243 (2006)
6. Askari, N., Heys, H.M., Moloney, C.R.: An extended visual cryptography scheme without pixel expansion for halftone images. In: Proceedings of IEEE 26th Annual Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–6, May 2013
7. Nerella, S.K., Varma, K., Chaganti, G.R.S.: Securing images using color visual cryptography and wavelets. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **2**(3), 163–168 (2012)
8. Kaur K., Khemchandani V.: Securing visual cryptographic shares using public key encryption. In: Proceedings of 3rd IEEE International Conference on Advance Computing (IACC), pp. 1108–1113, February 2013
9. Yadav, G.S., Ojha, A.: A novel visual cryptography scheme based on substitution cipher. In: Proceedings of 2nd IEEE International Conference on Image Information Processing (ICIIP), pp. 640–643, December 2013
10. Chen, Y.-H., Hsu, P.-Y., Huang, W.-L.: Tagged visual cryptography with access control. In: Proceedings of IEEE International Conference on Multimedia and EXPO Workshops (ICMEW), pp. 1–5, July 2014
11. Liu, F., Guo, T., Wu, C.K., Yang, C.-N.: Flexible visual cryptography and it's application. In: Transactions on DHMS IX. LNCS, vol. 8363, pp. 110–130. Springer, Berlin, Heidelberg (2014)
12. Arumugam, S., Lakshmanan, R., Nagar, A.K.: On (k, n)*-visual cryptography scheme. J. Des. Code Cryptogr. **71**(1), 153–162 (2014)
13. Tzeng, W.-G., Chi-Ming, H.: A new approach for visual cryptography. J. Des. Code Cryptogr. **27**, 207–227 (2002)
14. Liu, Y.-X., Harn, L., Yang, C.-N., Zhanga, Y.-Q.: Efficient (n, t, n) secret sharing schemes. J. Syst. Softw. Sci. Direct (Elsevier) **85**, 1325–1332 (2012)

Design and Implementation of Adders and Multiplier in FPGA Using ChipScope: A Performance Improvement

Suhas B. Shirol, S. Ramakrishna and Rajashekar B. Shettar

Abstract In recent years, graph theory, digital design, signal processing, image processing, MIMO and digital control systems, and in many more applications and portable computing systems, design for high speed and low power has become the challenge in the field of VLSI design. The main focus of the paper is a performance improvement in designing architecture of adder and multiplier, where speed is increased by reducing the number of gates involved in designing architecture of adder and multiplier consumption analysis is done using XPower analyzer.

Keywords Linear feedback shift register (LFSR) · Built-in self-test (BIST) Circuit under test (CUT) · Automatic test equipment (ATE) · Signature analyzer (SA) · Test pattern generation (TPG) · Artificial intelligence (AI)

1 Introduction

In the present electronic era, speed plays a very crucial role. Speed size and cost are the present keywords in the electronic industry. Any designer will have to take care of all the three factors considered into account before starting any design. The trend nowadays is miniaturization coupled with fast access time. In digital adders, time taken to propagate the carry from one stage to another and then to addition in next stage depends on previous generated carry; thus, the delay in carry generation increases as the number of addition bits increases. Thus, an attempt is made to reduce the delay in carry generation of adders, and the same adder is used to design the multiplier. The technique used here uses three RAMs designated as RAM-A, RAM-B, and RAM-C, and the address generator is designed using up counter, multiplier, and a carry-select adder. Carry-select adder (CSA) is one of the fastest adders used in many data-processing processors to perform faster arithmetic

S. B. Shirol (✉) · S. Ramakrishna · R. B. Shettar
Department of Electronics & Communication Engineering, B V Bhoomaraddi College of Engineering & Technology, Hubballi 580031, Karnataka, India
e-mail: suhasshirol@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_2

operations. The carry-select adder is designed by using a binary-to-excess-1 converter, and even the Vedic multiplier is also designed using CSA binary-to-excess-1 converter which reduces the delay and number of gates in the design [1, 2].

2 Adders

Adders are the digital circuits which are used to add two numbers. Adders play a very important role in many digital design applications in case of arithmetic operations, and in many more digital design applications, adders are key ingredients in digital design systems. We have many different adders which are used depending upon the system specifications and based on better performance. Some adders are like ripple-carry adder, carry-lookahead adder, BCD adder, carry-save adder, carry-skip adder, and carry-select adder. In case of ripple-carry adder, the delay in carry generation increases as the size of the adder increases and also the addition of next bits depends on carry of the previous bit, thus the delay in carry generation and the addition also increases. In other adders, also the delay in carry generation makes the adder operation slow, thus many researches are going to reduce the delay in adders. Nowadays, the area, speed, and cost are the main buzz words in electronic industries, thus designing an efficient adder with reduced delay, area, and power is the trending research [3].

2.1 Ripple-Carry Adder

Ripple-carry adder is one of the types of adder which is used to add binary numbers. The ripple-carry adder is designed using full adders; we can design higher sized adders same as 4-bit adder where the structure increases with number of full adders. And as from the below block diagram, it is evident that the structure involves huge delay in addition of numbers. As the current bits needs the carry generated from previous bits thus which makes larger delay in carry generation as the size of the adder increases. Cin will be assigned to 0, S is the sum output bits of the two numbers. And cout is the final carry out of the ripple-carry adder. As this ripple-carry adder involves huge delay, it will not be reliable to use this adder where we need maximum speed in circuit operation. Thus, many adders have been designed after this, but the carry-select adder is one of the most reliable adders which is discussed below.

2.2 Conventional Carry-Select Adder

Carry-select adder is one of the fastest adders where the addition of the 2 block digits does not depend on previous carry, thus the delay could be reduced in this adder. Here the addition of the block bits will be carried out for both cin = 0 and cin = 1 condition, and the sum of the output will be selected using mux depending on the previous block carry; thus, the delay will be reduced compared to ripple-carry adder and carry-lookahead adder. But the carry-select adder involves many ripple-carry adder blocks for both cin = 0 and cin = 1 condition; thus, the structure is quite complicated and bulky, and in it still can be reduced. Figure 1 shows 16-bit conventional carry-select adder.

2.3 Carry-Select Adder-BEC

The conventional carry-select adder can still be reduced in order to reduce the design complexity and number gates in designing, which in turn reduces delay. Thus, an attempt is made to reduce the number of gates in designing, i.e., the conventional carry-select adder involves many ripple-carry blocks for both cin = 0 and cin = 1 condition which can be reduced by replacing ripple-carry adder block when cin = 1 by binary-to-excess-1 converters which reduces the number of gates involved in designing and reduces the delay. The below block diagram shows the carry-select adder using binary-to-excess-1 converter which is used in our project to implement the main block of our design, i.e., for matrix multiplication. Thus, the above study shows that the carry-select adder using binary-to-excess-1 converter is one of the fastest adder and which has been used to implement our project design. Thus, this 16-bit adder design has been implemented in verilog HDL and Spartan-6 FPGA using ChipScope Pro. Figure 2 shows 16-bit conventional carry-select adder using BEC.

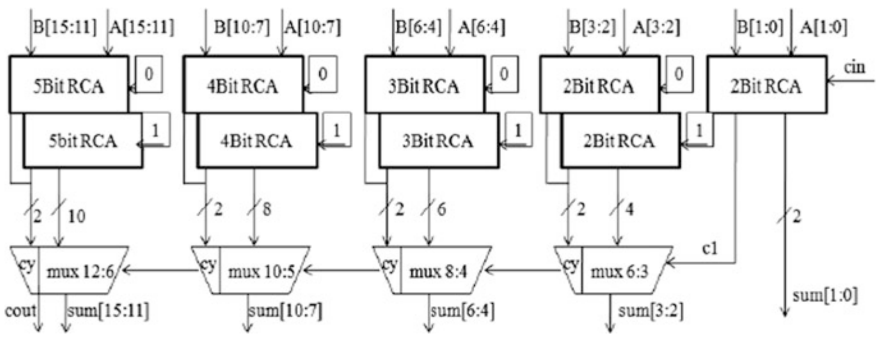


Fig. 1 16-bit conventional carry-select adder

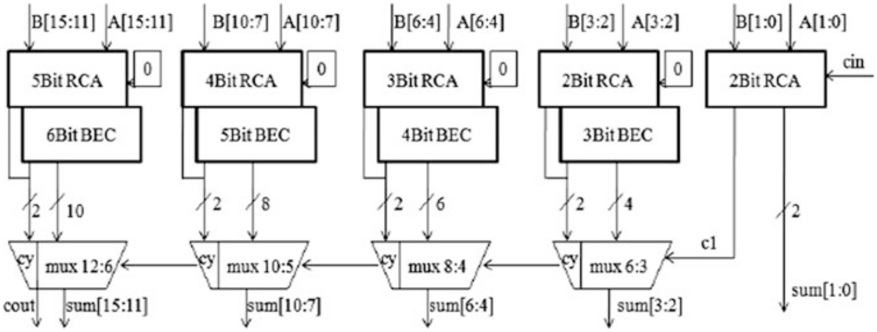


Fig. 2 16-bit conventional carry-select adder using BEC

3 Multiplier

Multipliers are the digital circuits which are used to multiply the binary numbers and are used in many areas of digital design like in multiply accumulator unit, arithmetic and logical unit, and in filters and in many more areas to multiply the binary numbers. There are many types of multipliers such as array multiplier, Vedic multiplier, booth multiplier. But the Vedic multiplier is reliable and efficient multiplier for our design. Thus in our design, an 8-bit input and 16-bit output Vedic multiplier has been designed using carry-select adder binary-to-excess-1 converter [4].

3.1 Vedic Multiplier

The Vedic multiplication is the ancient type of multiplication method, but it is the most accurate type of multiplication method. There are two types of Vedic multiplication sutras [5].

- Urdhva Tiryakbhyam sutra
- Nikhilam sutra

We have used Urdhva Tiryakbhyam sutra for the Vedic multiplier design. The 8-bit input and 16-bit output Vedic multiplier is designed using Urdhva Tiryakbhyam sutra, whereas the carry-select adder binary-to-excess-1 converter is used in this design for addition of partial products. It reduces the delay of the multiplier. The multiplier has been designed using verilog HDL and implemented on Spartan-6 FPGA using ChipScope Pro.

3.2 Matrix Multiplier

Matrix multiplication is most commonly used operation. The matrix multiplication consumes more time in digital designs for the processing, thus an attempt is made to reduce the delay in matrix multiplication. We have different matrix multiplication algorithms, and if we are multiplying N × N matrix then the standard matrix multiplication involves N3 operations and its the accurate one and Strassen's algorithm and coppersmith and Winograd algorithms involves $N^{2.81}$ and 2^2 operations, but the accuracy in the data generated will be less. Thus as our system involves more accuracy toward the matrix multiplication, we are using standard matrix multiplication algorithm for our design [6].

Here it uses carry-select adder binary-to-excess-1 converter (CSA-BEC). The multiplier is designed by using CSA-BEC which reduces the delay of carry generation in the addition of partial products of the multiplied numbers. The matrix A is stored in RAM-A and matrix B in RAM-B which is fetched using decoder and given to the matrix multiplier. Let's take a 2 × 2 matrix multiplication as example which involves matrix A, matrix B as given below. Here the matrix multiplication involves two input matrices and one output matrix, i.e., Mat-A, Mat-B, and Mat-Y. The inputs of the matrices are stored in RAM-A, RAM-B depending upon the address generated by the address generator. And after matrix multiplication, the outputs generated will be stored in RAM-C. Matrix operations are most commonly used in many applications like digital signal processing, image processing, and graph theory in arithmetic and logical units and in many more applications. The matrix operations involve multiple additions and multiplications. We have implemented 2 × 2, 4 × 4, and 8 × 8 matrix multiplication using verilog HDL language, and simulated and synthesized using Xilinx 14.2 ISE simulator software. A single-bit matrix multiplication is designed using above formulae and the same unit is used for other bit generation.

Design and Implementation

The standard algorithm used for matrix multiplication consumes most of time, as it involves reading data from memory, multiplying data, and again storing the result back in memory. And also computation in the adder while multiplication using adders will also take lot of time for computation which must be reduced. The method used here involves memory units, multiplier is designed by using carry-select adder-BEC, and a 16-bit carry-select adder is used which reduces the area and delay [7, 8].

The block diagram of our design consists of three memory modules. The matrix A and matrix B are stored in RAM-A and RAM-B, respectively, which depend on the address line. These matrix inputs are used for matrix multiplication. The carry-select adder is designed by using BEC which reduces the delay and number of gates. And the block diagram is as shown in Fig. 3.

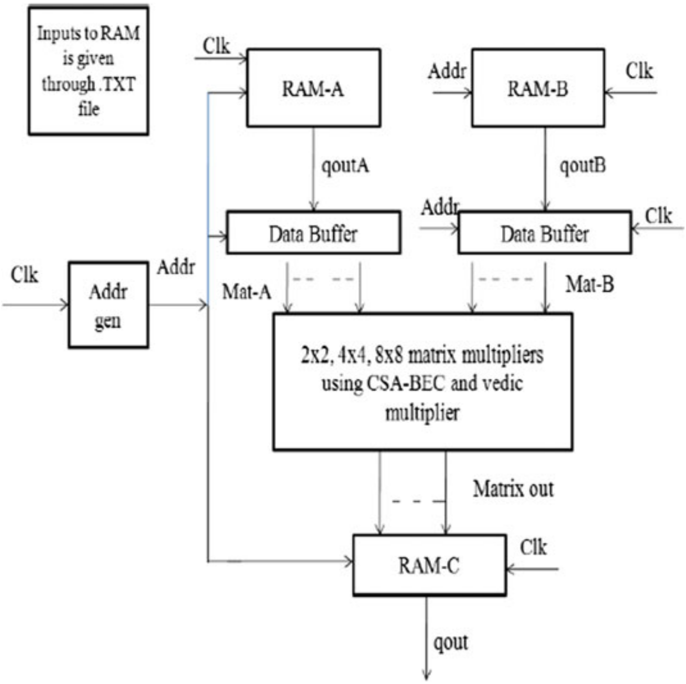


Fig. 3 Block diagram of matrix multiplication

4 Results

The design is done by using Verilog HDL code, and the simulations and synthesis are performed by using Xilinx 14.2 ISE simulator using hardware description language (verilog). Adders and multipliers are implemented on Spartan-6 FPGA using ChipScope Pro. The results are given in below Sects. 4.1 Adders Analysis. The result analysis of different adders is shown in Table 1.

The path delay obtained by using conventional carry-select adder is more compared to path delay obtained by carry-select adder using BEC. Thus, this adder reduces the carry generation delay in matrix multiplication and also reduces the gate count and area in the design.

4.1 Multiplier Analysis

The result analysis of different multiplier is shown in Table 2.

As the CSA-BEC is with reduced delay, the Vedic multiplier is designed using the same and the delay of other multiplier is compared.

Table 1 Adder analysis 16-bits

S. no.	Adders	Previous path delay (ns)	Proposed path delay (ns)
1	Ripple-carry adder	39.23	13.13
2	Carry-lookahead adder	28.98	12.83
3	Conventional carry-select adder	28.64	5.90
4	Carry-select adder using BEC	5.96	4.67

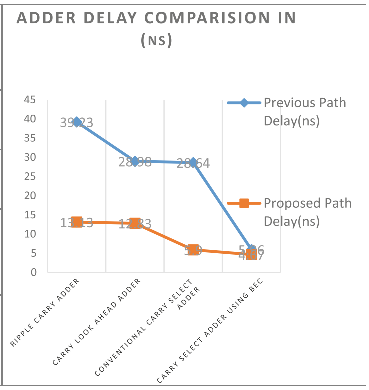
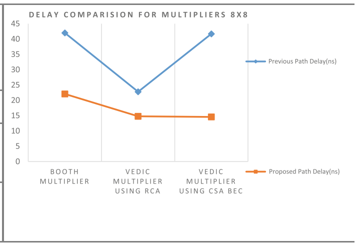


Table 2 Multiplier analysis

S. no.	Multiplier	Previous path delay (ns)	Proposed path delay (ns)
1	Booth multiplier	42	22.07
2	Vedic multiplier using RCA	22.78	14.77
3	Vedic multiplier using CSA-BEC	41.69	14.56



4.2 Matrix Multiplication Design Analysis

The result analysis of matrix multiplication for 2×2, 4×4, and 8×8 is shown in Table 3.

Our design has been implemented using verilog HDL and simulated in Xilinx 14.2 ISE simulator, and after the simulation and synthesis, the generated synthesis report shows the below maximum obtained path delays.

Table 3 Matrix multiplication design analysis

S. no.	Order of matrix multiplication	Previous path delay (ns)	Proposed path delay (ns)
1	2×2	5.13	5.28
2	4×4	9.94	7.73
3	8×8	15.40	15.08

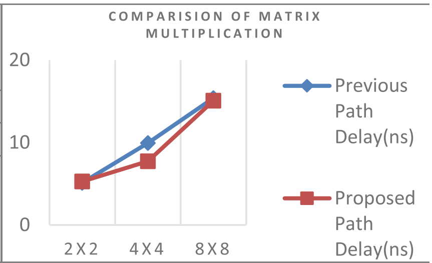
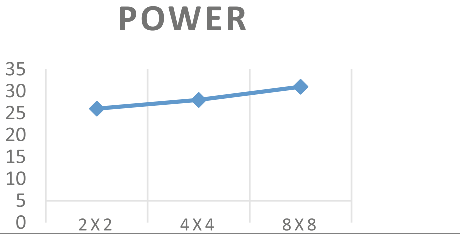
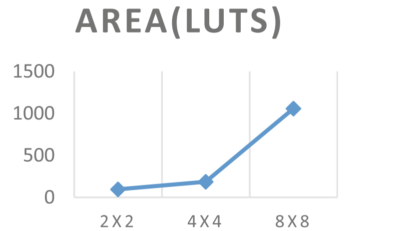


Table 4 Power and area analysis for different order of matrix

S. no.	Order of matrix multiplication	Power (mW)	Area (LUTs)
1	2 × 2	26	96
2	4 × 4	28	185
3	8 × 8	31	1058





Power Analysis

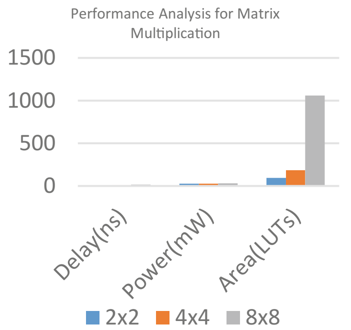
Our design has been implemented using verilog HDL and simulated in Xilinx 14.2 ISE simulator, and after the simulation and synthesis, the power dissipation is calculated for the system by using XPower analyzer which is shown in Table 4.

Area Analysis

Our design has been implemented using verilog HDL and simulated in Xilinx 14.2 ISE simulator, and after the simulation and synthesis, the synthesis report shows the utilization summary where the used LUTs show the area in Table 4.

Performance analysis for the different matrix multiplication

S. no.	Order of matrix multiplication	Delay (ns)	Power (mW)	Area (LUTs)
1	2 × 2	5.28	26	96
2	4 × 4	7.73	28	185
3	8 × 8	15.08	31	1058



Future Scope

This design can be used for signed numbers matrix multiplication by using booth multiplication algorithms in the design.

References

1. Ramkumar, B., Kittur, H.M.: Low-power and area-efficient carry select adder. IEEE Trans. Very Large Scale Integr. (VLSI) Syst. **20**(2) (2012)
2. Mohanty, B.K., Senior Member, IEEE, Patel, S.K.: Area delay power efficient carry-select adder. IEEE Trans. Circuits Syst. II: Express Briefs **61**(6) (2014)
3. Elangadi, S., Shirol, S.: Design and characterization of high speed carry select adder. Int. J. Ethics Eng. Manag. Educ. **1**(6) (2014)
4. Jais, A., Palsodkar, P., SM-IEEE: Design and Implementation of 64 Bit Multiplier Using Vedic Algorithm. IEEE, 6–8 April 2016
5. Saha, P., Banerjee, A., Bhattacharyya, P., Dandapat, A.: High Speed ASIC Design of Complex Multiplier Using Vedic Mathematics. IEEE (2011)
6. Jain, S., Kumar, N., Singh, J., Tiwari, M.: High speed matrix multiplication implementation using field programmable gate array. Int. J. Eng. Trends Technol. **7**(2) (2014)
7. Aslan, S., Desmouliers, C., Oruklu, E., Saniie, J.: High performance reconfigurable pipelined matrix multiplication module designer. IEEE **10**(1) (2010)
8. Ghissoni, S., Costa, E., Monteiro, J., Reis, R.: Combination of constant matrix multiplication and gate-level approaches for area and power efficient hybrid radix-2 DIT FFT realization. IEEE **8**(3) (2011)

Signaling Protocol Specification for Signaling Approach-Based Virtual Output Queue Router Architecture

Jaya R. Surywanshi and Dinesh V. Padole

Abstract Network communication protocol is a set of rules used by the network designer while designing a packet in computer network. It is a mechanism through which the network and users are connected together and designer can easily expand and manage the capability of network. In this paper, work is presented on a new communication protocol with signaling mechanisms. This protocol is used for a new Virtual Output Queue (VoQ) router architecture by considering 2D mesh topology to support Quality of Service (QoS) requirements between all hardware resources in Network-on-Chip (NoC). This system will be responsible for forwarding the instantaneous information between the neighbor's routers. With this new protocol, each router will get the present status information of its neighbor's router in advance. Major issues of routing like static routing and congestion in the routing path will try to avoid by using this protocol.

Keywords Network communication protocol · NoC · QoS · SoC
VoQ · Router · Signaling block

1 Introduction

The field of electronics is widely developing in modern times. With advancement in technology, the size of electronics has reduced numerous fields. Electronics these days have become much more compact accommodating a ton of services in a tiny area. The integrated architecture is facilitated things. All the different components are developed as a single chip according to the requirement of the application. According to predictions as per Moore's law, the density of an integrated circuit roughly doubles with every passing year [1]. The term System-on-Chip (SoC) is

J. R. Surywanshi (✉) · D. V. Padole
G.H. Raisoni College of Engineering, Nagpur, India
e-mail: abhijaya19@gmail.com

D. V. Padole
e-mail: dvpadole@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_3

highly popular for compact systems and fundamentally used to refer the chip which has the integration of software with on-chip silicon hardware and is specially designed so as to perform specific functions. The integrated chip which is a system on the chip may have digital signals, analog signals, mixed signals, or even radio frequency functions—all capable from a single chip developed on a single substrate—thereby making it compact as well as functional. They are widely popular in a variety of applications; the most common of it being smart phones. They are also found in digital television sets, wearable computing devices, and many others [2].

The popularity of SoC could be attributed to the variety of applications they offer on a single integrated module. However, there are certain drawbacks for the SoC such as unpredictable performance, scaling problems, and reduced efficiency to name a few. There have been numerous researches undertaken globally to tackle these issues and improve performance [3]. NoC is generally used to interconnect multi-cores in SoC. The router is the prime factor of NoC architecture. The foremost dispute in these NoCs routers is related to the QoS functionalities like better bandwidth, inadequate transaction times, the rate of losses, and low congestion. Throughput, latency or delay, errors, bandwidth, dropped packets, jitter or variation in delay, packet loss, and out-of-order packet delivery are certain parameters that are to be considered while designing the NoC for better QoS. The goal of QoS is to provide guarantees on the ability of a network to deliver predictable performances [4]. A defined QoS may be required for some types of network real-time traffic or multimedia application [5–7]. To achieve the good performance of a network, it is essential to down the consistency of these problems. Regarding these issues, the main work presents here is signaling approach specification. Presenting a new protocol through which each router will get the neighboring router status information in advance. Signaling mechanism is proposed in order to get router current status information. Router connected by the 2D mesh architecture and wormhole switching techniques for NoC. A higher rate of packet loss and congestion problem is presented in 2D mesh architecture, with the wormhole switching techniques and static routing algorithm. But the advantages of the wormhole routing technique is that it has a small buffer requirements in the routers, and it also reduces latency parameter [8, 9]. The flow of this paper is as follows: After the introduction, the proposed signaling approach-based new Virtual Output Queue router architecture is explained in Sect. 2; Sect. 3 presents singling protocol specification for the proposed router and at the last result and conclusions.

2 Router Architecture

The router is the most significant and integral part of the NoC system. The function of the router in a network is to direct the traffic from the source to destination. To facilitate the same, it coordinates the flow of data, which is considered as a very crucial step in communication networks. Thus, it can be rightly considered the backbone of the communication taking place in a NoC system. Keeping in mind the importance of a router, it should be designed so that it provides maximum efficiency and throughput.

Fundamentally, basic blocks of the architecture of a router are:

1. One input port
2. One output port
3. A switching matrix in order to connect the input port and the output port
4. A local port which helps in the connection of the router to the corresponding IP core.

These routers can be deemed as intelligent devices because they receive incoming data packets, then they inspect their destination and finally, and they figure out the best path (shortest and most efficient) for the data to transfer from the source to destination. The architectural design of the router is a significant factor in the determination of the critical path delay of a router; this affects per-hop delay as well as network latency. Thus, the architectural design of the router should be carefully planned to meet the required latency and throughput requests even in the tight constraints of area and power. This router design efficiency is hence the deterministic factor when considering the overall performance of the network [10].

The router design in the proposed method will be based as per the following stages like router function, the input controller, switch, and switch allocator. The router function mainly determines the routing path for message transfer. The other stages are employed to initiate the routing process. The output queuing signaling scheme depends on router architecture. The router includes parts such as input ports [I1…I5], output ports [O1…O5], crossbar unit [S1…S5], and the arbitration unit [A1…A5]. The control logic consists of routing logic which is used to check for the availability of buffers in the adjacent router before switching packets. The output port connected with arbiter which sends out packets using round-robin arbitration scheme. Figure 1 shows the proposed router architecture with signaling mechanism. Here each port informs an arbiter of the output queues that it has available through signaling scheme. Then each port generates an additional input queue for each possible output queue reported from the previous step. When a frame is ready to be switched between ports, the input port queues the frame in a Virtual Output Queuing. Then the input port requests to transmit the frame over the crossbar to the output port. When the output queue is available, the arbiter allows the request from the input port and takes one record from the available output queues, and then the frame is transmitted. Once the output port transmits the frame, then it notifies the arbiter that the outlet queue is now available.

Table 1 represents the detail specification of router architecture. Internal architecture consists of five ports (input/output), arbiter block, routing block, and signaling block. The arbiter block categorizes the flits coming from the ports. The signaling block is responsible for the generation of signaling flit to update its neighbor's routers current state. Also, this block collects the information of the neighbor's routers which is updated in case of incoming signaling flit. These stored states are required at the header flit management during the path establishment

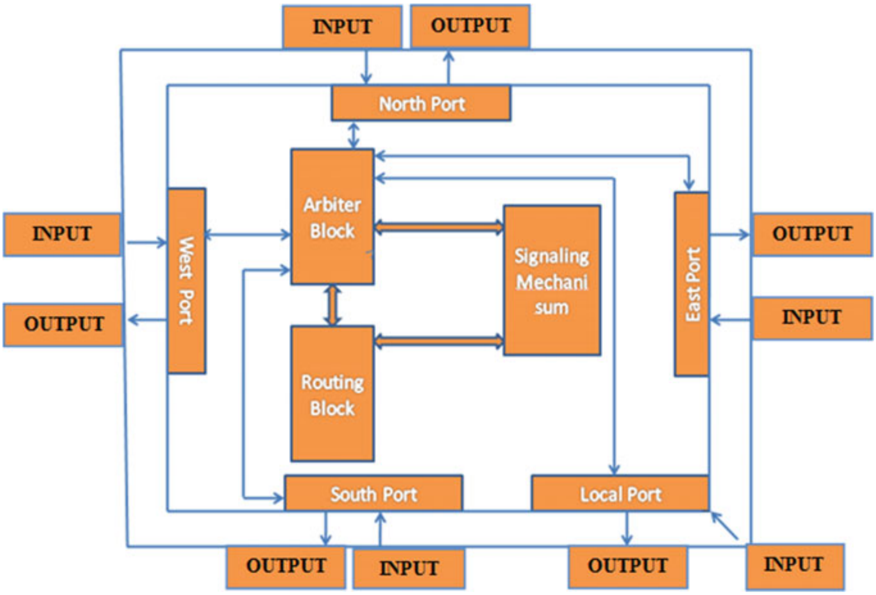


Fig. 1 Router architecture with signaling block

Table 1 Proposed router architecture specifications

Sr. no	Components	Specifications
1.	Topology	4 × 4 2D topology
2.	Routing algorithm	Dy XY routing
3.	packet switching	Wormhole packet switching
4.	Flit size	32 bits
5.	Packet size	16 flits
6.	Link width	32 bits
7.	No of virtual channel	4/port
8.	Depth of virtual channel	4 flits/VC
9.	Width of the communication port	32 bits
10.	Channel width of the network	32 bits

processes. Finally, this block is cooperating with the routing block during data flit routing. Routing block is the main block of the router. It ensures the routing processes of the different incoming flits. This block uses the wormhole technique and a Dynamic X-Y routing algorithm basing on signaling information in order to find the best path for the data processes.

The new signaling feature we have introduced inside the router. This is responsible to generate and broadcast the signaling information in order to update neighboring routers present states with the new traffic load states. This block will analyze newly received signaling flits and update the internal routing table information. Routing decision is depended on the information stored in the routing table. This block collects congestion information from neighbor port by checking status of Virtual Output Queues. If the buffer has free space for storing packets, it means no congestion is there; otherwise congestion is present. All neighbor blocks send its current status information to the signaling block. So this signaling block has the information of all neighboring router. This is required to manage a head flit during the path establishment process. Signaling block gets the free buffer space information from all neighbors' routers and forms signaling flit. This flit is nothing but the current status of neighbor's router. Every time the signaling block communicates with the routing block during data flit routing. This signaling block connected to the arbiter block and the routing block. This information is then forward to the routing block. The routing block then decided the new route to forward the packets. It always prefers the free route where the no congestion is present.

3 Signaling Protocol Specification

Singling-based communication protocol structure is defined in two major flit types: data flit and signaling flit. Flit size is 32 bits fixed for all flit to maintain the scalability and matching the hardware data bus and for easy implementation of internal router buffer constraints. Data flits are used to send the actual data between processor. For the communication process between processors, three basic types of data flit are required:

- Head data flit: The wormhole commutation technique is based on the path establishment between the source and destination ports. This information find out the free memory space in router and the waiting flits length in-queue, is provided through the signaling process between neighboring routers.
- Communication data flit: Communication data flit type is processed after the communication path establishment between source port and destination port. It transports the actual data payload. The process identifier (Proc ID) role is important here to identify the output port of the flit. Proc ID uses here a local switching table present inside router and regularly update at the start and end of every process.
- Signaling flit: Signaling flit generates the signal for congestion information on the basis of current buffer status of router and broadcasts this information to another router through packet routing mechanism. Signaling flits hold useful information of neighboring routers in order to provide present status of each router with a virtual representation of its neighbors. It has generated by a router when it changes its status. Messages are mainly divided into three flits, and they are encapsulated in packets. The main structure of these contents is head flit, body flit, and last tail flit. In wormhole techniques, buffers and physical channels are allocated to flits instead of packets [11].

(a)

00	Header
01	Data
10	EOP
11	Signaling

(b)

Type	@S	@D	Ser ID	Flit No	Proc ID	Port ID	Reserve

(c)

Type	Flit ID	Proc ID	Data

(d)

Type	Sw ID	State_R	Sch Tech	Proc No	Reverse

(e)

Type	Flit ID	Proc ID	EOP

Frame structure details.
Type: Flit type
@S : Source address
@D : Destination address
Ser ID : Service Identifier
Flit No: Flit number
Proc ID: Process identifier
Port ID: Port ID
Flit ID: Flit ID
Sw ID: Switch identifier
State_R: State of router
Sch Tech : Scheduling Techniques
Proc No : Process No
Data: Payload data
EOP : End of the Process

Fig. 2 **a** The flit types, **b** header flit structure, **c** communication data flit structure, **d** signaling flit structure, **e** end of packet flit

The packet frame structure is presented here. The first prefix informs the type of flit. It represents by 2 bits given in the Fig. 2a. Header flit specification has defined in Fig. 2b. Header flit carries the information of Source address, Destination address, Service ID, the flits number of the packet, the Process identifier, port ID, and communication process identifier. A field port identifier (Port ID) can distinguish the input port while routing. Figure 2c communication data flit structure which holding the information of actual data payload. Actual data carries through these flits. The data flit will simply fallow the same path, routed by the header flit.

A data flit contains Flit ID, Process ID, and actual information in data form. The details of structure of signaling flit are given in Fig. 2d. This carries the information of routing status. The field required to this flit is switch identifier (Swh id), status of the router, the scheduling technique (Sch Tech) used, and process number (Proc no). Last flit is the end of packet flit (EOP) shown in Fig. 2e. It has the similar structure as the data flit, but it is used to represent the end for current communication processes.

4 Results and Discussion

Considering the router as the prime factor of NoC, the study is undertaken to improve the Quality of Service in NoC architecture. The study adopts a Virtual Output Queuing signaling scheme that depends on router architecture. The same is implemented with software, and the results of the same are discussed. The performance factors like throughput and latency are estimated to analyze the effectiveness of our proposed research and will be compared with existing techniques. Xilinx 13.1 simulator is used for implementation of this router. Table 2 shows the primary-level results of implementation. 500 MHz frequency is considered for 99% throughput where the packet injection rate is 0.1 flit/node. The proposed protocol tested with proposed router with different routing algorithms. 'From obtained results, Dy-XY algorithm is having the best latency results as compared to others.

Comparison of proposed work results with others results is concluded in the Table 3. After the comparison, we can say that we have good results as compared to others.

Table 2 Throughput and latency results for different routing algorithms

Sr. no	Topology	Routing algorithm	Average latency in clocks	Average throughput in %
1	2D Mesh	XY-Routing algorithm	9	78.8
2	2D Mesh	IX/Y (XY and X) algorithm	8	81.8
3	2D Mesh	XYX	7	80.6
4	**2D Mesh**	**DyXY**	**3**	**99.4**

Table 3 Comparison of results

Sr. no	Title	Work done	Operating frequency (MHz)	Latency in clock cycles
1	HIBI communication network for System-on-Chip	Support for hierarchical topologies with several clock domains, flexible scalability, and runtime configuration of network parameters	435	4 clock cycles
2	An efficient on-chip NI offering guaranteed services, shared-memory abstraction, and flexible network configuration	Use a transaction-based protocol to achieve backward compatibility with existing bus protocols such as AXI, OCP, and DTL	500	4 clock cycles
3	Benchmarking mesh and hierarchical bus networks in System-on-Chip context	Worked on hierarchical bus mechanism	435	4 clock cycles
4	Results of the proposed system	Signaling protocol mechanism for Virtual Output Queue router architecture	500	3 clock cycles

5 Conclusion

Signaling protocol which proposed here will clear the idea how the flit can deliver the congestion information to each router. The congestion information is available in advance to each router so the router will find out the best route. This work is a contribution to the Quality of Service management in NoC based on a signaling approach mechanism between the neighbor's routers. The solution being developed is based on information exchanges in order to quantify router status to its neighbors. This information is useful to the path establishment between processors. The developed router architecture and designed protocol tested with different routing algorithm with 2D mesh topology. We found the best results. From the results, we can conclude that proposed protocol is more suitable for proposed router. Good results will automatically affect the performance of NoC. It will update the QoS for NoC at certain level.

References

1. Patti, R.S.: Three-dimensional integrated circuits and the future of system-on-chip designs. Proc. IEEE **94**(6), 1214–1224 (2006)
2. Wang, L.T., Stroud, C.E., Touba, N.A.: System-on-Chip Test Architectures: Nanometer Design for Testability. Morgan Kaufmann (2010)

3. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: vision, applications and research challenges. AdHoc Netw. **10**(7), 1497–1516 (2012)
4. Nasri, S.: New approach of QoS metric modeling on network on chip. Int. J. Commun. Netw. Syst. Sci. **4**, 351–355 (2011). https://doi.org/10.4236/ijcns.2011.45040. Published Online May 2011. http://www.SciRP.org/journal/ijcns
5. Liu, Y., He, H.: Grid service selection using QoS model. In: Third International Conference on Semantics, Knowledge and Grid, pp. 576–577 (2007)
6. Luo, J., Jiang, L., He, C.: Finite analysis for energy and QoS tradeoff in contention-based wireless sensor networks. In: International Conference on Communication (ICC). IEEE Communications Society, pp. 1–6 (2007)
7. Kunavut, K., Sanguankotchakorn, T.: Multi-constrained path (MCP) QoS routing in OLSR based on multiple additive QoS metrics. In: InterSium on Communications and Information Technologies (ISCIT-IEEE), pp. 226–231 (2010)
8. Karim, F., Nguyen, A., Dey, S., Rao, R.: On-chip communication architecture for OC-768 network processors. In: DAC01 Proceedings of the 38th Annual Design Automation Conference, DAC 2001, pp. 678–683, 18–22 June 2001
9. Moraes, F., et al.: HERMES: an infrastructure for low area overhead packet-switching networks on chip. Integr. VLSI J. **38**(1), 69–93 (2004). www.Elsevier.Computer.com
10. Swapna, S.: Efficient router design for network on chip (2013)
11. Dally, W.J., Seitz, C.L.: The torus routing chip. J. Distrib. Comput. 187–196 (1986)
12. Salminen, E., Kangas, T., Hama La Inen, T.D.: HIBI communication network for system-on-chip. J. VLSI Signal Process. **43**, 185–205 (2006)
13. Adulescu, A.R., Dielissen, J., Pestana, S.G., Gangwal, O.P., Rijpkema, E., Wielage, P., Goossens, K.: An efficient on-chip NI offering guaranteed services, shared-memory abstraction, and flexible network configuration. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. **24**(1) (2005)
14. Salminen, E., Kangas, T., Lahtinen, V., Riihima ki, J., Kuusilinna, K., Timo D.: Benchmarking mesh and hierarchical bus networks in system-on-chip context. J. Syst. Archit. (Elsevier) **53,** 477–488 (2007)

Prediction-Based Parallel Clustering Algorithm for M-Commerce

Likhesh N. Kolhe, Vaishali Khairnar and Ashok Kumar Jetawat

Abstract A rapid increase of mobile commerce (M-Commerce) with sensing devices has resulted in enormous services from different service providers. M-Commerce services provide numerous ranges of emerging services. Also, different qualitative matrices are provided with similar functionality. Automatically the service flow is combined with other services. M-Commerce stakeholders are ambient, dynamic in nature, which requires efficient techniques to enhance the output. Major challenge is to select appropriate optimization technique or algorithm to provide a numeric set of services with dynamic qualities. It is difficult to propose a method directly to predict M-Commerce. Hence, this research proposes a method of prediction in M-Commerce techniques proposed a prediction based parallel clustering algorithm using hybrid optimization technique for M-Commerce. Hybrid optimization technique or algorithm can be developed by applying cross-mutation technique in adaptive ant colony optimization with particle swarm optimization to improve the efficiency and throughput of the system in M-Commerce. To predict the optimum service it runs in parallel using MapReduce on a Hadoop platform. Parallel processing services reduce the time factor, which is essential for processing the massive amount of unstructured data in a mobile environment. Relevancy, correctability of this proposed system would be validated through simulation and modeling on real-time existing data sets.

Keywords MapReduce · Particle swarm optimization · Adaptive ant colony optimization · M-Commerce

L. N. Kolhe (✉) · A. K. Jetawat
Pacific Academy of Higher Education and Research University, Udaipur, India
e-mail: likhesh8@gmail.com

V. Khairnar
Terna Engineering College, Navi Mumbai, Maharashtra, India

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_4

1 Introduction

Clustering algorithms have been employed in various fields like medicine, manufacturing, nuclear science, and radar scanning. Data clustering algorithms may be partition or hierarchal. In hierarchal clustering algorithms, the data are not clustered into particular cluster in one step [1]. Instead, a sequence of partitions can occur from single cluster having all objects to 'n' number of clusters each having a single object. The partition clustering algorithms attempt directly to decompose the dataset into a disjoint cluster's set [2]. The criterion function in which the clustering algorithm attempts to reduce the local structure of data by allocating clusters to the peaks in global structure or probability density function. The global criteria can reduce some dissimilarity measures in samples inside each cluster while reducing the dissimilarity of various clusters. To enhance the result of M-Commerce, the basic particle swarm optimization algorithm (PSO) technique is use to solve problem of wingman optimal control should be rectified in order to make sure a fast, autonomous solution or result to the hybrid optimization technique. The application of hybrid optimization technique includes: forming rapid, feasible as well as autonomous solutions to solve the optimal control problems. The revelation of this hybrid optimization technique through PSO algorithm with two-dimensional model in static threat atmosphere gives assurance in the method when employed to continuing future research.

Internet has modified the way in which people think, communicate, and do business with one another. The Internet-based models for business to customer e-commerce are the best application of e-commerce, which permits firms to conduct business trade-off with customers over the Internet-based information systems like portals, Web-based systems, and shopping malls [3]. A hybrid optimization technique using genetic algorithms as the prediction of controls with the help of system environment. The genetic algorithm is used as the pattern-directed search method in order to evaluate the independent variables exponent. The hybrid optimization technique combines two or more algorithms for attaining the solutions [4].

For the enhancement of the research work, here MapReduce technique is amalgamated with particle swarm optimization technique to extract the relevant results from the datasets. In the first phase, data extraction is taken place with the help of Apache Flume. For data extraction, Amazon API is used to extract results from amazon Web site or the same concept can be applied to existing M-Commerce datasets also. In the second phase, PCHOT algorithm is applied to extract the final result.

MapReduce is a framework used for managing large amount of data, in parallel or in large clusters in reliable manners. As the name suggests, in MapReduce, first the data are mapped with resources, and then by performing operations, it reduces into final result (Fig. 1).

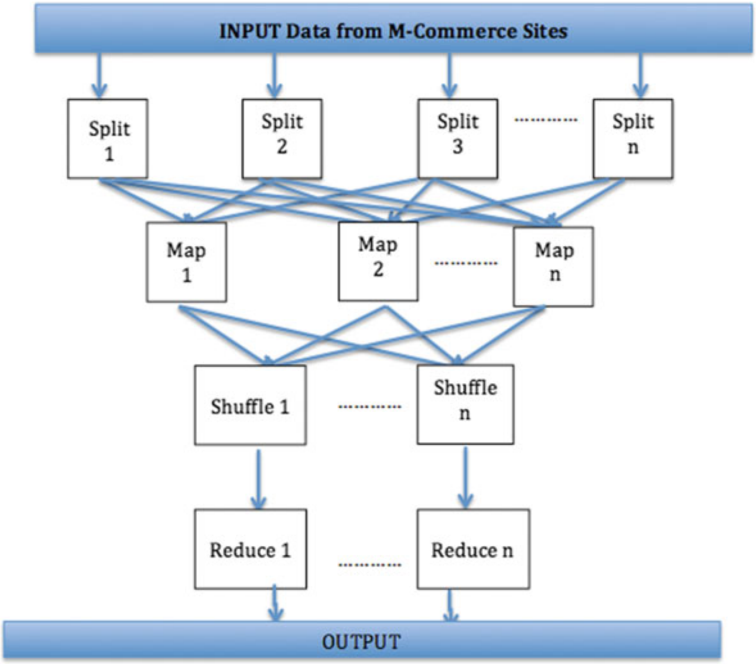


Fig. 1 Framework for MapReduce methodology

2 Related Work

2.1 M-Commerce Using Parallel Clustering Algorithms

Avasthi and Dwivedi [5] carried out the prediction of mobile users' behavior with the help of clustering algorithms. Temporal property and user relations were employed in this paper. The CTMSP-mine technique was employed to extract the users' behaviors. CO-Smart-CAST has created the user clusters as well as similarities among the mobile users. The CTMSP-mine algorithm has used both the results of time intervals and CO-Smart-CAST. These patterns were employed to forecast the future behavior of mobile users and then service recommendations [5].

Evangeline et al. [6] presented the analysis on service pattern prediction and mining the personal mobile commerce environment using 2-DML-association rule mining technique. This research has performed theoretical analysis on the mining as well as prediction of patterns in the mobile commerce environment. In this research, the 2-DML-association rule mining technique was employed to determine the location-based service patterns in the mobile commerce. The result of this algorithm has revealed the best performance efficiency as well as memory efficiency [6].

Papamichail et al. [7] analyzed the modified data clustering in electronic commerce using k-means range algorithms. This research has employed the k-means range algorithm and a mixture of partition k-means clustering technique together with a familiar spatial data structure that is range tree, which permits quick range searches. It has provided a synchronized result for the growth of distributed interactive decision that helps in electronic commerce. Since, it permits the customers to model their better choices along with many dimensions, explore for product information and form the products' data clusters that retrieved to improve their purchase decisions. This research has also described the implications of both customer as well as online store approach [7].

Priyadharshini and Kalimuthu [8] determined an effective framework for forecasting the mobile commerce patterns using graph diffusion technique. The problem of mobile commerce pattern mine technique has been overcome through the effective framework with the help of graph diffusion technique. The main intention of this research was to build the graph for the purchased items through mobile users and after that to determine the repeatedly purchased items. With the help of ranking technique, the researcher has ranked the number of items depending upon their trade-off or transactions. By estimating the behavior of mobile users as well as recommended the ranked items, this framework has produced more effective and precise item recommendation when compared to mobile commerce explorer (MCE) framework [8].

3 Research Gap

From the above literature survey, it is found that several applications were proposed for M-Commerce. At the same time, there were no studies conducted to propose a method for the prediction in I-Commerce. Hence, it is most important to conduct the research to identify a method for the prediction in mobile commerce.

It is difficult to propose a method directly to predict M-Commerce. **Hence, this research proposes a method of prediction in M-Commerce techniques by using prediction-based parallel clustering algorithm using hybrid optimization technique for M-Commerce**. This research will help the future researchers to conduct further research about the proposed topic. The gaps identified in the previous researches were addressed in the literature survey and explained it clearly. After analyzing the existing researches, this research aims to propose a prediction-based parallel clustering algorithm using hybrid optimization technique for M-Commerce.

4 Proposed Algorithms for M-Commerce

This research mainly concentrates on the prediction in M-Commerce. The proposed system is aiming at enhancing the performance efficiency of optimization in predicting the reviews of customers in Internet commerce. In this research, parallel clustering algorithm is used for developing the service composition. By applying this technique, some related services are chosen from the service collection. Many of the parallel clustering algorithms are employed as the message passing model in the master–slave architecture, and also, it may be used for effective computing. There is the amalgamation of parallel clustering algorithm and hybrid optimization techniques (adoptive ant colony optimization and particle swarm optimization) in proposed system.

4.1 Particle Swarm Optimization

The particle swarm optimization (PSO) is a calculative method that optimizes the problem by frequently developing a customer solution linking to a provided range of value. Each solution is indicated by particle's position vector. The particles travel in the space of multi-dimensional solution. After the iterations, the positions and velocities of the particles are altered and it attempts to move nearer to the point of optimum solutions. The PSO technique is simple to implement, and it has been widely employed in many real-world applications. There are some characteristics of PSO technique includes: each particle indicates a feasible solution; at the time of iteration, each particle requires to update the current position, local best location and current velocity; after each and every iterations, values of current position, velocity, and global best position may be altered; all particles present in the swarm have some best global position; and this best position is considered to be the optimal solution.

Pseudocode for PSO Algorithm

1. Input {starting values of the particles}
2. Output {optimum solution}
3. Initiate
4. for n = 1 to n^{total} do
5. for par = 1 to par^{total} do
6. // Update the particle
7. $VV^{new}{}_{par} = w_0 \circ VV_{par}$
8. $+ \, x_1 \circ R_1 \circ (PV^{pbest}{}_{par} - PV_{par})$
9. $+ \, x_2 \circ R_2 \circ (PV^{gbest} - PV_{par})$
10. $PV^{new}{}_{par} = PV_{par} + VV^{new}{}_{par}$
11. Update PV_{par} and VV_{par} by new values
12. fv_{par} = fitness (PV_{par})
13. // Update the particle's best fitness
14. if ($fv_{par} < fv^{pbest}{}_{par}$) then
15. $fv^{pbest}{}_{par} = fv_{par}, \; PV^{pbest}{}_{par} = PV_{par}$
16. end if

17. // Update the global best fitness
18. if ($fv_{par} < fv^{gbest}$) then
19. $fv^{gbest} = fv_{par}$,
20. $PV^{gbest} = PV_{par}$
21. end for
22. end for
23. end for
24. Optimum Solution = PV^{gbest}

4.2 Parallel Clustered PSO

There are two main operations such as Map and Reduce operations as follows:

$$\text{Map operations: Map}\,(j,\ v) \rightarrow \left[\left(j',v'\right)\right] \tag{4.1}$$

$$\text{Reduce operations: Reduce}\left(j',\ \left[v'\right]\right) \rightarrow \left[\left(j',v'\right)\right] \tag{4.2}$$

Features of PCPSO algorithm include:

- A particle indicates the feasible solutions.
- Hence, particle must contain position information of centroids.
- At the time of iteration, a MapReduce task is generated for every particle.
- par^{total} map tasks are generated to update the entire particles and estimate their fitness values.
- The reduce function arranges the entire updated particles that are output of different MapReduce tasks.

4.3 PCHOT

The hybrid optimization technique (HOT) is the best expansion technique of (ant system and particle swarm optimization), and it provides better capability and efficiency. There are some best features of PCHOT technique includes: a novel observation method for finding status is accepted to predict whether it is entrapping at most relevant output or not. Hence, the proposed technique is divided into two phases.

The two-phase methodology is used to determine the optimum service sets. These services include set of relative mobile services like big data storage, computation, and other capabilities related to the media service (such as social as well as entertainment), payment service, and messaging service are given by various service providers in Internet commerce environment. The parameters to be used in this proposed system are as follows: fitness function, computation time, and number of particles, number of services, number of iterations, and number of parallel clusters.

In the first phase, some relative services are chosen from service pool and customer reviews in Internet sites will be mined from both online system as well as mobile sites. It may be employed for the prediction to suggest the items or products to the users. In the second phase, the dataset is mined by parallel clustering algorithm with the help of modified fuzzy c-means technique integrated with hybrid optimization techniques in which adoptive ant colony optimization and particle swarm optimization techniques to improve the performance efficiency of the system in Internet commerce. The performance estimation is executed by measuring the recall, precision and prediction accuracy. Finally, the result will be compared with the existing system.

Commonly, the following factors are employed to simulate adoptive ant colony systems: path length and intensity of phenomenon. In order to choose the next path, the state alteration probability is measured as follows:

$$P_{xy} = \frac{\left(\tau_{xy}\right)'\gamma_x\left(\frac{1}{L_{xy}}\right)'\gamma_y}{\sum_{\substack{y=1 \\ y \neq x}}^{NA} \left(\tau_{xy}\right)'\gamma_x\left(\frac{1}{L_{xy}}\right)'\gamma_y} \tag{4.3}$$

where L_{xy} and τ_{xy} are the path length and pheromone intensity between nodes y and x, respectively. $'\gamma_x$ and $'\gamma_y$ are considered as control parameters for identifying the mass of path length and trail intensity, respectively. NA indicates number of ants. After choosing the next path, the trail intensity is updated as follows:

$$\tau_{xy}(\mathrm{m} + 1) = \mathrm{p}\tau_{xy}(\mathrm{m}) + \Delta\tau_{xy} \tag{4.4}$$

In the above equation, p is indicated as a coefficient of $(1 - \mathrm{p})$ indicates the trail between time m and m + 1 and $\Delta\tau_{xy}$ is considered as the quantity of pheromone trail included to τ_{xy} by ants.

Pseudocode for PCHOT

1. Intialize all the $\mathrm{par}^{\mathrm{total}}$ particles
2. using random values
3. for n = 1 to $\mathrm{n}^{\mathrm{max}}$ do
4. call MapReduce
5. // Spawn $\mathrm{par}^{\mathrm{total}}$ map tasks
6. Read the updated fv
7. from MapReduce output
8. for par = 1 to $\mathrm{par}^{\mathrm{total}}$ do
9. Apply hybrid optimization technique for population having a probability P_{xy}

10. Analyze prediction accuracy and relevancy
11. end if
12. end for
13. end for
14. Optimum solution = PV_{gbest}

In the above algorithm first initialized the particle value and taken random values. Then by applying MapReduce algorithm, we have developed a output as a population for further process. Then in next stage, we have applied hybrid optimization technique to get the probability, and in the final phase, we have analyzed prediction accuracy and precision and compared it with different algorithms.

5 Performance Analysis

The performance estimation is executed by measuring the

- Recall
- Precision
- Prediction accuracy

Comparison done between existing algorithms with proposed methodology we have found that as compared to adaptive ant colony optimization algorithm and particle swarm optimization techniques performance measure is more. When will perform cross-mutation over AACO and PSO (Fig. 2).

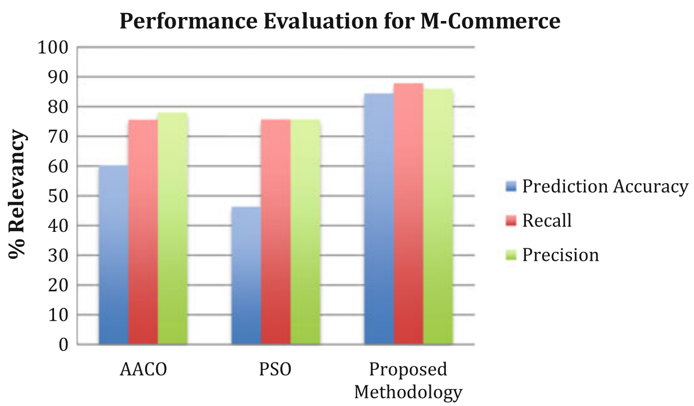


Fig. 2 Performance evaluation

6 Conclusion

Rapid increase of mobile commerce (M-Commerce)) with sensing devices has resulted in enormous services from different service providers. M-Commerce services provide numerous ranges of emerging services. By applying the above technique, we can extract relevant data with high precision and prediction accuracy which will help in social contextual behavior to society environment for diverse choice of goods and services.

References

1. Niknam, T., Nayeripour, M., Firouzi, B.B.: Application of a new hybrid optimization algorithm on cluster analysis. Int. J. Electr. Comput. Eng. **4**(4), 238–243 (2009)
2. Kim, W.: Parallel clustering algorithms: survey (2009). https://s3-us-west-2.amazonaws.com/mlsurveys/46.pdf. Accessed 9 March 2017
3. Lee, S., Ahn, H.: The hybrid model of neural networks and genetic algorithms for the design of controls for internet-based systems for business-to-consumer electronic commerce. Expert Syst. Appl. **38**, 4326–4338 (2011)
4. Humphreys, et al.: A hybrid optimization technique applied to the intermediate-target optimal control problem. Glob. J. Technol. Optim. Open Access J. **7**(2), 1–8 (2016)
5. Avasthi, S., Dwivedi, A.: Prediction of mobile user behavior using clustering. Int. J. Sci. Res. Publ. **3**(2), 1–5 (2013)
6. Evangeline, S.J., Subramanian, K.M., Venkatachalam, K.: Efficiently mining the frequent patterns in mobile commerce environment. **5**(1), 30–39 (2014)
7. Papamichail, et al.: The k-means range algorithm for personalized data clustering in e-commerce. Eur. J. Oper. Res. **177**, 1400–1408 (2007)
8. Priyadharshini, R., Kalimuthu, M.: An efficient framework for predicting and recommending M-Commerce patterns based on graph diffusion method. Int. J. Comput. Appl. Technol. Res. **2**(3), 311–314 (2013)

Effects of Cognitive Ability, Trust and Time-Saving: Predicting Further Amelioration and Successive Usage of E-ticketing with TAM, TPB and Cognitive Frameworks

Rifat Hossain, S. M. Hasan Mahmud, Md. Altab Hossin, Touhid Bhuiyan and Yin Xiao Hua

Abstract Prior research was perused the technology usage on limited factors; our aim is to investigate factor to adumbrate e-ticketing systems accuracy with further improvement and successive usage in the context of Bangladesh. To do so, we designed a research model with the integration of Technology Acceptance Model (TAM), Theory Planned Behaviour (TPB), trust theory, time-saving theory and cognitive framework. Survey data were amassed from the different profession of consumers who are using e-ticketing system; 145 valid data were tested the research model to find structural relationship and hypothesis result by using Partial Least Squares (PLS). Cognitive ability has a significant impact on time-saving and trust to purchase the ticket on Internet, while a wispy impact has cited between cognitive ability and perceived usefulness. In addition, trust has positively influenced on attitude towards of buying an online ticket, and time-saving has a positive relationship with TAM. Moreover, the study assists e-ticketing authorities and individual consumer to understand which factor gradually affects the progressive usage of e-ticketing.

R. Hossain (✉) · S. M. Hasan Mahmud · T. Bhuiyan
Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh
e-mail: rifat35-176@diu.edu.bd

S. M. Hasan Mahmud
e-mail: hasan.swe@daffodilvarsity.edu.bd

T. Bhuiyan
e-mail: headswe@daffodilvarsity.edu.bd

Md. A. Hossin · Y. X. Hua
University of Electronic Science and Technology of China, Chengdu, China
e-mail: altabbd@163.com

Y. X. Hua
e-mail: yinxiaohua2014@uestc.edu.cn

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_5

Keywords Cognitive ability · Trust · Time-saving · E-ticketing
E-commerce

1 Introduction

For all the developing countries, technology development had less consideration in previous decades, the abundance growth rate of new technology is the era of the Internet, and consumers are morphing new technology gradually. E-ticketing system takes most convenient place in the Internet activities [1]. Though the incipient of e-ticketing and Internet banking was minuscule popular, the usage of e-ticketing care of the Internet banking is becoming adoptive in Bangladesh.

E-ticketing is the preeminent achievement for Internet commerce. Consumers can purchase or reserve their ticket at anywhere anytime; the process is fathomable and attractive. Internet banking ensures consumer reliability with more confidentially and securely; payment gateway provided more security when a customer sends the ticket payment. Most renounced Internet banking gives the privilege to payment. Above 62 million consumers use the Internet in Bangladesh (BTRC, April 2016), while 58.045% clients are using Internet by the mobile phone (bdnews24.com, April 2016). Moreover, a customer can easily purchase a ticket at anytime from anywhere.

The world's topmost countries are taking the lucrative advantage of purchasing the ticket via the Internet including the USA, Australia, Europe and Canada. A copious proportion of US consumers used e-ticketing while Europe adopts less than 50%; however, 58% Canadian Web users using e-ticketing (CBCnews, 28 October 2013). In Bangladesh, comparing of those countries; a moderate proportion of consumer usage e-ticketing system.

An ample amount of investment in IT sector of Bangladesh look forward to acquiring the sharp development; online electric bill payment, online train ticketing, online game ticketing and online bus ticketing are more accountable. As stated in the Bangladesh Business News (23 July 2015), more than 90% clients of who are using Internet via mobile phone (58.045%), accessing the Web with the mobile phone to purchase the ticket that accretion the e-ticketing client.

The challengeable point is to identify the cognitive ability of e-ticketing users that how they adopt new technology. To govern the factor which explains e-ticketing customer's behaviour, according to the TAM and TPB. Analysing these theories, making an accord of which factors are grasping the increment of e-ticketing usage and mitigate the dilemmas. Followed by the TAM, TPB and cognitive theories, varieties of research have been found in the positive relationship between attitude and behavioural intention to use [2]. Trust positively influenced online shopping [3]. We aimed that cognitive ability is predicting the technologies present and future usage depending on systems trust and time-saving consequences.

Is there any impact of external human factors like cognitive ability, trust and time-saving on TAM and TPB to use e-ticketing in Bangladesh?

Based on the above research questions, our research objectives are:

RO1: To examine the effect of cognitive ability to purchase a ticket on e-ticketing in Bangladesh.

RO2: To examine the effect of time-saving to purchase a ticket on e-ticketing in Bangladesh.

RO3: To examine the effect of trust to purchase a ticket on e-ticketing in Bangladesh.

2 Theoretical Background and Hypothesis Development

Alzahrani et al. [4] considered TPB with the explanation of online gaming payment system which ascribes customer attitude factor [5]; mobile payment adoption attributed customer attitude and behaviour intention [6].

Cognitive science identifies consumer cognitive behaviour and physiology to adopt new technology [7]. Cognitive ability endorses foreteller to foretell the integration of device and technology, evaluating the positive and negative integration of the proposed system. Consumer trust depends on the system security, efficiency, time-saving and the operations, cognitive science ensure system performance depending on human and device integration.

TAM theory grosses over the usability and adaptability of perceived usefulness and perceived ease of use between attitude and behaviour intention to use [8]. From meta-analysis, it helps fortune teller to portend new technologies acceptation and exception; WeChat payment [9] significantly adopts TAM theory, Internet banking [10], mobile payment [11]. The research proposed (Fig. 1) a model with the integration of TAM, TPB and cognitive framework.

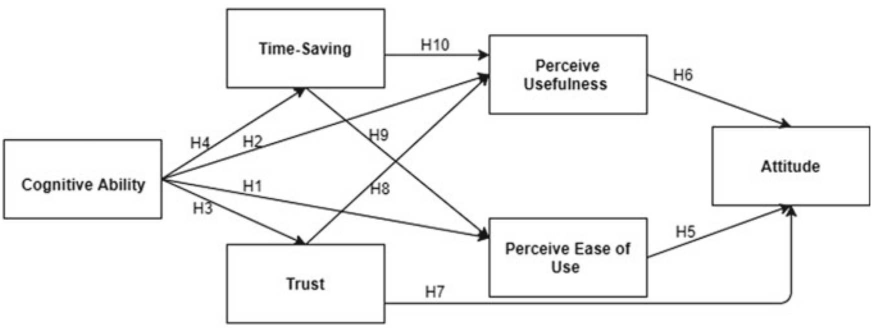


Fig. 1 A proposed research model

2.1 Cognitive Ability

Cognitive ability is considered as a prediction factor which measures the human interaction of new device and new technology [7], predicts what kind of method and operation should make flow to increment the system performance. In e-ticketing system, e-payment contributes pivotal role when a client will purchase ticket; cognitive ability ensures the technologies' future usage [12]. Cognitive style made a positive influence on online payment system [11]; the intention to attempt murder is also discerned by a cognitive factor [7]. Spüler et al. [13] integrated EEG with the prediction of cognitive workload.

H1 There is a positive relationship between cognitive ability and perceived usefulness of purchase ticket using the Internet at present and in future.

H2 There is a positive relationship between cognitive ability and perceived ease of use of purchase ticket using the Internet at present and in future.

H3 There is a positive relationship between cognitive ability and trust to purchase tickets using the Internet at present and in future.

H4 There is a positive relationship between cognitive ability and time-saving of purchase ticket using the Internet at present and in future.

2.2 Perceive Ease of Use

Perceived ease of use is referred as the degree to which a person believes that using a particular technology, predicting the system would be free of the attempt [8]. In existing research, perceived ease of use was strong affection on attitude towards to use online marketing, to forecasting online food ordering clients behaviour has a significant effect on the attitude [2]. Eventually, the use of e-ticketing system achieves customer trust and acceptance.

H5 There is a positive relationship between perceived ease of use and attitude towards to purchase online ticket.

2.3 Perceive Usefulness

Perceived usefulness is mentioned as the degree to which a person believes that using a particular system to enhance his or her job performance [9]. In online ticketing, customer purpose, how to improve the customer performance by using the given adoptive technology. Perceived usefulness gives the reliability to use the particular technology [10]. Based on the perceived usefulness, e-ticketing makes positive factor on customer expectation.

H6 There is a positive relationship between perceived usefulness and attitude towards purchasing online ticket.

2.4 *Trust*

Trust refers to the degree to which a person believes that new technologies' usage depends on the reliability of his or her job performance [9, 14, 15]. When a new technology is launched, a consumer does not know the usability, reliability and security; system performance could ensure these factors efficiency. Online payment, Internet banking makes more confident to find the positive effect on trust [10] and trust influence consumer attitude to purchase travel online [16]. Trust is contributing a long-term evolvement between stakeholders [17]. So, trust has significantly yielded positive intention to use e-ticketing.

H7 There is a positive relationship between consumer trust and attitude towards to purchasing online ticket.

H8 There is a positive relationship between consumer trust and perceive usefulness to purchase online ticket.

2.5 *Time-Saving*

In the urban life, the consumer does not pay more time to purchase the ticket by itself. The process of getting a ticket from the agency is more time consuming than purchase ticket on the Internet. While e-ticketing takes less time and saves time that becomes a great factor for a customer. Time-saving factor mostly followed in the USA and UK consumers to order food using the Internet at home [2].

H9 There is a positive relationship between time-saving and perceived ease of use of purchasing online ticket.

H10 There is a positive relationship between time-saving and perceived usefulness to purchase online ticket.

3 Research Method and Data Analysis

The objective of this paper is to test the research model and simulating factors philosophy to predict e-ticketing usage through Internet banking. Factors are integrated with theoretical framework these already significant numerous Internet commerce contextual research. Survey method with the questionnaire tools used to evaluate the development model.

Around 145 sample size were hoarded from university students, Govt. employees and garment workers in Bangladesh. A questionnaire conducted in two sections; section one contained demographic information and another section enlightening proposed research questionnaires. The response was justified by seven numerous variables, questions rate being started from "strongly disagree = 1" to "strongly agree = 7".

PLS path-modelling is a variance-based method for measurement composed-based path model [18]. Illustrating the validity of cross-loading criteria, the average variance extracted; convergent validity and composite reliability exhibit the hypothesis result. Therefore, analysing the validity and measurement, smart PSL software followed PLS method and attained the fame from the scholar.

3.1 Measurement Model

The measurement model experiments the outer model by ascertained item loading, the average variance extracted (AVE), composite reliability (CR) and Cronbach's alpha. Cronbach's alpha measures each questions internal reliability.

3.2 Average Variance Extracted and Composite Reliability

Each construct should have needed more than 50% average variance extracted [18] and more than 0.7 of composite reliability [19]. Thus, both criterions touched our variables (See Table 1). For Cronbach's alpha, 0.6 is the minimum value to significant variable reliability [2]. Table 1 covered the criterion positively.

Table 1 Cronbach's alpha, composite reliability (CR) and AVE

Research construct	Cronbach's alpha	Composite reliability (CR)	Average variance extracted (AVE)
Attitude	0.674	0.818	0.601
Cognitive ability	0.801	0.870	0.626
Perceived ease of usefulness	0.796	0.864	0.615
Perceived usefulness	0.796	0.881	0.713
Trust	0.758	0.861	0.674
Time-saving	0.726	0.824	0.544

Table 2 Fornell–Larcker criterion (discriminant validity)

Research construct	Attitude	Cognitive ability	Perceived ease of use	Perceived usefulness	Trust	Time-saving
Attitude	**0.775**					
Cognitive ability	0.736	**0.791**				
Perceived ease of usefulness	0.723	0.755	**0.784**			
Perceived usefulness	0.607	0.557	0.553	**0.844**		
Trust	0.678	0.638	0.498	0.633	**0.821**	
Time-saving	0.806	0.710	0.712	0.608	0.650	**0.737**

3.3 Discriminant Validity

From the discretion of discriminant validity, Fornell–Larcker criterion exposes the compatibility with the proposed model. Therefore, discriminant validity is being positively constituted our designed model (See Table 2).

3.4 Structural Model

Structure model is assessed with path coefficient, standard deviation and statistical testing. The bootstrapping procedure allows the calculation of p-value with lots of sample for the determination of path coefficient. Finally, T-statistics and path coefficient examined the hypothesized model. According to the Table 3, Path co-efficiency 0.638 and T-statistics of 7.894 supported the relationship between cognitive ability and trust. Trust significantly established positive relationship with an attitude towards (path coefficient 0.371, T-statistics 3.245) and H7 strongly supported, while H4 (cognitive ability -> perceive usefulness) with path coefficient 0.119 and 0.841 was not supported. Moreover, Table 3 shows the structural relationship and hypothesis test result.

4 Discussion

Cognitive ability established the presage validity and factorial validity of governance to individually purchase online ticket. How confidently consumer can use the system to present and future, cognitive ability established this statement. Pappas et al. [20] found customers cognitive ability and affective perceptions augment

Table 3 Path coefficient and hypothesis test result

Hypothesis	Path	Path coefficient	Standard deviation	T-statistics	Result
H1	Cognitive ability -> perceive ease of use	0.504	0.119	4.246*	Supported
H2	Cognitive ability -> perceive usefulness	0.119	0.142	0.841*	Not supported
H3	Cognitive ability -> trust	0.638	0.086	7.376*	Supported
H4	Cognitive ability -> time-saving	0.710	0.065	10.911*	Supported
H5	Perceive ease of use -> attitude	0.478	0.071	6.753*	Supported
H6	Perceive usefulness -> attitude	0.108	0.116	0.930*	Not supported
H7	Trust -> attitude	0.371	0.114	3.245*	Supported
H8	Trust -> perceive usefulness	0.375	0.154	2.436*	Supported
H9	Time-saving -> perceive ease of use	0.354	0.115	3.090*	Supported
H10	Time-saving -> perceive usefulness	0.280	0.165	1.709*	Supported

purchase intention and online shopping behaviour. Therefore, hypotheses H1, H3 and H4 positivism significantly assessed that system accuracy and regular usage importantly depend on trust, time-saving and perceived ease of usefulness. Lipnevich et al. [21] tested this cognitive ability with personality factors on mathematics implementation and importantly influenced. Though perceive usefulness does not support with cognitive ability (H2), overall assessment is committed to purchase the ticket confidently on the Internet.

From prior research, public bike system designed a model followed by the integration of TAM and TPB which meet a positive relationship between perception pleasure and attitude towards [22]. Guritno and Siringoringo perused [23] perceived ease of use influenced attitude towards online airlines ticket reservation. Online shopping had positively impacted with customer perceived value [24]. Table 3 illustrates perceive ease of use has a significant factor (Path coefficient 0.478, T-statistics 6.753) with strong positivism to attitude towards. This ensures a moderate plausibility to purchase the ticket from online with the perceived ease of use and consumer attitude to usage. Therefore, H5 has been accepted, while H6 has not been accepted. Perceived usefulness has not been established a significant relationship with attitude towards of e-ticketing.

Based on previous research, trust is the strongest clairvoyant on online travel websites usage [16]. Online supermarket with the relationship of e-trust significantly supported to the loyalty [25]. Wang et al. [17] committed trust explored a key determination of stickiness intention. Trust directly impacted on consumer attitude towards of e-ticketing usage [23] (See Table 3). Trust positively adhered with the attitude towards and perceived usefulness (H7 and H8 supported) to purchase the online ticket via e-banking.

Cheow et al. [2] found an unequivocal relationship with time-saving to make an unambiguous improvement on online food delivery service. A large proportion of clients demand ticket within minor time and wanted the privilege to acquire a ticket at anywhere anytime. Time-saving positively impacted with perceived ease of use and perceived usefulness. Therefore, H9 and H10 are supported to purchase online ticket.

Moreover, time-saving and trust disclose consumer intention to use e-ticketing and cognitive ability constructs accurate decision making and consistency of e-ticketing.

5 Conclusion and Future Work

Retrospect to the previous research, a limited number of research merged with cognitive ability, time-saving and TAM. We designed a new model which defines the consumer physiology, adoption behaviour to purchase an online ticket with the usage of Internet. Followed by the result, the Cognitive ability is the most significant factor to predict the usage of device and technology with the interruption of human, time-saving and trust. All axiom positively influenced to corroborate consumer to purchase the ticket in online. The result is ostensible for foreteller which factor is staring vitality and its further improvement will be promoting the service quality to end-user acceptance.

Approximately 150 responses were calculated in this study. More response may represent the different result which would be more calculative to detect the proposed hypothesis. This research designed cognitive factor and time-saving factor with the integration of TAM and TPB; however, different integration will be providing more significant to purchase a ticket in online via Internet banking. Here, this study only conducted their survey in urban areas, nevertheless conducting the survey in the rural area will have been the future research in Bangladesh.

References

1. Jalil, M.S.: E-service innovation: a case study of Shohoz.com. Procedia Soc. Behav. Sci. 531–539 (2016)
2. Cheow, V., Yeo, S., Goh, S., Rezaei, S.: Consumer experiences, attitude and behavioral intention toward online food delivery (OFD) services. J. Retail. Consum. Serv. 150–162 (2017)
3. Bilgihan, A.: Computers in human behavior Gen Y customer loyalty in online shopping: an integrated model of trust, user experience and branding. Comput. Human Behav. 103–113 (2016)

4. Alzahrani, A.I., Mahmud, I., Ramayah, T., Alfarraj, O.: Extending the theory of planned behavior (TPB) to explain online game playing among malaysian undergraduate students. Department of Computer Science. Telemat. Inform. (2016)
5. Nasri, W., Charfeddine, L.: Factors affecting the adoption of Internet banking in Tunisia: an integration theory of acceptance model and theory of planned behavior. J. High Technol. Manag. 1–14 (2012)
6. Martins, C., Oliveira, T., Popovič, A.: Understanding the Internet banking adoption: a unified theory of acceptance and use of technology and perceived risk application. Int. J. Inf. Manag. 1–13 (2014)
7. Butler, M.J.R., Senior, C.: Critical essay: organizational cognitive neuroscience drives theoretical progress, or: the curious case of the straw man murder. Human Relat. (2017)
8. Ozturk, A.B., Bilgihan, A., Nusair, K., Okumus, F.: What keeps the mobile hotel booking users loyal? Investigating the roles of self-efficacy, compatibility, perceived ease of use, and perceived convenience. Int. J. Inf. Manag. 1350–1359 (2016)
9. Qu, Y., Rong, W., Ouyang, Y.: Social aware mobile payment service popularity analysis: the case of WeChat payment in China. 289–299 (2015)
10. Al-sharafi, M.A., Arshah, R.A., Abu-shanab, E., Fakhreldin, M., Elayah, N.: The effect of security and privacy perceptions on customers trust to accept internet banking services: an extension of TAM literature review. COMSCET 1–9 (2016)
11. Hossain, R., Mahmud, I.: Influence of cognitive style on mobile payment system adoption: an extended technology acceptance model. 5–10 (2016)
12. Huang, C.: Cognitive factors in predicting continued use of information systems with technology adoption models. Int. Electron. J. (2017)
13. Spüler, M., Gerjets, P., Klein, E.: EEG-based prediction of cognitive workload induced by arithmetic: a step towards online adaptation in numerical learning. ZDM 267–278 (2016)
14. Ozturk, A.B., Nusair, K., Okumus, F., Singh, D.: Understanding mobile hotel booking loyalty: an integration of privacy calculus theory and trust-risk framework. Inf. Syst. Front. 1–15 (2017)
15. Pantano, E., Priporas, C.: Computers in human behavior the effect of mobile retailing on consumers â€™ purchasing experiences: a dynamic perspective. Comput. Human Behav. 548–555 (2016)
16. Agag, G.M., El-masry, A.A.: Why do consumers trust online travel websites? Drivers and outcomes of consumer trust toward online travel websites. J. Travel Res. 347–369 (2016)
17. Wang, W., Wang, Y., Liu, E.: The stickiness intention of group-buying websites: the integration of the commitment-trust theory and e-commerce success model. Inf. Manag. 625–642 (2016)
18. Cepeda, G., Henseler, J., Ringle, C.M., Luis, J.: Prediction-oriented modeling in business research by means of PLS path modeling: introduction to a JBR special section. J. Bus. Res. 4545–4551 (2016)
19. Mahmud, I., Ramayah, T., Kurnia, S.: As User Resistance in Pre-implementation Stage of pt us cr. Elsevier Ltd. (2017)
20. Pappas, I.O., Kourouthanassis, P.E., Giannakos, M.N., Chrissikopoulos, V.: Explaining online shopping behavior with fsQCA: the role of cognitive and affective perceptions. J. Bus. 794–803 (2015)
21. Lipnevich, A.A., Preckel, F., Krumm, S.: Mathematics attitudes and their unique contribution to achievement: going over and above cognitive ability and personality. Learn. Individ. Differ. 70–79 (2016)
22. Chen, S.: Using the sustainable modified TAM and TPB to analyze the effects of perceived green value on loyalty to a public bike system. Transp. Res. 58–72 (2016)
23. Guritno, S., Siringoringo, H.: Perceived usefulness, ease of use, and attitude towards online shopping usefulness towards online airlines ticket purchase. Procedia Soc. Behav. Sci. 212–216 (2013)

24. Wu, L., Chen, K., Chen, P., Cheng, S.: Perceived value, transaction cost, and repurchase-intention in online shopping: a relational exchange perspective. J. Bus. Res. 2768–2776 (2014)
25. Moriuchi, E., Takahashi, I.: Satisfaction trust and loyalty of repeat online consumer within the Japanese online supermarket trade. Australas. Mark. J. 146–156 (2016)

Knowledge Management Systems Leveraging the Competitive Advantage of Top IT Organizations: A Multi-case Study of Benchmarking Practices

Geeta Sachdeva and Seema

Abstract Purpose: In today's scenario, the essential factors for successful designing and implementation of knowledge management system at organizational level with respect to their benchmarking practices have not been critically evaluated. The needs of a large-scale initiative of knowledge management have not been investigated from the stakeholders' perspectives, and only perspective of firms has been studied. Moreover, it has not been investigated how knowledge management systems have transformed the brand of top IT organizations continually and how they have sustained their growth by using their KM systems strategically. This paper is targeting to abridge this space. Design/methodology/approach: A qualitative multi-case analysis technique has been used in this paper for data synchronization to gain fruitful insights into this issue. Collection of data has been done from real case studies in the literature, and same was categorized and analysed. Findings: The overall results analysed from these real-time case studies have been constructive, thereby projecting the relevance of building and sustaining a synchronized KM system in the ever-evolving business environment and some critical success factors have also been proposed. The support of Human Resource Development wing has been found very critical for successful implementation and maintenance of knowledge management system in an organization.

Keywords Knowledge management · Benchmarking practices of KM
Knowledge economy

G. Sachdeva · Seema (✉)
National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India
e-mail: seemapanchal07@gmail.com

G. Sachdeva
e-mail: geeta_schdv@yahoo.co.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_6

1 Introduction

The competencies possessed by the employees of an organization decide the worth of the intellectual capital of the firm. These competencies are developed over a period of time, and lot of investment with respect of time, resources and efforts are required for their diffusion into human intellect. When these competencies are acquired by human capital of an organization in a systematic manner, the organization becomes a successful learning organization and when these learning practices become a culture of the organization, it becomes a small knowledge economy. Now its up to that organization how to manage that knowledge and its human capital which are called knowledge workers. This leads to the designing and development of full-fledged knowledge management systems of an organization which is well synchronized with the evolving needs of the businesses. An effective KM system means creating a framework of knowledge in which its knowledge workers could build and utilize their potential to their maximum utility.

KM provides such an environment within which it is amusing working, also where the human capital could learn and contribute to their peers, partners and customers. In a world of ever-changing marketplaces, firms are continuously finding new ways to become accustomed to changing circumstances so that they are equipped to endure and thrive in this stringent competition. The propagation of knowledge economy stressing upon the worth of information as an enabler of building and maintaining competitive advantage has been definitely and obviously driving many firms to review the ways they were treating their knowledge resources previously and to distinguish methods in which they could take their advantage more resourcefully in the future. In this scenario, it is not surprising that knowledge management and information management have been emerging as two of the most sought-after tactical management approaches in the twenty-first century. Its followers are arguing that firms especially operating in information technology industry may attain significant competitive advantages by investigating the data and information they acquire and possess that often remains untapped and transforming them into productive and moreover actionable knowledge. In short, knowledge management concept can be explained as an incorporated, orderly and vibrant method to identify, manage and share all or particularly the firm's information resources, including databases, documents, policies and procedures, as well as previously unarticulated expertise and experience which has been retained with individual employees.

Bergeron (2003) defines knowledge management in this way: "Knowledge Management is a deliberate, systematic business optimization strategy that selects, distils, stores, organizes, packages, and communicates information essentials to the business of a company in a manner that improves employee's performance and corporate competitiveness."

In today's dynamic environments when technology is evolving every day and organizations also face challenges to sustain in global marketplace, it becomes a necessity to gain competitive advantage to proceed towards sustainable

Table 1 Prominent reasons for launching of KMS systems in some of the top MNCs

Sl. no.	Reason	% of companies for this reason (A total of 20 companies investigated by Dr. Chaudhuri S.)
1	Gaining competitive advantage	80
2	Improve customer retention/satisfaction	80
3	Retain key talent/ expertise	75
4	Develop new services	70
5	Improved brand image	25
6	Avoid loss of key personnel	40

Source Excerpts from Chaudhuri, S. (2011). Knowledge Management in Indian IT industries. In 3rd International Conference on Information and Financial Engineering, IPEDR (Vol. 12)

development. This leads to a question what should be the prime focus of investment of a global organization which is operating in information technology industry to achieve and maintain service excellence. The answer lies in the question itself. The main focus of an IT firm should be to build and maintain its own knowledge management system and also participate in building the knowledge economy of the industry and the nation at large. While studying the evolution of the knowledge management systems of various IT firms, it was witnessed that most of the firms did launch the system for gaining competitive advantage which can also be seen in the study done by Dr. Chaudhuri Santwana, 2011.

The most prominent reasons found out by Dr. Chaudhuri, S. in his research investigation are listed in Table 1.

2 Literature Review

Following literature was studied during the research work investigation and gaining insights into work done till date on the captioned subject matter—

In "Barriers to Effective Use of Knowledge Management Systems in Software Engineering" by Desouza [1], the author has tried to explore the use of KM system in IT industry to know the significant issues which hold back the effective utilization of the knowledge management systems in computer hardware and software engineering. In "Knowledge Management for Business Performance Improvement" by Alhashmi et al. [2], the purpose is how the decision-making process exists in the intellect of workforce and why a system is looked-for to take into custody and effectively codify this knowledge. Decision-making is made the basis of knowledge, and the management of this has been investigated in detail by this study. In "Key Factors in Knowledge Management" by Babu [3], the researcher has qualitatively investigated the critical success factors and some of the versatile practices in the industry for knowledge codification for its management through analysing the experience of several organizations. He has studied the knowledge management system of various companies and provided the best selection out of them. In "Exploring the Failure Factors of Implementing Knowledge Management System in the Organizations" by Akhavan et al. [4], the focus was to determine whether all the system specifically employed for implementing the knowledge management system in the organizations are worthy and fetch satisfied results and if they fail, what are the main factors contributing towards the same. In "Knowledge Management Research at the Organizational Level" by Bray [5], the researcher has tried to investigate the knowledge management system with respect to the information system employed by the company but the scope of furtherance lies with difficulty to communicate and retain contextual information. In "The Balanced Scorecard as a Knowledge Management Tool: A French Experience in a Semi-public Insurance Company" by Wegmann [6], the researcher has studied the KM system with respect to the controlling measure of the company performance, i.e. balanced scorecard. He has accumulated planning decisions with control ones. In "Minding the cognition: toward a strategic knowledge management for competitive advantage" by Muthusamy [7], he has studied the KM at various levels of the organization and particularly some of the strategies to obtain, generate and propagate the knowledge across the different hierarchies of the organization. In "Exploration, Exploitation, and Knowledge Management Strategies in Multi-Tier Hierarchical Organizations Experiencing Environmental Turbulence" by Bray [8] (Date posted: 6 February 2007 Last revised: 5 January 2009), this study is an extension of March's model of considering exploration and exploitation in a multi-tier firm as to how hierarchies matter in the execution and functioning of KM system in the firms. In "Critical Success Factors of Knowledge Management Systems: a Multi-Case Analysis" by Akhavan et al. [9], the main focus of this study was to determine all possible factors in a synchronized way for designing and implementing knowledge management system in organizations. They have done it with the help of various case studies of different companies. In "Intellectual Capital Evaluation: Relationship between Knowledge Management Implementation and Company's Performance" by Shakina and Bykova [10], they have focused upon the degree of improvement in the performance of the company by applying the knowledge management system. Their work shows both the external and internal factors which affect the performance of the company and the complementarity between the components of KM system like human, institutional and market resources. In "The Use of Knowledge Management Systems to Support Knowledge Creation and Sharing Activities among Employees—A Survey based Study of IT Shared Services Company" by Hoong and Lim [11], the basic work done in this paper is to investigate the utilization of KMS in organizations and the outcome of its scientific features.

3 Objectives

1. To understand the benchmarking practices of knowledge management systems implemented in ten reputed IT organizations and understand their working for capacity building of the organizations.
2. To analyse the critical factors to be studied carefully while implementing knowledge management in an organization.

Research Methodology Data was collected from the various studies done in the past on the topics related to the captioned subject matter. Various practices of knowledge management of different IT organizations available in literature were studied thus making this study as exploratory in nature. In-depth study with categorizations **was particularly selected as the methodology ascertaining that this study particularly is a multi-case analysis**. For the whole study, from data collection to data synchronization and further data analysis are inter-related. The study examines the various aspects of designing, implementing and maintaining the KM systems in ten reputed IT organizations. It also includes discussion of various tools which have been incorporated in the KM system of such organizations and their benefits in providing and maintaining their competitive advantage.

Conceptual Framework and Case Studies
As far as concepts of KM are concerned, multiple explanations of knowledge have been discussed in the narrative, and various debates regarding this theory have been articulated from a series of aspects and views of differing experts. Actually from the ancient Greek time, the hunt of mankind's thought process has been to seek what knowledge actually is. Also, early philosophers such as Plato and Aristotle, followed by Hobbes and Locke to name, have given various philosophies towards the same. However, one should not and must not look at it from a philosophical point of view. As observed by Alavi and Leidner (2001), the knowledge-based theory of the firm was never built on a universal truth of what knowledge really is but on a pragmatic interest in being able to manage organizational knowledge. In current times, we are witnessing an enlarged vision and interest in knowledge from other domains also.

Knowledge Management Systems
Following the above explanations, it can be stated that knowledge and expertise which exist in firms which are generating additional and fruitful significance when rapidly applied give emphasis to mainly on building and maintaining a sustainable knowledge economy which supports extensive knowledge and technology transfer. Undeniably, the knowledge is of inadequate significance if it is not pooled. As a result of which firms are starting to put into practice such information systems which are premeditated specially to smoothen the progress of the **production, incorporation, distribution and propagation** of knowledge of organization. Such frameworks can be referred to as knowledge management systems and fall into following four categories:

1. **Content administration equipments**: Platform that can offer capabilities to synchronize, categorize and codify knowledge from different sources.
2. **Knowledge contribution equipments**: Platforms which are used to support contribution of knowledge among employees and various stakeholders of the organizations.
3. **Knowledge exploration and retrieval systems**: Tools providing the platform that enables the organizations for exploration, recovery and acquiring the knowledge discovering abilities.
4. C**ommon Knowledge Management system**: Platforms which put forward an efficient and comprehensive key for all general needs of a firm from people management to quality assurance. Among these generalized KM systems, **corporate portals** give the impression to make available organizations with a rich and complex shared information service which acts as a workspace for the conception, exchange, preservation and reprocessing of knowledge.

Portals Support for Knowledge Management Processes

The terms Employees Portals, Enterprise Intranet Portals, Corporate Portals, Business-to-Employees Portals and Business-to-Employees Systems are many a times worn alternatively for the same meaning to consign to the general grouping of portals, whose focus is to provide employees with relevant information at required time specifically which they generally require to execute their duties and formulate proficient trade decisions. While looking for some good knowledge management lessons from Indian IT firms, we can notice that every firm wants to build its own knowledge base to remain competitive even when turnover is very unpredictable in this industry.

Microsoft started a project called skills planning and development (**SPUD**) in 1995 through its internal IT personnel which involved building an online platform for performance management of employees whose goal was to build an online repository of competency profiles for various jobs at Microsoft. This platform later evolved into a full-fledged KM shop providing various facilities to build and maintain a global knowledge economy within an organization. According to **Bill Gates, "Knowledge Management is a fancy term for a simple idea. You're managing data, documents, and people efforts"**. Some of the technologies used by Microsoft for maintaining its KM system are **My web and Microsoft SharePoint**. Launched in 2001, Microsoft SharePoint facilitates content management and various cutting-edge similar capabilities. As many as 75–78% Fortune 500 Companies use SharePoint. From 2006 to 2011 only, Microsoft figures of sold user licenses were over 36.5 million. Apart from SharePoint, some unique employee engaging KM practices include **Breakfast Meetings** and **Biz-talk** (a knowledge library open to everyone in the organization).

Google started its KM system by initializing Knowledge Creation at first and then integrating all the phases of KM into it. The organization believes that in order to create knowledge, we need to create enabling conditions through innovations for the same. The same is proposed by Nonaka Theory also. Google's innovation ecosystem is composed of Content providers, Advertisers, Consumers and

Innovators. **Knowledge Creation (Nonaka)** in Google rests on five pillars mainly which are called **Intention, Autonomy, Fluctuation and Creative Chaos, Redundancy and Requisite variety**.

Apple started its KM Journey quite earlier from most others in the industry. It started its KM system with extensive planning and back up. It enhanced its organizational intelligence by creating such systems which enable knowledge workers to get better work processes and improve their quality of work life. The knowledge management tools used by Apple Inc. are **Data Warehouse, Company Intranet and Expert System**. Data Warehouse is a repository of information which is validated, reformatted, reorganized, summarized, restructured and supported with data from relevant sources. Company Intranet uses an Internet Protocol technology to distribute information, operational systems or computing services within Apple. Artificially intelligent platform was designed to provide solutions for complex problems of organizational working.

HP Consulting has devised its own state-of-the-art KM system focusing mainly on the needs of consultants which has been named as **K-Net**. It is an electronic platform for accessing structured knowledge and participating and following discussion forums where the consultants can browse, explore and submit the queries and capabilities. It also serves as a tool for capacity building of the consultants for the organizational brand formation. It also has a populated **Solution store** for consistency and standardization of knowledge across various domains for similar problems. For project knowledge sharing, it has **Project workspace, Project document manager and Project Discussion Forum**.

Taking the case of **Infosys**, the firm is trying to excel in improving teamwork, refining the softwares, reusing codes and meeting the growth expectations of the shareholders through its unique knowledge management system based on Knowledge Management Maturity Model. Its knowledge management portal is serving as a knowledge shop wherein the existing and recently posted knowledge can be capitalized into **Knowledge currency units** (KCUs). The employees can rate this knowledge into tangible KCUs according to their relevance in their working and can also earn the same on the content posted by them. This serves as a very good incentivization scheme for the management to keep the employees engaged in maintaining this knowledge shop thereby fulfilling organizational goals of making a competitive KM system. The ultimate **motto** of the firm in promoting this KM system is “**Learn Once, Use Anywhere**”.

Oracle started its formal KM system in 1997. Consultants find themselves convenient for getting the reference material, and customers can obtain the technical assistance on a self-service mode by using its **web-based project libraries**. The firm puts KM system into action through their state-of-the-art technological platforms like **My oracle.com, GlobalXchange, Knowledge Areas and Community Areas**. Technology serves as a critical enabler in implementing these platforms; such is the significance of the technology in today’s era. They are extending this KM system beyond the enterprise by planning the **Oracle Technology Network (OTN) and Oracle Partner Network (OPN)** to integrate its supply chain network through their knowledge network.

Open Text had launched a KM network as an Extranet especially for its affinity partners beyond the enterprise needs in 1997 to improve their inter-collaboration. It hosts a Corporate Intranet also, called **OLLIE** which possesses **Global knowledge library**, and three practicing communities called **Competitive intelligence forum, Customer's Dashboard and knowledge centre**. It has also provided a special KM platform for its field knowledge workers which has been named as **Livelink Wireless**.

Sun Microsystems Philippines (Sunphil) was made through a joint venture between Sun Microsystems and the famous erstwhile distributor Philippine Systems Products in 1999, and they integrated their knowledge shops very effectively through the Digital KM system called **Sunphil Corporate Portal** which has transformed the traditional knowledge management of heavy paperwork into seamless paperless documentations and knowledge transfer. It has unique features which work without the hassle of manual reporting. It includes some of the very versatile features of knowledge management like standard document rating, standard profiling and filtered search and standard collaborative authoring. Project proposal preparation and documentation turnaround time (TAT) have been reduced significantly by its KM portal. Also, they are exploring the innovative approaches of harnessing information mobility and real-time expert contact through their KM platforms like Personal Digital Assistant (PDA) and short messaging service as Philippines has been considered as the world's SMS capital.

Siemens' wing of Information and Communication Networks devised **Sharenet** in 1999 as a business development framework which supported knowledge management while helping sharing knowledge of projects through different stages of maturity and across differing technologies. This has enriched the role of sales personnel into strategic management consultants who can interpret trends according to their requirements and explore new opportunities while collaborating with the customers. Sharenet is a global platform which shares the local innovation across the globe. Some of its unique applications include standard project debriefings, standard manuals, standard and customizable codified databases, structured questionnaires, customizable chat rooms and hotlines which help its personnel to work smartly removing the time of troubleshooting and redundancy in everyday technical work.

Tata Consultancy Services offers a portfolio of various IT and IT-enabled services to other organizations through outsourcing contracts across the globe. It has approximately over one and a half lakhs IT consultants on its rolls. In 1998, it first felt the need of a KM system and developed a **Corporate Groupware** and further did pilot testing of a standardized KM system in 1999. All the operating locations of the firm in the world are connected through Internet specifically using Lotus Notes Domino Servers. KM system covers every function from Quality Assurance to People Management. Knowledge Repository is called **Kbases** which possesses information about various processes, product lines and projects. **Process Asset Libraries** also known as Pals contain case studies for reference of project leaders. Pals and Kbases ultimately are merged for providing User Interface sub-portals in Quality Management, Software Productivity Improvement, Materials of Training and tools of information. **TCS also uses Microsoft SharePoint** for customizing its knowledge needs through **KnowMax** which provides the access of best practices and approximately four decades of experience to the consultants. Also, the Knowledge bank also called **K-Bank** can be maintained and contributed by any associate or officer thereby maintaining content quality. The various tools used for KMS in TCS are **Infinity** (for messaging, IP telephony, video conference), **Blogs** and **Wikis**, **Idea storms** and **TIP** (product innovation and problem-solving), **Mysite** (for social networking).

4 Analysis

All the ten IT organizations studied are practicing state-of-the-art knowledge management systems with their own customisations according to their business and clientele needs. After studying their KM systems, it has been evident that selecting or designing an efficient KM system in an organization whether small, medium or large is not a cake walk. It entails various steps which should include all the stages listed below. Required additions or alterations can be done depending upon particular requirement. These steps can lead towards an efficient thought process of making a comprehensive and integrated KMS in any organization:

1. Careful evaluation of Infrastructure;
2. Designing and development of KMS integrated with all organizational functions;
3. Identification of key knowledge workers and key knowledge assets;
4. Designing team accountable for sustaining KMS;
5. KM blueprint creation and system implementation;
6. Using evidence-based approaches for change management and cultivation of learning culture;
7. Evaluation of return on investment in tangible worth;
8. Collecting feedback and refining the KM system.

After understanding the framework to build a comprehensive KMS,

1. The **responsibility of maintaining** this extensive system lies **primarily with that organizational function which deals with** the assets responsible for handling knowledge, i.e. **intellectual assets**.
2. The phenomenon of **identifying intellectual assets** or key knowledge workers is also a key element in building and sustaining a KM system.

Both the tasks described above have been identified as the responsibilities of Human Resource Development (HRD) Function and hence it becomes a critical function of HRD to initiate, create and drive the knowledge management systems in the organizations to stay competitive in today's ever-evolving business environments.

For this purpose, HRD requires:

1. Support and strong commitment from the management;
2. Reinforcement strategies for adoption and driving learning and sharing culture in the organization;
3. Understanding of business processes to synchronize with KMS;
4. Culture which promotes information sharing and cooperative troubleshooting;
5. Flexible approach of management for learning from failures;
6. Incentivization support from management to retain and enhance the knowledge of key knowledge workers;
7. Inculcating and driving a culture for continual identification and diffusion of Best Practices.

5 Conclusion

In the list of emerging challenges currently being faced by HRD function in new age organizations, knowledge management is becoming a critical issue for maintaining competitive position in the industry. It becomes even more critical when it comes to IT industry where the basic units of operations are information databases generally referred to as knowledge. IT organizations need to maintain a well organized and integrated knowledge management system to meet all its demands in digital era. That is why it is stated that the role of HRD in today's scenario has uplifted from shop-floor people management to the board where it is actively involved in developing organizations into digital knowledge economies which are able to make informed decisions about their future. Knowledge management systems are proving as strong equipments to empower HRD function, thereby empowering the organizations as a whole to compete and strive for business excellence.

The survival of IT industry seems impossible in this digital era without possessing efficient KM systems which remain abreast with

1. Evolving technologies like artificial intelligence;
2. Virtual marketplaces;
3. Unpredictable ownership switchover and mergers and acquisitions;
4. Fluctuating figures of employee turnovers;
5. Evolving HR policies and procedures, etc.

Limitations and Future Scope of Study

1. Accessibility to the information regarding case studies included in the literature review of knowledge management of other companies was limited.
2. This study was done only pertaining to the literature review accessed. The findings are not generalizable to other areas of study.

Moreover, typical research areas for further study can be suggested as methodologies of knowledge management, its ontologies, technologies of language engineering, issues related to human and organizations in implementing knowledge management systems and much other.

Depending on the aim of the study, diverse interdisciplinary studies can be conducted which can range from information and big data science, organizational behaviour through its psychology, business management, business process re-engineering, etc. Thus, the knowledge management concept is beyond doubt a multidisciplinary domain which can only be carried forward if the various domains and domain specialists collaborate and share collectively the understanding for creation of such pioneering and trustworthy platforms for the same.

References

1. Desouza, K.C.: Barriers to effective use of knowledge management systems in software engineering. Commun. ACM **46**(1), 99–101 (2003)
2. Alhashmi, S.M., Siddiqi, J., Akhgar, B.: Knowledge management for business performance improvement (2006). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=670289. Accessed 13 Oct 2017
3. Babu, K.V.S.N.: Key factors in knowledge management (2012). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2150567. Accessed 13 Oct 2017
4. Akhavan, P., Jafari, M., Fathian, M.: Exploring the failure factors of implementing knowledge management system in the organizations (2012). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2188273. Accessed 13 Oct 2017
5. Bray, D.A.: Literature review-knowledge management research at the organizational level (2007). http://ssrn.com/abstract=991169. Accessed 13 Oct 2017
6. Wegmann, G.: The balanced scorecard as a knowledge management tool: a French experience in a semi-public insurance company (2007). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1012699. Accessed 13 Oct 2017
7. Muthusamy, S.: Minding the cognition: toward a strategic knowledge management for competitive advantage. Strateg. Change **17**(5–6), 145–153 (2008)
8. Bray, D.A.: Exploration, exploitation, and knowledge management strategies in multi-tier hierarchical organizations experiencing environmental turbulence (2005). https://papers.ssrn.com/sol3/papers.cfm?abstract-id=961043. Accessed 13 Oct 2017
9. Akhavan, P., Jafari, M., Fathian, M.: Critical success factors of knowledge management systems: a multi-case analysis. Eur. Bus. Rev. **18**(2), 97–113 (2006)
10. Shakina, E., Bykova, A.: Intellectual capital evaluation: relationship between knowledge management implementation and company's performance (2011). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1999904. Accessed 13 Oct 2017
11. Hoong, A.L.S., Lim, T.-M.: The use of knowledge management systems to support knowledge creation and sharing activities among employees—a survey based study of it shared services company. In: 2012 Ninth International Conference on Information Technology: New Generations (ITNG). IEEE (2012)

Performance Analysis of Parallel Implementation of Morphological Operations

Ashwin Geet D'Sa and B. G. Prasad

Abstract Morphological operation is a most common technique used in image processing because of its low complexity. Despite its simplicity, morphological operations usually take a long time since the same operations are repeated on every pixel of an image. Since the processing technique for each pixel in an image remains the same, the morphological operation can be carried out in parallel for different pixel, by effectively using the threads over the multi-core processors, which are now common even in the commodity hardware. The use of the API's provided by OpenMP allows parallelization of the serial program, by allowing us to create and control the threads. Hence, performance analysis can be made by comparing the effectiveness of carrying out the task in parallel over the serial implementation of the same task by varying the number of threads over the processor. Further, the memory usage by the program can be analyzed the varying the number of threads.

Keywords Morphology · Performance analysis · Multi-threading
OpenMP

1 Introduction

Morphology is an image processing technique which is usually applied to binary images, which is now extended to the grayscale images [1]. Morphology works based on the set theory where the operations are applied over two sets. One set is processed by another one, where one set decides the shape and size, and known as the structuring element (SE), which is translated over the image. Few of the morphological operations which are most commonly used are Erosion, Dilation,

A. G. D'Sa (✉) · B. G. Prasad
Department of Computer Science, BMS College of Engineering, Bengaluru, India
e-mail: win.12894@gmail.com

B. G. Prasad
e-mail: drbgprasad@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_7

Closing, and Opening, which are important building blocks of many other image processing algorithms.

1.1 Mathematical Morphology

Morphological operation is one of the most used techniques in image processing as well as pre-processing of the input image. Mathematical morphology is the set theory operation developed by Kowalczyk et al. [2]. It provides the neighborhood relation of the pixels such that p is a neighbor of q if and only if (p, q) $\in$ K, where K is kernel or structuring element.

Morphological techniques probe an image with a small shape or template called a structuring element. The structuring element is moved across all possible pixels in the image, and it is compared with the corresponding neighboring pixels. The fundamental operations of morphology are dilation and erosion and form the primary operations [3]. For the binary images, the operations are defined as below.

The erosion operation is defined as follows:

With A and B as sets in Z^2 (2-d integer space), the erosion of A by B, denoted $A \ominus B$, is defined as

$$A \ominus B = \{z \mid (B)_z \subseteq A\}$$

It indicates the erosion of A by B is the set of all points z such that B, translated by z ($(B)_z$), is contained in A. It can also be expressed as

$$A \ominus B = \{z \mid (B)_z \cap A^c = \varnothing\}$$

where A^c is the complement of set A and $\varnothing$ is the null set.

The dilation operation is defined as follows:

With A and B as sets in Z^2 (2-d integer space), the dilation of A by B, denoted $A \oplus B$, is defined as

$$A \oplus B = \{z | (B^\wedge)_z \cap A \neq \varnothing\}$$

It is based on reflecting B about its origin and shifting this reflection by z. The dilation of A by B is the set of all displacements, z, such that $B^\wedge$ and A overlap by at least one element. It can also be written as below

$$A \oplus B = \{z | [(B^\wedge)_z \cap A] \subseteq A\}$$

where $B^\wedge$ is the reflection of B.

The erosion operation for gray scale is defined as follows:

The erosion of f by a flat structuring element b at any location (x, y) is the minimum value of an image in the region coincided with b when the origin of b is at (x, y).

$$[f \ominus b] = \min_{(s,t) \in b} \{f(x+s,\ y+t)\}$$

The dilation operation for gray scale is defined as follows:

The dilation of f by a flat structuring element b at any location (x, y) is the maximum value of an image in the region coincided with b when the origin of b is at (x, y).

$$[f \oplus b] = \max_{(s,t) \in b} \{f(x-s,\ y-t)\}$$

The boundary extraction or the edge detection using morphology is given below:

$$\beta(A) = A - (A \ominus B)$$

where β(A) is the boundary set of A.

1.2 Parallel Computing

Parallel computing is a technique of dividing a large task into multiple small tasks and executing them simultaneously. Dividing a large task into smaller ones and executing them simultaneously usually result into the speedup of the application.

Multi-processing. In order to achieve parallel execution in software, hardware must provide a platform that supports the simultaneous execution of multiple threads [4]. One way to achieve the speedup of the applications is to exploit the concurrency in the software applications through multi-tasking or time-sharing of the threads executing the software. In multi-threading, time-sharing or time-slicing works by interleaving of the execution of the threads. This model does not allow for true parallelism as no two threads are executed simultaneously. Only one instruction stream can run on a processor at a single point in time.

Hyper-threading. By duplicating the architecture space, we can create a logical processor; however, the physical processor executing the instruction is same for those logical cores. This technique is known as simultaneous multi-threading [4], or hyper-threading technology, or HT technology. Hyper-threading technology provides multiple processors to the software applications; however, the number of physical processors remains the same. Through this technology, the operating systems can schedule multiple threads on the logical processors. However, this approach is a concurrent approach and does not account to true parallelism.

Multi-core. Since the creation of logical cores does not account for true parallelism, the improvement in the manufacturing technology can be considered.

Hence, more than one processing resources can be embedded on the single processor to form multiple "execution cores" within a single processor. These cores are essentially two individual processors on a single die. These cores have its own execution resource as well as the architectural resources.

OpenMP. The OpenMP provides APIs for user-directed parallelization, through the compiler directives, where the programmer specifies the actions to be taken by the compiler and runtime system in order to execute the program in parallel, also provides the constrains that has to be considered while executing the program in parallel. The OpenMP API uses the fork-join model of parallel execution [5] where a master thread executing the application spawns multiple threads; all the spawned threads execute the piece of program and then returns the control to the master thread. The order of execution of multiple threads can also be specified by the OpenMP compiler directives. The OpenMP provides directives such that the programs that are intended to execute serially can be executed in parallel easily.

2 Literature Review

The survey on the related work where either the morphological operations were applied in parallel or where other image processing techniques were applied in parallel is made, and the techniques used are reviewed; few of the notable works are as follows.

Görgünoğlu et al. in their work [6] proposed sequence of preprocessing steps involved in minutiae based fingerprint systems; here the performance analysis was made for the various preprocessing algorithms such as grayscale conversion, thinning implemented in parallel using OpenMP and CUDA, which resulted in the conclusion that few algorithms performed better in CUDA and few performed better in OpenMP.

Liu and Gao in their work [7] showed that a speedup was achieved when a serial program of cubic convolution interpolation algorithm was implemented in parallel on dual-core machine, which further showed the increase in speedup on a quad core machine. A comparison of the Thread Building Blocks (TBB), i.e., API available in C++ language versus OpenMP, was made with the change in the image sizes.

Valencia and Plaza in their work [1] implemented morphological opening and closing reconstruction operations for hyper-spectral image processing on different processors with different number of cores. Here the size of the structuring element was varied, and execution time on various numbers of cores was compared. Here on certain processors, few kernel sizes performed better than the other kernel size.

A. Plaza and J. Plaza in their work [8] implemented morphological operation over hyper-spectral images using NASA's thunderhead clusters, demonstrated the speedup with the increase in number of nodes in the cluster, and also proposed a method to reduce the inter-process communication among the nodes in the cluster.

It is notable that the above works have analyzed the parallel execution of the algorithms either on various processors or by changing the number of nodes in the cluster or on the different APIs; however, the works have not made the study on the number of threads executing the parallel program. Hence, through this work a comparison of number of threads executing the code and its execution time can be analyzed.

3 System Design

Here the overall design is to build a system that implements the morphological operations in parallel, then to analyze the hot spots using the Intel VTune Amplifier software, and then to analyze those hot spots that can be implemented in parallel; if the hot spots can be implemented in parallel, then we parallelize those sections of the program one by one and then compare the serial and parallel versions of the program by changing the number of threads.

Step 1. Read the input image file, which is the grayscale image.
Step 2. Implement morphological operations such as dilation, erosion, and edge detection on the given image.
Step 3. Using Intel VTune Amplifier, analyze the hot spots in the execution of the program. Hot spots are the parts of the program that take long time to execute.
Step 4. Identify the hot spots that can be implemented in parallel.
Step 5. If the hot spots can be implemented in parallel, build the parallel version of the serial program by parallelizing the hot spots. Repeat Step 3 to check whether any other section of the program becomes the hot spot.
Step 6. Compare the execution times of the serial versus parallel program and also vary the number of threads executing the program and compare its execution time with the serial version.

4 Results and Discussion

The test bench used to implement this work had the following specification:

Operating System: Ubuntu 16.04 LTS, Memory (RAM): 16 GB.

Processor details: Model: Intel(R) Core(TM) i7-6500U CPU @ 2.50 GHz, Number of Physical Cores: 2, Number of Logical Cores: 4, CPU max MHz: 3100.0000, CPU min MHz: 400.0000.

Cache details: L1d cache: 32 K, L1i cache: 32 K, L2 cache: 256 K, L3 cache: 4096 K.

The implementation contains the three morphological operations, i.e., dilation, erosion, and edge detection. The execution time by varying the number of thread is the main portion of the results; however, the memory utilization has also been taken into the consideration, and the results are discussed below. Intel VTune Amplifier's basic hot spot analysis feature is used to analyze the hot spots in the implementation. It was observed that the main hot spots occurred when the operations over the matrix containing the pixels were carried out. Since there were no dependencies between the pixel values, it could be parallelized easily, for which OpenMP compiler directives were used. The sections of the programs that could be parallelized were erosion, dilation as well as edge extraction operation.

Tables 1 and 2 show the various execution times in seconds when the input image of 3000 * 3000 pixels and 6000 * 6000 pixels were considered. In the studies, the sections that were parallelized were 'erosion,' 'erosion and dilation,' and 'erosion, dilation, and edge detection'. The readings considered are: total elapsed time of the program, idle time for the program, and the difference between the elapsed and idle times, with varying thread count with different parallel sections of the program.

Figure 1 shows the graphs for elapsed time, and Figs. 2 and 3 shows the graphs of idle time and difference between the elapsed and idle time for the execution of operations on image of size 3000 * 3000 pixels, by varying the thread count (2, 4, 8, 16) and by parallelizing the sections of 'erosion,' 'erosion and dilation,' and 'erosion, dilation, and edge detection'. Here we can observe that the performance is higher in parallel execution than in serial execution; however, the execution time increases with increase in the number of threads. Idle time is the time when the processor does not work; it occurs due to pipeline stalls due to demand load and store of the memory. Here, it can be evidently seen that Figs. 1 and 3 are identical, as the idle time is almost negligible.

Figures 4, 5, and 6 are the results obtained on image of size 6000 * 6000 pixels; however, Figs. 4 and 6 show that the performance reduces when the program is parallelized and further the performance reduces with the increase in the number of threads, hence further studies can be made by looking into the memory-related issues for further analysis.

The following are the results observed when the memory analysis for the same programs was made.

Tables 3 and 4 show the memory-related issues when the input images of 3000 * 3000 pixels and 6000 * 6000 pixels were considered. The readings considered are memory bound in percentage and sum of loads and stores, with varying thread count with different parallel sections of the program.

Figure 7 is the graph showing the memory bound for image of size 3000 * 3000 pixels, i.e., the percentage of time spent by the program performing memory-related operation and no other processing work, and it includes the demand load and store operations, and cache miss's where the data has to be fetched from the memory. It can be observed that increase in thread count as well as the number of parallel sections also increased the memory bound of the program. Figure 7 is the graph showing the load and store count for image of size 3000 * 3000 pixels. It can be observed that increase in thread count as well as the number of parallel sections also increased the load and store operations. Figures 9 and 10 are the graph showing the memory bound and total loads and stores count for image of size 6000 * 6000. It can be observed that, memory-related issues such as memory bound percentage, load and store count increases with increase in number of thread, which may be one of the overhead for the program, increasing its latency, thereby reducing the performance.

Table 1 Various execution times in seconds for image size 3000 * 3000 pixels

	Parallelizing erosion					Parallelizing erosion and dilation				
Thread count	1	2	4	8	16	1	2	4	8	16
Elapsed time	1.749	1.925	0.299	0.353	0.636	1.749	0.215	0.275	0.439	0.737
Idle	0.048574	0.053179	0.012244	0.014306	0.11317	0.048574	0.003893	0.00726	0.02838	0.010659
Elapsed-idle	1.700426	1.871821	0.286756	0.338694	0.52283	1.700426	0.211107	0.26774	0.41062	0.726341

	Parallelizing erosion, dilation and edge detection				
Thread count	1	2	4	8	16
Elapsed time	1.749	0.23	0.282	0.515	0.764
Idle	0.048574	0.007305	0.001023	0.063985	0.034982
Elapsed-idle	1.700426	0.222695	0.280977	0.451015	0.729018

Table 2 Various execution times in seconds for image size 6000 * 6000 pixels

	Parallelizing erosion					Parallelizing erosion and dilation				
Thread count	1	2	4	8	16	1	2	4	8	16
Elapsed time	7.094	7.146	8.186	9.717	14.397	7.094	8.259	8.842	12.352	19.991
Idle	0.40042	0.0456	0.3518	0.02803	1.05969	0.40042	0.731	0.0016	0.02898	0.03175
Elapsed-idle	6.69358	7.1004	7.8342	9.68897	13.33731	6.69358	7.528	8.8404	12.32302	19.95925

	Parallelizing erosion, dilation and edge detection				
Thread count	1	2	4	8	16
Elapsed time	7.094	7.299	9.238	13.397	20.684
Idle	0.40042	0.00944	0.12689	0.06548	0.0291
Elapsed-idle	6.69358	7.28956	9.11111	13.33152	20.6549

Table 3 Various memory-related operations for image of size 3000 * 3000 pixels

	Parallelizing erosion					Parallelizing erosion and dilation				
Thread count	1	2	4	8	16	1	2	4	8	16
Memory bound	1.7	2.7	8.5	8.2	15.3	1.7	3.6	7.4	15.9	17
Loads + stores count	7,500,390,000	8,571,815,034	11,553,094,588	15,023,088,416	23,559,818,807	7,500,390,000	9,154,132,995	13,865,258,890	22,551,091,297	40,256,672,386

	Parallelizing erosion, dilation and morphology				
Thread count	1	2	4	8	16
Memory bound	1.7	2.3	6.5	16	17.2
Loads + stores count	7,500,390,000	9,022,526,485	14,782,281,654	24,622,450,538	40,797,084,797

Table 4 Various memory-related operations for image of size 6000 * 6000 pixels

	Parallelizing erosion					Parallelizing erosion and dilation				
Threads	1	2	4	8	16	1	2	4	8	16
Memory bound	1.9	2.5	6.2	7.6	16	1.9	2.6	6.8	15	17
Loads + stores count	29,844,410,000	34,101,155,417	42,185,174,381	59,112,335,121	93,530,863,300	29,844,410,000	37,195,442,585	55,252,427,307	8.8278E+10	1.5697E+11

	Parallelizing erosion, dilation and morphology				
Threads	1	2	4	8	16
Memory bound	1.9	2.4	6.9	15.1	16.6
Loads + stores count	29,844,410,000	37,674,958,650	56,051,151,835	91,636,610,443	1.64E+11

5 Conclusion and Future Works

Through this work, from the Figs. 1, 2, 3, 4, 5, and 6, it can be observed that the increase in the number of threads may or may not improve the performance of the parallel program. Also, it is observed that increase in the number of parallel portions in the serial program need not improve the performance of the program. The main reason is the memory-related issues, which are time consuming; from Figs. 7, 8, 9, and 10, we can infer that the memory-related operations may increase with the

Fig. 1 Graph showing the elapsed times for image of size 3000 * 3000 pixels

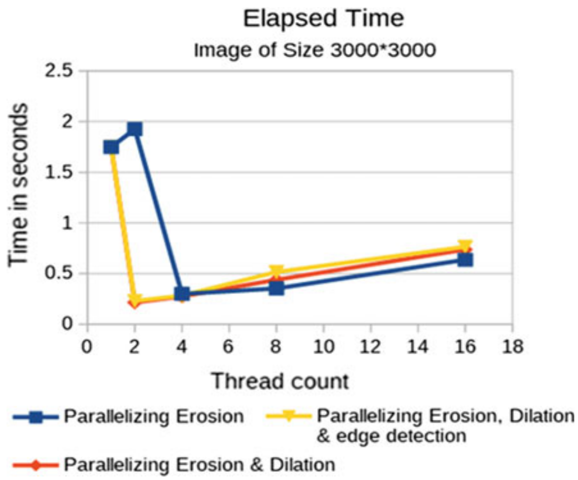


Fig. 2 Graph showing the elapsed times image of size 3000 * 3000 pixels

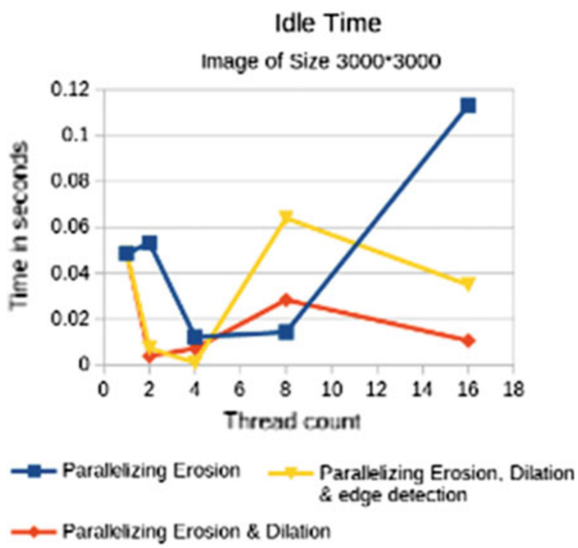


Fig. 3 Graph showing the difference in elapsed and idle time for image of size 3000 * 3000 pixels

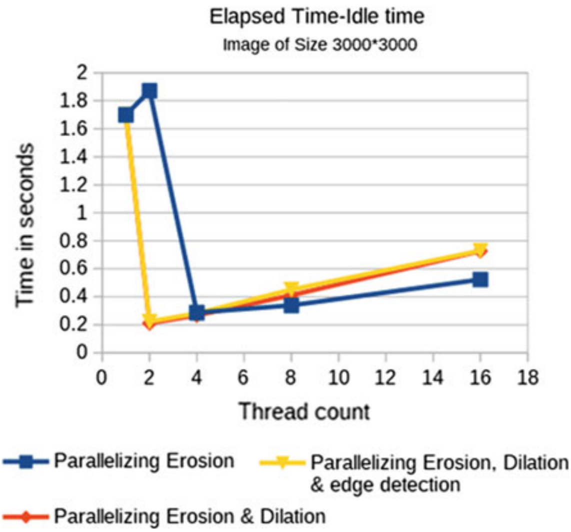
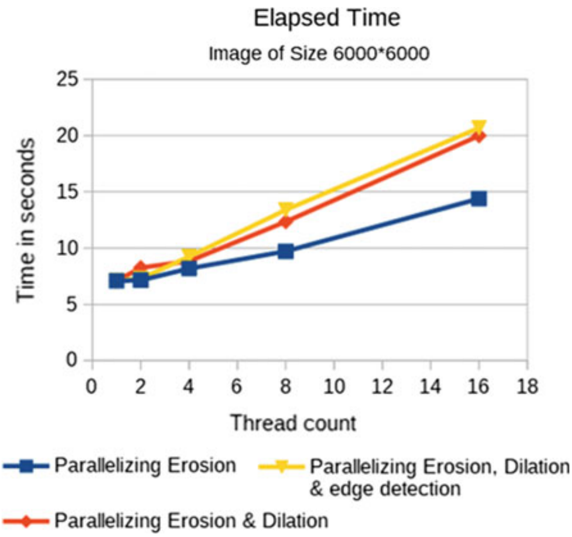


Fig. 4 Graph showing the elapsed times for image of size 6000 * 6000 pixels



increase in the number of threads. The increase in the memory-related operation when the number of threads increases, usually occurs due to the false sharing of the cache, which is the one of the most common phenomena observed in multi-threading applications. This work can be further carried out in future by varying the image size considered, thereby providing a better analysis. The results

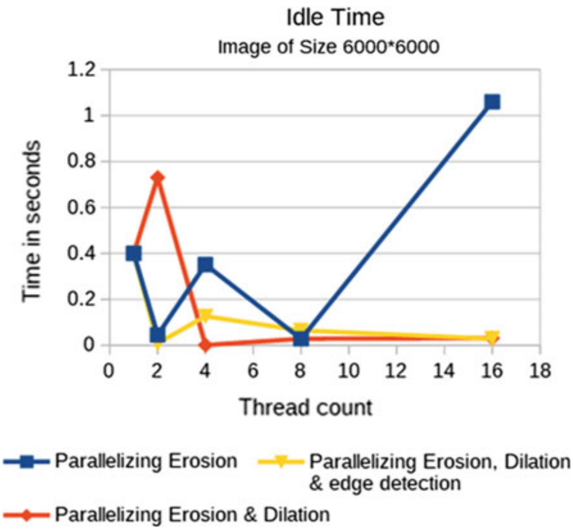


Fig. 5 Graph showing the idle times for image of size 6000 * 6000 pixels

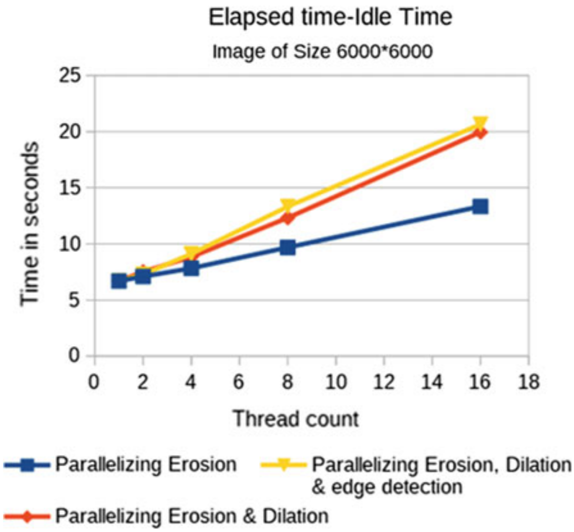


Fig. 6 Graph showing the difference in elapsed and idle time for image size 6000 * 6000 pixels

obtained is not generic, but implementation, APIs and platform dependent. Hence, carrying out of this work on different platforms, APIs would yield a generic solution. An optimization can be performed in the implementation, so that multi-threading would yield a better solution than the serial program.

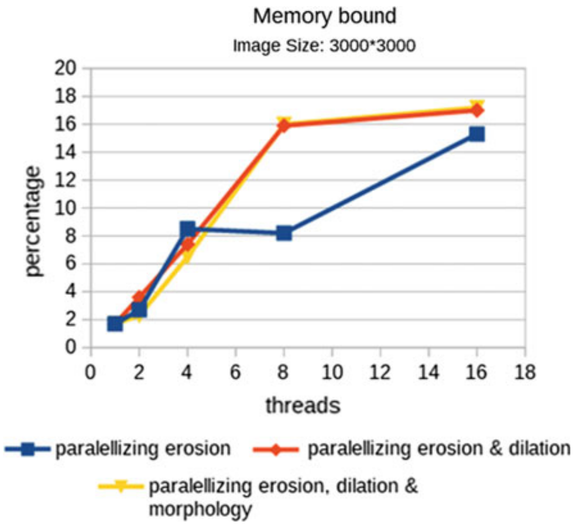


Fig. 7 Graph showing the memory bound in percentage for image of size 3000 * 3000 pixels

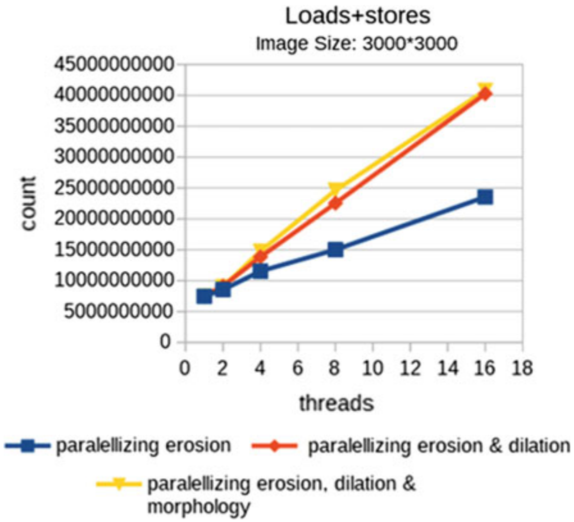


Fig. 8 Graph showing the load and store count for image of size 3000 * 3000 pixels

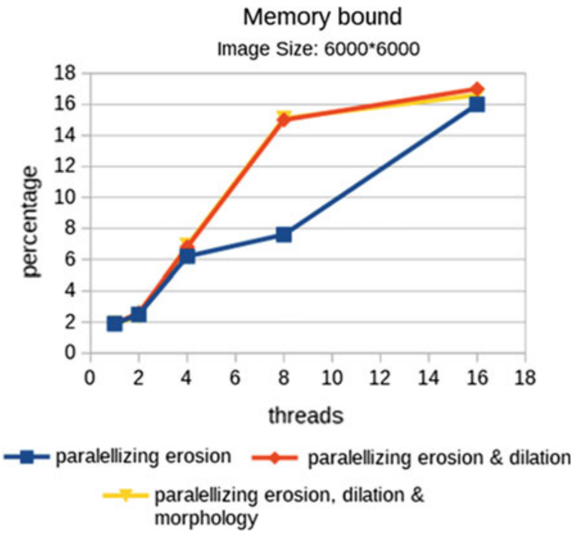


Fig. 9 Graph showing the memory bound in percentage for image of size 6000 * 6000 pixels

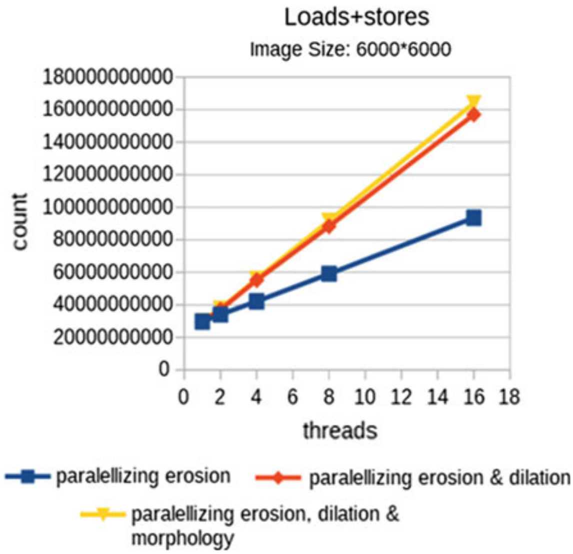


Fig. 10 Graph showing the load and store count for image of size 6000 * 6000 pixels

Acknowledgements The work reported in this paper is supported by the college [BMSCE, Bengaluru] through the TECHNICAL EDUCATION QUALITY IMPROVEMENT PROGRAMME [TEQIP-III] of the MHRD, Government of India.

References

1. Valencia, D., Plaza, A.: Efficient implementation of morphological opening and closing by reconstruction on multi-core parallel systems. In: IEEE Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (2009)
2. Kowalczyk, M., Koza, P., Kupidura, P., Marciniak, J.: Application of mathematical morphology operations for simplification and improvement of correlation of image in close-range photomography. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVII, Part B5 Beijing (2008)
3. Gonzales, R.C., Woods, R.E.: Image segmentation. In: Digital Image Processing, 3rd edn. Pearson Printice Hall, New Jersey (2008)
4. Akhter, S., Roberts, J.: Multi-core Programming—Increasing Performance Through Software Multi-threading
5. OpenMP Application Program Interface, Version 4.0, July 2013. http://www.openmp.org/wp-content/uploads/OpenMP4.0.0.pdf
6. Görgünoğlu, S., Orak, İ.M., Çavuşoğlu, A., Gök, M.: Examination of speed contribution of parallelization for several fingerprint pre-processing algorithms. In: Advances in Electrical and Computer Engineering, vol. 14, no. 2 (2014)
7. Liu, Y., Gao, F.: Parallel implementations of image processing algorithms on multi-core. In: IEEE Fourth International Conference on Genetic and Evolutionary Computing, pp. 71–74 (2010)
8. Plaza, A., Plaza, J.: Parallel morphological classification of hyperspectral imagery using extended opening and closing by reconstruction operations. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008 (2008)

US Dollar's Influence on Indian Gold Price: Assessment Using Artificial Neural Network

Deepa Bogale, Aniket Muley, Parag Bhalchandra and Govind Kulkarni

Abstract The main focus of this study is to highlight the inter-relationship between US dollar's rate and gold price in India. To study the influence of US dollar on gold price in India. The aim of this investigation is to identify the behavior of gold price data pattern. To study the functional relationship between the gold price and influencing parameter viz. US dollar price using artificial neural network (ANN) modelling technique. The back propagation algorithm is employed for analyzing the gold prices. In this research, taking the output of the neural network, it has been implemented using R software.

Keywords ANN · Back propagation · Gold price · US dollar price

1 Introduction

In India, gold rates depend on a host of factors, including international gold prices, currency rate movements, and also the local tariffs. In India, gold prices tend to move based on international prices due to currency fluctuations and a host of other factors, viz., decision on interest rates and inflation. However, physical demand also plays a big role in the movement of gold. If there is excess liquidity in the system

D. Bogale · P. Bhalchandra · G. Kulkarni
School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded 431606, Maharashtra, India
e-mail: latadsy@gmail.com

P. Bhalchandra
e-mail: srtmun.parag@gmail.com

G. Kulkarni
e-mail: govindcoolkarni@gmail.com

A. Muley (✉)
School of Mathematical Sciences, Swami Ramanand Teerth Marathwada University, Nanded 431606, Maharashtra, India
e-mail: aniket.muley@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_8

then it could move higher, as Gold Exchange Traded Funds tend to mop-up gold. Another important factor is central banks make their purchases. The USA has the highest gold reserves in the world. When many central banks start buying gold, it tends to affect gold prices across the globe including India. However, they rarely sell in tandem, to avoid disrupting prices of the precious metal. So, all these factors influence gold rates in India today and always whatever may be the today's gold price, it would always be different than yesterday's price. However, on Sunday in India, gold prices do not change, as there is no trading that is done [1, 2]. The results of the research showed that the former can predict better in comparison to the latter [3–19].

In the next subsequent sections, methodology, result and discussion, and conclusion, these topics are discussed in detail.

2 Methodology

Artificial neural network (ANN) (Fig. 1) converges the mathematical model of the ANN by means of simple neuron with a set of input nodes and associated weights pass through the nonlinear sigmoid activation function for data normalization and controls the output values (Fig. 2).

In the model, simple neuron with a set of input nodes are evaluated with its associated weights to process each input value. Further, it is multiplied by associated weights and apply the summation to get the single value is then pass through

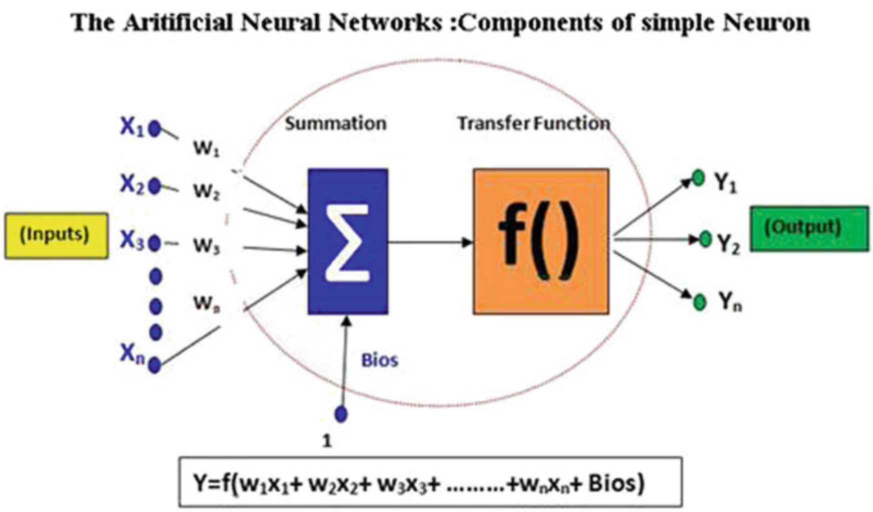


Fig. 1 Mathematical aspects of ANN [19]

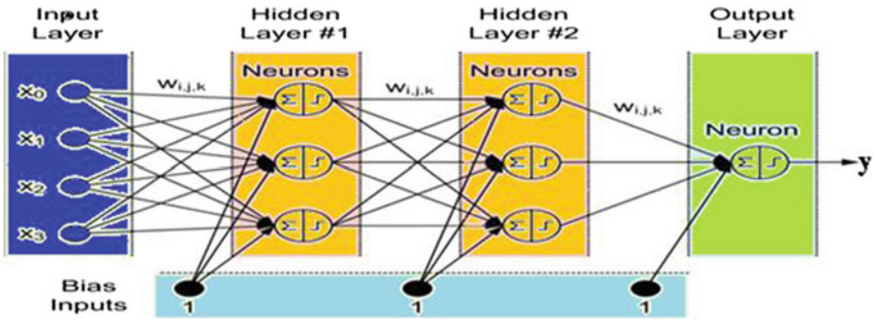


Fig. 2 Different layers of the ANN architecture [9]

the nonlinear activation function called sigmoid activation function (Eq. 1) for data normalization and control the result values [20–26].

Set of input nodes = $(x_1, x_2, x_3, \ldots, x_n)$, associated weights = $(w_1, w_2, w_3, \ldots, w_n)$, and output nodes = $(y_1, y_2, y_3, \ldots, y_n)$.

$$logsig(n) = \frac{1}{(1 + e^{(-n)})} \tag{1}$$

The ANN is presented in terms of the input layer, hidden layer, and the output layer [9].

The proposed algorithm for optimal ANN architecture is represented in the following way:

1. To predict US Dollar's price connection with Gold price by ANN model based on the seasonal data by splitting it into training set (70%) and test set (30%).
2. Standardize the data to reduce in weights.
3. Develop R program to find the optimal ANN model with hidden layer and 1–10 hidden neurons on trial and error bases.
4. The Levenberg–Marquardt back propagation algorithm with weight backtracking algorithm and cross-entropy differentiable error function is considered, and logistic function used as an activation function is a proposed ANN model.
5. The data is normalized before subjected to training and testing by transforming the data to the range of 0–1 by Eq. (2)

$$\mathrm{Xn} = \frac{\mathrm{x} - \mathrm{Xmin}}{\mathrm{Xmax} - \mathrm{Xmin}} \tag{2}$$

6. After the completion of neural networks for training dataset, the network's output values are normalized and it needs denormalization transforming into the actual value. The equation is defined as Eq. (3)

$$X = X_n \times (X_{max} - X_{min}) + X_{min}. \tag{3}$$

where X is the original value, X_{min} and X_{max} are the minimum and maximum values in the series, respectively, and X_n is the normalized data.

7. Test the efficiency of sensitivity coefficients of input variables.
8. To obtain optimal ANN model by using different structures to choose the best network. ANN model is validated mathematically by RMSE, MAE, and MAPE values. Simulate the process until obtaining least RMSE, MAE, and MAPE values closer to one with least number of steps and error to conclude the process. These are calculated through Eqs. (4–6), and the best representation is recognized according to the performance criteria counting by RMSE, MAE, and MAPE as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (A_i - P_i)^2}{N}} \tag{4}$$

$$MAE = \frac{\sum_{i=1}^{N} (A_i - P_i)}{N} \tag{5}$$

$$MAPE = \frac{\sum_{i=1}^{N} (A_i - P_i)}{N} \times 100 \tag{6}$$

where P_i—forecasted values, A_i—observed values, and N—number of datasets.

3 Result and Discussion

In the present study, secondary data has been composed of Reserve Bank of India's (RBI's) data warehouse [16]. The 25 years' gold price and dollar prices' monthly data is considered from March 1992 to July 2016; it consists 293 observations. We have used Indian rupee/US dollar exchange rate for our study keeping in view the fact that the USA is the major trading partner of India and of course the US dollar is regarded as the best hedging currency in the world. The data is collected from Reserve Bank of India's database. The price of gold in US dollar is taken as base price and the price of gold is converted into rupees, and import duties are added to the converted value. The study period has been chosen purely on the basis of availability of data and keeping in mind that neural network estimation requires a long-time user.

In our research, we have used R software to develop our model. The neuralnet package of the R software is specifically used for implementation. The dataset contains information on the variables: gold price (GP) and USD exchange rate (DP). Our goal is to devise a model which predicts, based on the input variables, whether or not a default will occur within given years. The neuralnet package uses back propagation with weight backtracking as its standard algorithm.

Tables 3 and 4 are obtained by applying smallest absolute ratio (slr) algorithm. The results are as follows.

Based on the Tables 1 and 2, obtained results of the proposed optimal model are represented in Figs. 3 and 4. It is observed that, the efficiency of the proposed model is shown through line diagram at the ninth hidden neuron.

Table 1 Hidden neurons from 1 to 10 for train set (Gold Price)

Hidden neurons	Number of steps	Error	Time (s)	RMSE	MAE	MAPE
1	5	0.9357	0.63	0.7177	0.696304	69.6304
2	22	0.9162	0.95	0.7092	0.688064	68.8064
3	5	0.9482	0	0.7197	0.698016	69.8016
4	7	0.9844	0.02	0.7372	0.715230	71.5230
5	7	0.9828	0.02	0.7352	0.713874	71.3874
6	11	0.9729	0.02	0.7330	0.711045	71.1045
7	6	1.0520	0.02	0.7519	0.703093	70.3093
8	4	0.9526	0.01	0.7199	0.698169	69.8169
9	**9**	**0.9051**	**0.03**	**0.7004**	**0.679800**	**67.9800**
10	5	1.0020	0.02	0.7425	0.72067	72.067

Table 2 Hidden neurons from 1 to 10 for test set (Gold Price)

Hidden neurons	Number of steps	Error	Time (s)	RMSE	MAE	MAPE
1	20	1.2760	0.49	0.70471	0.698441	69.8441
2	7	1.2117	0.92	0.68945	0.683195	68.3195
3	13	1.2827	0.09	0.70049	0.694086	69.4086
4	10	1.2730	0	0.70300	0.696654	69.6654
5	15	0.8560	0.03	0.58397	0.576890	57.6890
6	8	1.2038	0.02	0.68738	0.681174	68.1174
7	32	0.7143	0.05	0.51956	0.508653	50.8653
8	16	0.9306	0.12	0.61045	0.604052	60.4052
9	**17**	**0.6822**	**0.02**	**0.50753**	**0.495949**	**49.5949**
10	19	0.87434	0.12	0.57667	0.569140	56.9140

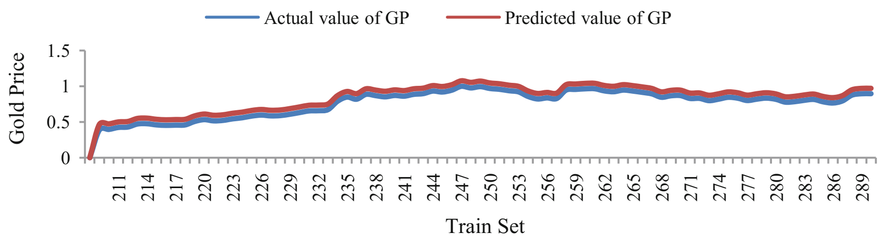


Fig. 3 Actual versus predicted observation line plot for training dataset

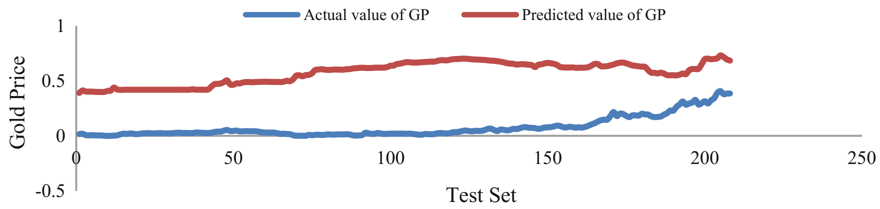


Fig. 4 Actual versus predicted observation line plot for testing dataset

4 Conclusion

In this study, it is observed that, gold prices are heavily dependent on the US dollar. Gold is an international traded commodity and US dollar is the preferred international currency. Any changes within the US dollar are bound to affect gold prices either directly or indirectly. Tables 1 and 2 represent the sensitivity analysis of the proposed algorithm for gold price (GP) and US dollar rate (DP) by the proposed algorithm, and further it is compared with Tables 3 and 4. It is observed that, the proposed algorithm gives more consistent results as compared with slr algorithm with nine hidden neurons as well as the RMSE, MAE, and MARE values. We have tried to get certain level up to ten hidden neurons for train and test set dataset. Table 1 illustrated that, for train set data at nine hidden neurons we have obtained the more accurate results; i.e., error (0.9051) is minimum and coefficient of determination RSME (0.7004) is least as compared to others. Similarly, Table 2 explores that the result obtained at the ninth hidden neuron found to be more accurate result; i.e., error (0.6822) is minimum; RSME (0.5075) is least as compared to others. The variation among the optimal network for train set (1-9-1) and test set (1-9-1) is different; it is due to the sample observation in the respective dataset. This model is helpful in forecasting gold prices. In future, one can increase

Table 3 Hidden neurons from 1 to 10 for train set (Dollar's Price)

Hidden neurons	Number of steps	Error	Time (s)	RMSE	MAE	MAPE
1	1	0.9312	0	0.7193	0.6980	69.80
2	**26**	**0.912**	**0.02**	**0.7076**	**0.6866**	**68.66**
3	14	1.017	0.01	0.7377	0.7154	71.54
4	11	1.072	0	0.7582	0.73580	73.58
5	5	1.033	0.01	0.7439	0.7216	72.16
6	9	0.937	0	0.7188	0.6974	69.74
7	3	0.949	0	0.7269	0.7055	70.55
8	10	1.003	0.01	0.7397	0.7182	71.82
9	6	1.035	0	0.7506	0.7286	72.86
10	5	1.062	0	0.7546	0.7328	73.28

Table 4 Hidden neurons from 1 to 10 for test set (Dollar's Price)

Hidden neurons	Number of steps	Error	Time (s)	RMSE	MAE	MAPE
1	11	1.299	0	0.7096	0.70327	70.327
2	8	1.211	0	0.6864	0.6803	68.03
3	8	1.298	0	0.7021	0.6957	69.57
4	6	1.262	0	0.7006	0.6943	69.43
5	21	0.668	0.01	0.4955	0.4813	48.13
6	23	0.701	0.01	0.5318	0.5232	52.32
7	6	1.015	0	0.6212	0.6150	61.50
8	19	0.852	0.14	0.5701	0.5623	56.23
9	**16**	**0.654**	**0.01**	**0.4839**	**0.4664**	**46.64**
10	8	1.187	0	0.6823	0.6761	67.61

the input parameters to check whether the accuracy level will be maintained or not. Also, one can perform the sensitivity analysis by taking number of iterations up to their desirable limits.

References

1. Ghezelbash, A.: Predicting changes in stock index and gold prices to neural network approach. TJMCS **4**(2), 227–236 (2012)
2. Nirkhi, S.: Potential use of Artificial Neural Network in Data Mining, vol. 2, pp. 339–343. IEEE (2010)
3. Mehr, A., Ahrari, A., Ghasemi, M.: Pattern making and forecast of economic growth of Iran using GMDH neural network. Econ. Res. Mag. **88** (2009)
4. Mehr, A., Ahrari, A., Kia, S.: Modelling and forecast of petrol price using GMDH neural network. In: Seasonal Magazine of Economic Researches of Iran, Faculty of economy, Allameh Tabatabaee, vol. 36 (2008)
5. Chen, W.Y., Lian, K.K.: A comparison of forecasting models for asian equity markets. Sunway Acad. J. **2**(4) (2005)
6. Solaimani, K.: Rainfall-runoff prediction based on artificial neural network. Am. Eurasian J. Agric. Environ. Sci. **5**(6), 856–865 (2009)
7. Sibi, P., Allwyn Jones, S., Siddarth, P.: Analysis of different activation functions using back propagation neural networks. J. Theor. Appl. Inf. Technol. **47**(3), 1264–1268 (2013)
8. http://zeenews.india.com/business/gold/story.aspx?newsid=397
9. http://www.mygoldguide.in/why-gold/gold-purity-tips-difference-between-24k-22k-18k-gold
10. http://www.rediff.com/money/report/gold1/20071116.htm
11. https://www.bankbazaar.com/gold-rate/gold-rate-trend-in-india.html
12. http://www.economist.com/node/9401752
13. http://www.goodreturns.in/gold-rates
14. https://www.rbi.org.in/scripts/annualPublications.aspx?head=Handbook%20of%20Statistics%20on%20Indian%20Economy
15. https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf
16. http://www.goldbullionpro.com/top-10-factors-affecting-gold-price-in-india-today
17. http://www.indiainfoline.com/article/article-latest/factors-affecting-gold-prices-in-india-117071800350_1.html

18. https://www.quora.com/What-impacts-gold-price-in-India-globally-How-is-the-U-S-dollar-related-to-or-affected-by-gold. What drives the price of gold?
19. http://www.investopedia.com/financial-edge/0311/what-drives-the-price-of-gold.aspx#ixzz4nCuUXhK2
20. Lin, J.: Empirical Study of Gold Price Based on ARIMA and GARCH Models. Stockholm's Universities (2010)
21. Brauner, E.O., Dayhoff, J.E., Xiaoyun, S., Hormby, S.: Neural network training techniques for a gold trading model. In: Computational Intelligence for Financial Engineering (CIFEr), vol. 23, no. 25, pp. 57–63. IEEE/IAFE (1997)
22. Cheng, B., Titterington, D.M.: Neural networks: a review from a statistical perspective. Stat. Sci. **9**(1), 2–30 (1994)
23. Eluyode, O.S., Akomolafe, D.T.: Comparative Study of biological and artificial neural networks. Eur. J. Appl. Eng. Sci. Res. **2**(1), 36–46 (2013)
24. Karsoliya, S.: Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. Int. J. Eng. Trends Technol. **3**(6), 714–717 (2012)
25. Soni, A.K., Abdullahi, A.U.: Using neural networks for credit scoring. Int. J. Sci. Technol. Manag. **04** (2015)
26. Swanston, D.J., Bishop, J.M., Mitchell, R.J.: Simple adaptive momentum: new algorithm for training multilayer perceptrons. J. Electron. Lett. **30**, 1498–1500 (1994)

Attribute Based Storage Mechanism with De-duplication Filter: A Review Paper

Amarja Hanumant Bhosale and Amrita A. Manjrekar

Abstract Cloud computing relies on sharing of resources to achieve coherence and economy of scale, similar to a utility. In existing system ABE (attribute based encryption) is used to maintain security. Attribute-based encryption is a type of public-key encryption in which the secret key of a user and the cipher-text are dependent upon attributes. In computing, data de-duplication is a specialized data compression technique for eliminating duplicate copies of repeating data. Related and somewhat synonymous terms are intelligent compression and single-instance storage. In this paper we studied the different attribute storage mechanisms as well as de-duplication system.

Keywords Attribute based encryption · De-duplication · Storage Access policy

1 Introduction

Cloud computing is one of the fast growing trends. Cloud provides effortless, well-organized area to accumulate data, sheltered data, and right to use information from any place with the assist of internet. It also gives bendable infrastructures, space for storing data and better performance.

Cloud storage has become an indispensable part for various applications in today's network. With the fast developing of cloud storage services [1], various encryption techniques turns into a vital system for securing the privacy of information.

A. H. Bhosale (✉) · A. A. Manjrekar
Department of Technology, M. Tech. Computer Science and Technology, Shivaji University, Kolhapur, Maharashtra, India
e-mail: amarjabhosale123@gmail.com

A. A. Manjrekar
e-mail: aam-tech@unishivaji.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_9

Important factor in cloud storage are Data secrecy and the space required to store it. Lots of cryptographic algorithms are available to keep data securely from illegal user. Data replication is important for make sure availability sometime data gets replicate in same node with different parameters and this kind of waste of storage is reduced by data de-duplication techniques.

Proposed systems uses the cipher text policies based/attribute based [2] encryption in which data secrecy maintained by using some factors like data itself, type of encryption algorithm, key size used while performing encryption.

Let sees some reviews and discussions connected to sharing and confidentiality of information over cloud which already exists.

2 Literature Survey

There are different policies defined on the subject of security and data sharing in cloud. Few important are described below.

2.1 Secure De-duplication with Attribute Based Storage Scheme

In this paper authors addressed [3] Attribute Based Storage Supporting Secure De-duplication of Encrypted Data. In Cloud Attribute based storage capacity supporting secure de-duplication Standard ABE framework does not support the protected de-duplication framework. In this paper author introduced a characteristic based framework for putting away information with de-duplication utilizing half kind cloud foundation. In this entire situation open cloud utilized for capacity reason on other hand private cloud is in charge of copy location This framework has two points of interest; first it can be utilized to secretly impart information to clients with indicating access arrangements and another is it accomplishes the standard idea of semantic security for information privacy while existing frameworks just accomplish it by characterizing a weaker security thought.

2.2 Data De-duplication with Randomized Tag

The proposed system [4] gives information regarding Secure and Efficient Cloud Data De-duplication with Randomized Tag. This framework presents secure framework manufacturing which connect together the progressions of video coding procedures and secures de-duplication. The plan is likewise precisely custom fitted to versatile-video-coding (SVC) methods for the help of wide-ranging systems and gadget for great versatile video diffusion.

2.3 Encrypted Cloud with Secure De-duplication

The authors discussed Encrypted Cloud Media Center with Secure De-duplication. As of late, Abadi et al. presented the primal MLE-2 [5] by way of satisfying security properties for secure and productive information de-duplication. Completely randomized plan (RMLE-2) requires the disorganized sense of balance testing calculation to distinguish between each and every one reproduction figures. In this approach, an appealing testing issue is the methods by which to decrease the overhead of RMLE-2 and proposed a competent advancement for RMLE-2.

2.4 ABE Scheme in Fog Communications

This article's writer proposed [6] an ABE Scheme for Fog Communications. In this paper creator proposed a proficient key trade system base on CP-ABE to set up secure communications in between the member. ABE methods used to achieve a high level of the client's protection and fine-grained information get to control. In this protocol, creators used the digitalized mark called signature and CPABE strategies to get the prime security objective: privacy, verification, verifiability, and access control.

2.5 ABE for Data Sharing

Authors implemented Collaborative Key Management Protocol for Cloud Data Sharing. Figure content display trait based encryption (CP-ABE) is a promising cryptographic method for fine-grained get to control of outsourced in arrange in the cloud. According to authors [7] outline there are a few downsides of key management damage the importance of its application. One downside in dire need of arrangement is the key escrow issue. In this paper, they proposed a community enter management convention in CP-ABE. In which creators acquainted trait bunch with produce a private key refresh calculation for fine-grained and quick property rejection.

2.6 Privacy-Preserving Online/Offline

In this article authors proposed [8] Privacy-Preserving Online/Offline and Outsourced Multi-Authority Attribute-Based Encryption. Furthermore, extensive ABE plans among different experts (multi-specialist ABE) are more reasonable for practical applications than essential single expert ABE plans. All things measured, existing multi-specialist ABE plots either can't save get to planning safety or support costly computational cost of encryption and decode stage. In this paper author gave the answer for resolve the issues about the calculation load for the two proprietor and clients. The main thought was to moderate the online calculation overhead for proprietor by part the encryption calculation to the online encryption and disjointed encryption.

2.7 A Dynamic Resource Sharing

A Dynamic Resource Sharing Mechanism proposed by the respective authors [9]. In this paper, author examine the benefit sharing issue in a front pull appreciative C-RAN, where different authority organization rent radio assets from the system administrator to give out their group. To give detention among various specialist co-ops, they acquainted a limit based agreement with control the impedance among RRHs, and characterized another metric to give least asset documentation to specialist co-ops. After that they figure the asset designation issue as a MINLP issue, and change it into a blended number direct programming (MILP) issue utilizing a linearization method. In this paper creator make sense of the scheme attack in the exit plan and they intended a solid preparation.

The following graph shows how the technique related to storage security are developed over the time (Fig. 1).

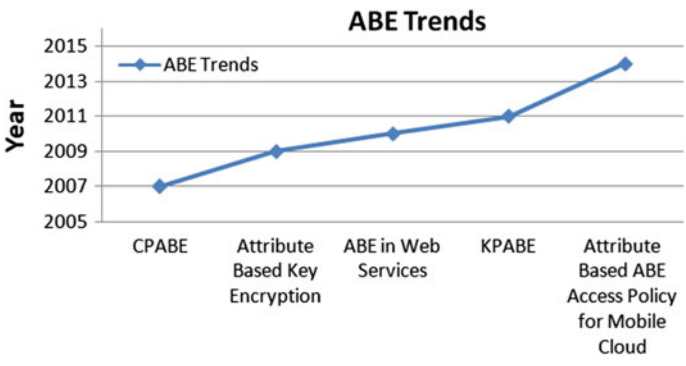


Fig. 1 Development of ABE trends

3 Comparisons Between Existing Algorithms

Here two main ABE algorithm are considered are as follows.

3.1 *KP-ABE (Attribute Based Encryption with Key-Policy)*

Attribute Based Encryption with Key Policy (KP-ABE) conspires is intended for one-to-many interchanges. KP-ABE plans are appropriate for organized associations with rules about who may read specific records.

3.2 *CP-ABE (Attribute Based Encryption with Cipher-Text Policy)*

In CP-ABE, every client is connected with an arrangement of characteristics. His mystery key is created in view of his characteristics. While scrambling a message, the encoded indicates the limit get to structure for his intrigued characteristics.

Table 1 compares these two algorithms based on different parameter.

Here K is the security parameter provide as input and generates two things in output first PK which is the public parameter and another one MK which is master private key for the system.

$$\text{Setup}\ (1^{k}) \rightarrow (\text{PK},\ \text{MK}).$$

Key Generation algorithm takes three things PK, MK and an additional set of attribute set (A) as the input, this algorithm generates an attribute based private key (SA) for the specific set of attribute A.

$$\text{KeyGen}\ (\text{PK},\ \text{SK},\ \text{A}) \rightarrow \text{SA}.$$

Here PK is the public Parameter, ‘M’ is the message and ‘A’ is an access structure, this algorithm done encryption and generate outputs ‘AS’ which is a

Table 1 Algorithm of ABE schemes

Sr. no.	Parameters	Algorithm	
		KP-ABE	*CP-ABE*
1	Initiation parameter	(K) (PK, MK)	(K) (PK, MK)
2	Encryption process	(M, PK, A) (CT)	(M, AS, PK) (CT)
3	Key generation	(MK, AS, PK) (SK)	(A, MK) (SK)
4	Decryption process	(SK, CT, PK) (M)	(CT, SK) (M)

trapdoor Key as well a tuple CT = (CT, T, L) where ‘T’ is the tag and ‘L’ is the label coupled with message M.

$$\text{Encryption (PK, M, A)} \rightarrow \text{(AS, CT)}.$$

Combination of label and cipher-text brace (L, CT), SK which is an attribute-based private-key, an attribute set—A as the input, this algorithm done decryption and provide output the original message M (Table 2 and Fig. 2).

$$\text{Decryption (PK, (L, CT), A, SK)} \rightarrow \text{M}$$

Table 2 Comparison of ABE schemes

Sr. no.	Parameters	Algorithm	
		KP-ABE	*CP-ABE*
1	Component linked with access policy	Data	Cipher text
2	Efficiency	Average	Average
3	Secured access control	Low	Average
4	Computational overhead	High	Average
5	Data confidentiality	No	Yes
6	User accountability	No	No
7	Scalability	No	Yes
8	User revocation	No	No
9	Collusion resistant	Yes	Yes
10	Restriction	Encrypter of data cannot decided	Decrypter allows only if user attribute that are organized logically

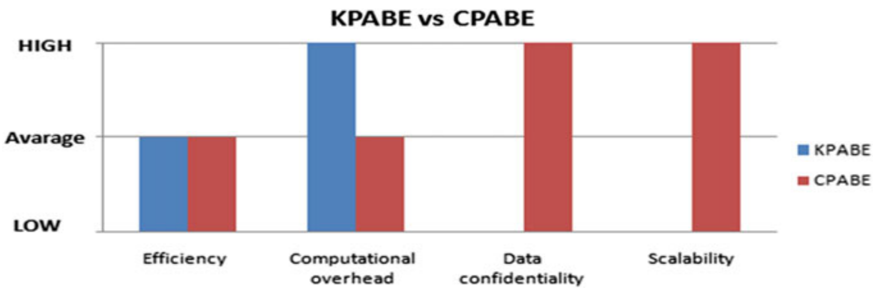


Fig. 2 Comparision graph between CPABE versus KPABE

3.3 Negative Aspect of ABE Schemes

3.3.1 ABE Restrictions

Owner of data needs to make utilize of each and every approved user's public key for encrypting information.

Due to monotonic attributes this scheme limited in the real environment.

3.3.2 KP-ABE Restrictions

The person who decrypts the encrypted data cannot decide.

Descriptive attributes required for the information.

Owner of data has to faith over the key issue party that's why this scheme unsuitable in some application.

3.3.3 CP-ABE Restrictions

Not able to fulfill the enterprise requirements such as considerable efficiency and flexibility.

Specifying policies and managing user attributes has some limitation (Table 3).

3.4 File-Level De-duplication

The file level de-duplication can be performed generally effortlessly. It requires less handling power since documents' hash numbers are moderately simple to produce. Be that as it may, there is the turnaround side of an award, if just a single byte of a record is changed, its hash number likewise changes.

3.5 Fixed-Size Block De-duplication

Block de-duplication requires more handling power than the document de-duplication, since the quantity of identifiers that should be prepared increments enormously. Correspondingly, its list for following the individual emphases gets likewise considerably bigger.

Table 3 Comparison of de-duplication schemes

Methods based on granularity	Place	Time
File-level de-duplication	Server-based de-duplication	Online de-duplication
Fixed-size block de-duplication	Client-based de-duplication	Offline de-duplication

4 Proposed Work

In this era efficiency matter that's why hybridness is one best solution over it. The combination of different technology, platform gives better result use as well private cloud used to host sensitive or critical workloads, and third party public cloud used to store actual data revocation module will be implemented to manage the data access. Data security and performance of the system is the key factors commencing the user's view point. The existing systems are made based on different encryption strategies such as ABE (attribute based encryption); CP-ABE (attribute based encryption with Cipher text policy) etc. duplication is also matters to minimize the storage cost that's why systems de-duplication are designed. The cost of storage space required to store data considered as an important factor.

Here we propose hybrid cloud based system to provide efficient data secrecy with attribute based policy. The system will also be consists of de-duplication filter which work either online or offline manner to avoid duplication of data, no matter file is encrypted or not. It will also consist of dynamic based revocation module to control defined data access.

5 Conclusion

There are different attribute based storage mechanism available in cloud computing. These policies are coupled with information and attribute related to that information. This information and attributes are associated with keys. This paper provides a review of CP-ABE (Cipher-text Policy Attribute Based Encryption) and KP-ABE (Key Policy Attribute Based Encryption) as well some de-duplication filter techniques. The comparison tables shows that CP-ABE better than KP-ABE with respect to some parameters. ABE schemes always preferred for providing security to sensitive information. These schemes provide additional scalability, flexibility and fine-grained way in control than any other. Storage factor problem can be minimized using different de-duplication techniques. In which hash value generated for comparing old data versus new data and avoid the unwanted replication of data over one single node.

References

1. Steiner, T.: SANS Institute Infosec reading room. An Introduction to Securing a Cloud Environment, Version June 2012. © 2012 The SANS Institute
2. Introduction to Cloud Computing. White paper, 2017. © Copyright 2017 Dialogic Corporation

3. Cui, H., Deng, R.H., Li, Y., Wu, G.: Attribute based storage supporting secure de-duplication of encrypted data in cloud. IEEE Trans. Big Data (2016). https://doi.org/10.1109/tbdata.2017.2656120
4. Zheng, Y., Yuan, X., Wang, X., Jiang, J., Wang, C., Member, IEEE, Gui, X.: Toward encrypted cloud media center with secure deduplication. IEEE Trans. Multimed. **19**(2) (2017)
5. Jiang, T., Chen, X., Wu, Q., Ma, J., Susilo, W., Lou, W.: Secure and efficient cloud data de-duplication with randomized tag. IEEE Trans. Inf. Forensics Secur. **12**(3) (2017)
6. Alrawais, A., Alhothaily, A., Hu, C., Xing, X., Cheng, X.: An attribute-based encryption scheme to secure fog communications. IEEE **5**, 2169–3536 (2017)
7. Lin, G., Hong, H., Sun, Z.: A collaborative key management protocol in cipher text policy attribute-based encryption for cloud data sharing, 28 June 2017, pp. 2169–3536. IEEE (2017)
8. Shao, G., Zhu, Y., Ji, Q.: Privacy-preserving online/offline and outsourced multi-authority attribute-based encryption, 978-1-5090-5507-4/17/$31.00, Wuhan, China. ©2017 IEEE ICIS 2017, 24–26 May 2017
9. Niu, B., Zhou, Y., Shah-Mansouri, H., Wong, V.W.S.: A dynamic resource sharing mechanism for cloud radio access networks. IEEE Trans. Wirel. Commun. **15**(12) (2016)

An Approach for Traffic Congestion Detection and Traffic Control System

Tarun Kumar and Dharmender Singh Kushwaha

Abstract The evolution of the computer technologies is playing important role in the development and planning of the smart transportation system for the public domain. The smart transportation management monitors and manages the traffic in real time. This paper proposes an approach for smart traffic monitoring and control system. The proposed system aims to provide least congestion routes for the emergency services. This approach detects the congestion and tracks the locations of the emergency vehicles in real time and notifies the traffic control room. The detection of the congestion is based on the images obtained from the cameras installed at various locations in the city. The RFID readers are used to detect and track the location of the emergency vehicles. Traffic control rooms control the traffic lights based on the congestion information and location of the emergency vehicles. The RFID module is simulated in NS2, and congestion detection module is implemented in OpenCV and Java. The proposed approach is suitable for creating green corridors for emergency vehicles.

Keywords Traffic congestion · Vehicle detection · RFID reader
RFID tags

1 Introduction

At present, the impact of the technologies on our life is increasing day by day. Many government bodies are integrating these technologies to provide better infrastructure, transportation, etc. Technologies such as the Internet of things (IoT) and the image processing bring revolutionary changes in smart traffic

T. Kumar (✉) · D. S. Kushwaha
Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, UP, India
e-mail: ertarun123@gmail.com

D. S. Kushwaha
e-mail: dsk@mnnit.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_10

surveillance and management system [1–3]. Merely adoption of these technologies to improve the transportation system may not be able to rule out the occurrence of traffic jams and rate of emergency vehicles getting stuck in traffic jams. The real-time detection of road traffic congestion may be used in traffic control system. The least congestion route may be provided to the emergency vehicles. The congestion of the road traffic may be obtained by detection of the vehicles and vehicles flow for the specific time period. The detection and recording of the road traffic congestion at various locations may be used for planning better infrastructure; i.e., the extra measures may be taken for highly congestion areas. The congestion conditions depend on the type of roads such as single lane road or multi-lane road, heavily crowded road or less crowded road. The system should be able to make the decision of the congestion based on the various real-time conditions.

2 Related Work

Terroso-Sáenz et al. [4] proposed an approach for road traffic congestion detection based on the vehicular ad hoc network and event-driven architecture. Road traffic congestion is recorded for the single and double lane roads. The approach uses vehicle-to-vehicle communication to estimate the road traffic congestion. Palubinskas et al. [5] used image processing-based approach for congestion detection. The image change detection technique is used for detection of the vehicles. The approach does not detect individual vehicle. The objective is to detect the density of the vehicle on the road by estimation of the space between the vehicles and average speed of the traffic flow, etc. Li et al. [6] estimated the road traffic congestion by processing the traffic video in time and spatial domain. The road occupation rate classifies the condition of the road as congested or non-congested regions. Idé et al. [7] detected the traffic flow by analyzing the real-time traffic videos. The approach uses image change detection for estimation of the traffic flow. An unsupervised vehicle counting algorithm is proposed to train the neural network. D'Andrea and Marcelloni [8] used GPS-based approach for the estimation of the real-time traffic congestion. The approach retrieves the GPS coordinates of the various smartphone users of the vehicles, and based on their location, the congestion and traffic flow are estimated. The estimated traffic congestion information is passed to the users. Most of these approaches are based on image processing, vehicle-to-vehicle communications, RFID readers, etc. These approaches provide the static information about the road traffic congestion. The approach proposed in this paper provides the real-time dynamic information for road traffic congestion. The congestion information is also utilized in controlling the real-time traffic. The status of the traffic signals may be changed based on the detection of any emergency vehicles.

3 Proposed Work

The proposed approach is based on the integration of the two different technologies as shown in Fig. 1a. An approach based on image processing is used to detect the traffic congestion, and RFID technology is used for tracking of the emergency vehicles. To detect the congestion, a background image is subtracted from the series of the frames. After background subtraction, pixels other than the background pixels are classified as foreground pixels. These foreground pixels represent the presence of the vehicles. The approach estimates the density of the traffic based on the total number of the detected vehicles and the width of the road. The congestion is classified into three different classes by using density value. The two-threshold policy is used to classify the congestion. The density below the first threshold is classified as least congestion road, the density between both thresholds is classified as moderate congestion, and the density above the second threshold is classified as heavy congestion road. The traffic control server controls the traffic based on the congestion matrix in real time. The congestion matrix is updated periodically. In the approach, RFID readers are also equipped with the cameras. These RFID readers detect the emergency vehicles in the real time and notify the traffic control server about the location of the emergency vehicle as shown in Fig. 1b. We assume that all the emergency vehicles must be equipped with passive RFID tag. The real-time congestion information and location of emergency vehicles are used by traffic control server. Traffic control server controls the traffic lights to provide the clearance for emergency vehicles. The congestion information is also stored into database that may be useful for planning the better transport infrastructure for future. The approach proposes an algorithm for detection of the vehicles estimation of the road traffic congestion.

Let there be total f_n frames captured in time t, at location l, and let the background image of the location l be b. Then, congestion can be defined as the ratio of

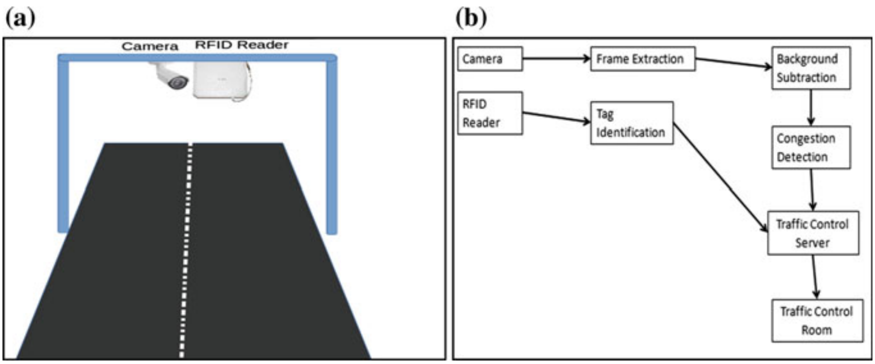


Fig. 1 **a** Installation of camera and RFID reader **b** work flow of proposed approach

total number of the pixels occupied by the vehicles P_f and total number of the pixels in background image P_b. So the density can be defined as:

$$d = \frac{1}{n}\sum_{0}^{n}\frac{P_{fi}}{P_{bi}} \tag{1}$$

Algorithm: Congestion Estimation
Input: density
Output: congestion classification

```
Procedure congestion_classification( f, b)
{

  If  d < threashold1  then
    Classifiy "least congestion"
  If  d >= threshold1 and d< threshold2 then
    Classifiy "modrate congestion"
  If d>= threshold2 then
    Classifiy "heavy congestion"
}
```

The estimation of the density at any location is performed periodically and updated to the database each time. The RFID reader detects the emergency vehicles continuously and updates the traffic control servers after detection of any emergency vehicles.

3.1 RFID-Based Emergency Vehicles Identification Module

The proposed RFID module comprises an antenna, RFID reader, and processing device. RFID reader read the tags of vehicles using an antenna. RFID antenna transmits the electromagnetic waves, and RFID tags receive these waves and transmit the tag information in the response [9]. RFID reader receives the tag information and passes the tag information to the processing device. In general, RFID tags contain Electronic Product Code (EPC) identifier [10, 11]. The EPC code is a unique identity of every tag. The basic EPC tag has four fields and the size of the tag is 12 bytes. The authorized database maintains the tags of the vehicle associated with their registration number. The vehicle can be identified using tag information of the vehicle. The proposed system reads the tag of the vehicle and processes them with the authorized database in order to identify the emergency vehicle.

4 Results and Analysis

In the proposed approach, the congestion detection approach is implemented on OpenCV libraries in Java and tested on the two pre-recorded videos. The first video is recorded when the traffic congestion is very low, and the second video is recorded for heavy traffic flow. To test the approach, each video is segmented into three different parts and each part is tested for estimation of the congestion. Based on the density value, approach classifies the traffics into least congestion or moderate congestion or heavy congestion. The two threshold for congestion classification is select t1 0.3 and t2 0.7. Based on these thresholds, the approach classifies the congestion (Fig. 2).

Tables 1 and 2 show the experiment results on both videos. The results reveal that approach is able to detect and classify the road traffic congestion.

The approach estimates the density of the congestion because traffic control server requires classification of the congestion to control the traffic.

Fig. 2 **a** Snapshot of the video 1 [12] and **b** snapshot of the video 2 [13]

Table 1 Estimated density of the all segments of the video 1

Video 1: Total frames 189, no. of pixels in background 244,520				
S. no.	Segments	Detected vehicles	Density	Classification
1	1–50 frames	4	0.082	Least congestion
2	63–113 frames	5	0.102	Least congestion
3	126–189 frames	3	0.061	Least congestion

Table 2 Estimated density of the all segments of the video 1

Video 2: Total frames 252, no. of pixels in background 267,550				
S. no.	Segments	Detected vehicles	Density	Classification
1	1–70 frames	32	0.83	Heavy congestion
2	85–155 frames	38	0.99	Heavy congestion
3	170–240 frames	35	0.91	Heavy congestion

4.1 Simulation of RFID-Based Emergency Vehicle Identification Module

The performance of the RFID reader during congestion scenario as well as its ability to handle the collision and the response time is measured for different scenarios.

Topology While simulating in NS2, a topology comprises one RFID reader agent. The performance of the RFID reader is evaluated against the different number of mobile nodes with the unique tag identifier. The numbers of the mobile tag nodes are 10, 20, 30, 40, 50, and 100. The speed of the mobile nodes is set to be random. Figure 3 shows the snapshot of the simulation where the collision may occur. Here, nodes 1, 2, 3, 4, and 5 lie within the range of RFID reader.

The objective of the simulation is to measure the efficiency of the system in order to avoid the effect of the collision and the computation of the average response time for a single tag. The simulation is carried out for 10, 20, 30, 40, 50, and 100 mobile nodes with a unique identifier. The efficiency is measured for three

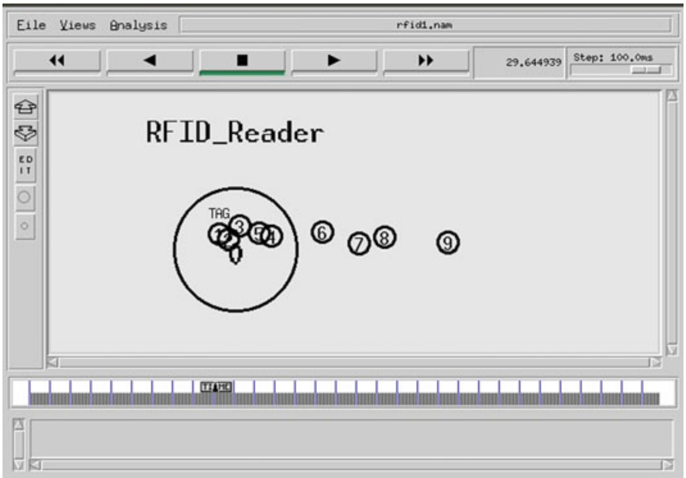


Fig. 3 Snapshot of the network animator (Nam), a view of the RFID simulation (collision scenario)

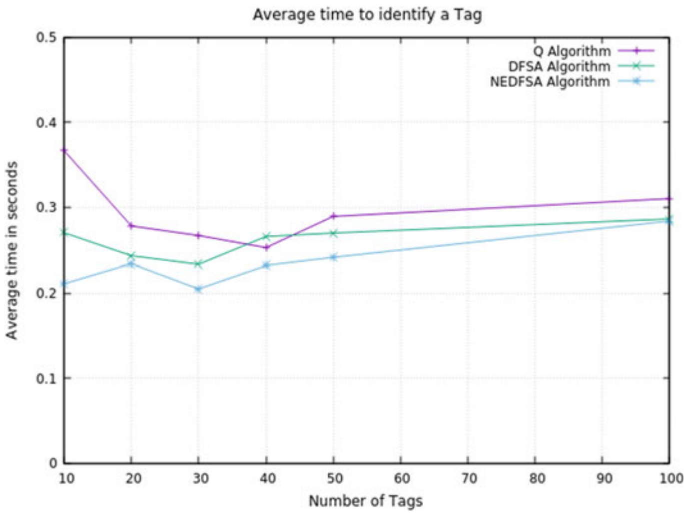


Fig. 4 Average response time of RFID module for different anti-collision algorithms

existing algorithms such as Q-Algorithm [14], Dynamic Framed Slotted ALOHA (DFSA) Algorithm [2], and Novel Estimated Dynamic Framed Slotted ALOHA (NEDFSA) Algorithm [1]. To compute the efficiency, the equation is given below in Eq. 2.

$$S_E = \frac{\sum Success\_slot\_count}{\sum frame\_Size} \tag{2}$$

Figure 4 shows the plot of efficiency and response time of the RFID module for different algorithms, respectively. In the simulation of the RFID module, all the tags are identified by RFID reader and the proposed simulation achieves 100% accuracy. The above figures establish the efficiency and the average response time of the system. The average response time to identify a tag is 0.29 s in Q-Algorithm, 0.26 s in DFSA, and 0.23 s in NEDFSA as shown in Table 3. The above results illustrate that RFID modules have average response time < 0.3 s. The timeout time based on the simulation is set to 0.5 s for RFID module. The anti-collision efficiency of the system is 0.2447 for Q-Algorithm, 0.2264 for DFSA and 0.2112 for NEDFSA. The average efficiency of the system is <1, means the system is capable to handle the collision. The simulation results reveal that proposed system has the quick response time as well as 100% accuracy in detection of the tags. The system is able to identify the tag of the emergency vehicle in case of occurrence of multiple tagged vehicles.

Table 3 Average response time of the system for different number of tags

S. no.	No. of tags	Average response time in seconds		
		Q-Algorithm	DFSA	NEDFSA
1	10	0.3667	0.2701	0.2107
2	20	0.2780	0.2432	0.2342
3	30	0.2668	0.2332	0.2047
4	40	0.2527	0.2657	0.2319
5	50	0.2892	0.2698	0.2413
6	100	0.3102	0.2863	0.2839

4.2 Comparison

Google is the most widely used for detection of the real-time traffic conditions in the public domain. It estimates the traffic conditions by retrieving the GPS coordinates of the various users in that location and by estimation of the speed flow. The approach requires Internet connectivity and is completely based on roadmap developed by Google itself. In developing countries such as India, Internet penetration is very low and Internet services are not reliable. People cannot depend on the services operated over the Internet. The proposed approach obtains the traffic information using a camera and RFID reader that should be deployed on roads. The approach does not require Internet connectivity every time. The approach proposed in [6, 7] is accurate in detection of the road traffic congestion but does not react to it. This proposed approach notifies the traffic control server in case of any emergency vehicle is detected. The proposed approach is more suitable for creating green corridors.

4.3 Limitations of Proposed Work

The implementation of the proposed approach detects the congestion by using image processing techniques. The image processing techniques highly depend on the environmental conditions [15].

5 Conclusions and Future Work

This paper extends the computer vision-based road traffic congestion detection to one level up in terms of response to congestion by integration of RFID-based technology. In this approach, RFID module identifies the emergency vehicle based on the RFID tags. The approach detects the real-time congestion and notifies the traffic control server. The classification of the congestion based on its density is also carried out by using two-threshold policies. The RFID readers have the quick response time to identify a tag. In the simulation, RFID reader achieves 100% accuracy in tag identification. The average response time to identify a tag is less than 0.3 s. The anti-collision efficiency of the system is also less than 1. The simulation results reveal that RFID module is quick and accurate in its job.

In future, our aim is to develop the smart traffic control system. The proposed system could be used for the smart traffic control system that is able to provide on-demand green corridor for the emergency vehicles. The system will utilize the congestion information generated by the proposed system and thereby identifying the least congested routes in real time.

References

1. Kumar, T., Kushwaha, D.S.: Traffic surveillance and speed limit violation detection system. J. Intell. Fuzzy Syst. **32**, 3761–3773 (2017). https://doi.org/10.3233/JIFS-169308
2. Kumar, T., Gupta, S., Kushwaha, D.S.: An efficient approach for automatic number plate recognition for low resolution images. In: The Fifth International Conference on Network, Communication and Computing (ICNCC) 2016, Kyoto, Japan, pp. 53–57 (2016). https://doi.org/10.1145/3033288.3033332
3. Zhang, E., Kuang, Y., Jiang, W., Umer, M.A.: Active RFID positioning of vehicles in road traffic. In: 11th International Symposium on Communications and Information Technologies (ISCIT) 2011, Hangzhou, China, pp. 222–227 (2011). https://doi.org/10.1109/iscit.2011.6089737
4. Terroso-Sáenz, F., Valdés-Vela, M., Sotomayor-Martinez, C., Toledo-Moreo, R., Gómez-Skarmeta, A.F.: A cooperative approach to traffic congestion detection with complex event processing and VANET. IEEE Trans. Intell. Transp. Syst. **13**, 914–929 (2012). https://doi.org/10.1109/tits.2012.2186127
5. Palubinskas, G., Kurz, F., Reinartz, P.: Detection of traffic congestion in optical remote sensing imagery. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2008, Massachusetts, USA, II, pp. 426–429 (2008). https://doi.org/10.1109/igarss.2008.4779019
6. Li, L., Chen, L., Huang, X., Huang, J.: A traffic congestion estimation approach from video using time-spatial imagery. In: First International Conference on Intelligent Networks and Intelligent Systems (ICINIS'08) 2008, Wuhan, China, pp. 465–469 (2008). https://doi.org/10.1109/icinis.2008.182
7. Idé, T., Katsuki, T., Morimura, T., Morris, R.: City-wide traffic flow estimation from a limited number of low-quality cameras. IEEE Trans. Intell. Transp. Syst. **18**, 950–959 (2017). https://doi.org/10.1109/TITS.2016.2597160
8. D'Andrea, E., Marcelloni, F.: Detection of traffic congestion and incidents from GPS trace analysis. Expert Syst. Appl. **73**, 43–56 (2017)
9. Bang, O., Kim, S., Lee, H.: Identification of RFID tags in dynamic framed slotted ALOHA. In: 11th International Conference on Advanced Communication Technology (ICACT) 2009, Gangwon-Do, Korea (South), pp. 354–357 (2009)
10. Cha, J.-R., Kim, J.-H.: Dynamic framed slotted ALOHA algorithms using fast tag estimation method for RFID system. In: Proceedings of IEEE Consumer Communications and Networking Conference 2006, Las Vegas, USA, pp. 768–772 (2006)
11. Kumar, T., Gupta, S., Kushwaha, D.S.: A smart cost effective public transportation system: an ingenious location tracking of public transit vehicles. In: 5th International Symposium on Computational and Business Intelligence (ISCBI) 2017, Dubai, UAE, pp. 134–138 (2017)
12. https://www.youtube.com/watch?v=Z3FYLP5TWCw

13. https://www.youtube.com/watch?v=O84FlZnP0qs
14. Maguire, Y., Pappu, R.: An optimal Q-algorithm for the ISO 18000-6C RFID protocol. IEEE Trans. Autom. Sci. Eng. **6**, 16–24 (2009). https://doi.org/10.1109/TASE.2008.2007266
15. Kumar, T., Kushwaha, D.S.: An efficient approach for detection and speed estimation of moving vehicles. Procedia Comput. Sci. **89**, 726–731 (2016). https://doi.org/10.1016/j.procs.2016.06.045

An Analytical Implementation of CART Using RStudio for Churn Prediction

Vani Kapoor Nijhawan, Mamta Madan and Meenu Dave

Abstract Data mining is a technique for finding new and undiscovered patterns, which help in predicting the future trends. Nowadays, it is being applied in all the fields, may it be, the field of medicines or credit cards or banking and insurance or telecommunications. Decision tree is a simple and popular technique of data mining (commonly employed for predictive analysis) which can be used to forecast the future trends. There are several algorithms for decision tree generation like ID3, C4.5, CART which can be applied with the help of different software tools like WEKA, Rapid Miner, R. This paper focuses on applying data mining in the field of telecommunications, to predict the churning behavior of the customers.

Keywords Data mining · Customer churn · CART algorithm · R

1 Introduction

Customer churn in telecom sector is one of the major areas of concern because it always involves a lot of overhead to acquire the new customers than to retain the existing customers of any organization. This work is about application of data mining techniques for telecom customers to predict their churning trends. There are

V. K. Nijhawan (✉) · M. Madan
Information Technology, Vivekananda Institute of Professional Studies,
Pitam Pura, Delhi, India
e-mail: vanikapoor28@gmail.com

M. Madan
e-mail: mamta.vips@gmail.com

V. K. Nijhawan · M. Madan
Information Technology, Guru Gobind Singh Indraprastha University, Dwarka, Delhi, India

M. Dave
Faculty of Engineering and technology, Jagan Nath University, Jaipur, Rajasthan, India
e-mail: meenu.s.dave@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_11

several algorithms for decision tree generation like ID3, C4.5, CART which can be applied with the help of different software tools like WEKA, Rapid Miner, R.

In this paper, the authors explored the CART algorithm by making use of a software tool R for a database of 300 records collected online through a survey Web site "www.surveymonkey.com." Through this survey, records were gathered, from a user base, belonging to different age groups, residing in different states, using one or more mobile phone connections of prepaid or postpaid mode and from different service providers.

1.1 Decision Tree

Decision tree is the technique used in data mining to streamline complicated strategic challenges. This method is a flowchart representing a tree in top-down fashion, starting from root node and moving toward the terminal nodes. Here, the internal nodes represent a test or decision and the branches represent the outcomes. It is a popular technique because of its simplicity.

Decision tree technique can be implemented by using any of the mentioned algorithms like ID3, C4.5, CART, CHAID. In this paper, the focus is on applying CART, i.e., Classification and Regression Trees' algorithm.

1.2 Software Tool Used: R

R is an open-source statistics and graphics package which is a command-line-driven software. It consists of hundreds of additional packages available for free, which are very useful for providing help in data mining, machine learning, and statistics [7].

RStudio is an IDE for using R. It consists of a console, syntax-highlighting editor for direct code execution, also tools for graph plotting, history, and debugging.

1.3 Why Choose R

KDnuggets, a leading site on Business Analytics, Big Data, Data Mining, Data Science, and Machine Learning and it is edited by Gregory Piatetsky-Shapiro and Matthew Mayo [8], conducted its 16th annual software poll in 2015, where about 2800 data miners and data scientists, voted for R as the most popular tool, out of 93 different tools (which is near about 46.9% share). Also, in the year 2016, R continues to be the top most used tool with a usage of 49% [8].

1.4 Rattle Package

Rattle stands for R Analytic Tool To Learn Easily (Rattle). It provides an interface to R functionality for data mining which is a simple and intuitive interface that helps the user to load data directly from a CSV file (or using ODBC), transform (if required) and explore the data, apply different techniques, and build models [9].

This paper makes use of RStudio and Rattle, as a tool for mining the data using CART algorithm.

1.5 CART Algorithm

CART is a decision tree algorithm, developed by Breiman et al. (1984), that constructs tree by splitting a node into two child nodes repeatedly, starting from the root node, that contains the complete learning sample. It applies a greedy approach to do the splits in the tree. A created tree can be used for predictions, by traversing the binary tree given a new input record [6].

Advantages

- CART can handle numerical or categorical variables easily.
- CART algorithm identifies the most significant variables and eliminates the insignificant ones.
- CART can handle outliers [6].

2 Related Work

CART Sonia et al. [6] have done a comparison of three popular algorithms of data mining, namely ID3 C4.5 and CART, on the basis of their advantaged and disadvantages and splitting criteria used and conclude that different algorithms can prove to be better than other two depending upon the types of datasets used.

Kaur and Mahajan in [2] say that if the data which is already available with the telecom companies is analyzed carefully, it can throw some light on churning patterns of the customers. This information can be used for the current and upcoming customers to design the retain policies.

In Ref. [5], authors have studied and applied CART using WEKA software to classify blood donors and they have also stated that CART-derived model along with the extended definition for identifying regular voluntary donors has provided a very good classification accuracy-based model.

Lakshmi et al. [3] have done a survey on different algorithms of decision tree like ID3, C4.5, CART, and CHAID. They describe and compare the performance of the above-mentioned algorithm.

Authors in Ref. [4] have developed and proposed a new algorithm, called CART for handling the data stream, i.e., dsCART. It has been shown mathematically by using the properties of the normal distribution and Taylor's Theorem that this new generated algorithm is accurate and fast and therefore appropriate for solving the problem of data stream classification.

Gordon in [1] makes use of CART and demonstrates their usefulness for data analysis mainly in the field of public health. Authors here have presented two illustrations for CART. For the first case, classification tree is used for categorical outcome, applied on the publically available dataset survival of lung cancer. In the second case, regression tree is used for continuous outcome on a dataset to assess the prevalence of injury and respiratory diseases in an agricultural population.

3 Methodology

This section describes the methodology used in the research process for the collection and preprocessing of the collected data.

3.1 Data Collection and Description

The data for the mining purpose was collected by designing an online questionnaire with the help of a survey Web site "www.surveymonkey.com." Through this survey, around 300 records were collected, from a user base, belonging to different age groups, residing in different states, using one or more connections of prepaid or postpaid mode and subscribing services from different service providers.

3.2 Data Preprocessing

This step involves cleaning the data by removing missing values and filtering the data so that it can be in a format accepted by R.

In this step, the attribute values were converted into the numeric categorical form and the age field in the dataset was converted from continuous to categorical.

4 Results and Implementation

After data cleaning and preprocessing step, the collected data file was given as input to RStudio, and then using the Rattle package, a set of some simple steps were required to finally generate the decision tree to predict the churning behavior of the upcoming customers.

4.1 Dataset Used

The attributes in the dataset used are the following (Table 1).

4.2 Description of Dataset Used

Attribute1: Age of the user (Table 2)
Attribute2: Gender of the user
Attribute3: State to which user belongs
Attribute4: No. of connections
Attribute5: Service provider
Attribute6: Type of payment plan
Attribute7: Duration of the connection
Attribute8: Type/Types of services being used
Attribute9: Whether the user is churning or not.

Table 1 List of attributes in the used dataset

S. No.	Attribute name
1	Age
2	Gender
3	State
4	No. of connections
5	Provider
6	Payment plan
7	Duration (how long the connection is being used)
8	Type of service used
9	Churn

4.3 Steps to Be Followed in Rattle

Step 1: Open RStudio and write the command library (Rattle) and press enter. After this, write *rattle*() to start the Rattle package.
Step2: Click on the data tab in rattle.
Step 3: Select file in source tab and browse the file.
Step 4: Now click on the execute tab on the top left corner, to see the file contents to verify the attribute types. Also, to verify that churn should be selected as the target attribute.
Step 5: Next step is to click on model tab and select tree option and click on execute once again to see all the decision rules for the tree.
Step 6: Last step is to click on draw button toward the middle right, and view the generated decision tree in RStudio.

4.4 Decision Tree Rules for Customer Churn (Using CART)

Summary of the Decision Tree model for Classification (built using 'rpart').

4.5 Decision Tree for Churn (Using CART): Generated Tree

The data file consisting of the attributes given in Tables 3, 4, 5, 6, 7, 8, 9, 10, and 11 was used as an input file, in RStudio. This led to the summary generation of the corresponding tree, which is shown in Table 12. After the summary generation, just a click on the DRAW tab in rattle leads to the designing of decision tree. Following is the tree generated by R for the given dataset (Fig. 1).

Table 2 Description of dataset attribute

```
> str (churn.data)
'data.frame':300 obs. of  9 variables:
$ Age               : int  1 1 0 2 1 0 1 1 1 1 ...
$ Gender            : int  1 0 0 1 1 1 1 1 1 0 ...
$ State             : int  1 1 0 1 1 1 1 0 2 3 ...
$ No.of.connections: int  1 1 1 0 1 0 0 0 1 1 ...
$ Provider          : int  0 3 3 1 2 0 1 1 2 1 ...
$ Payment.plan      : int  2 0 2 0 0 1 0 0 0 2 ...
$ Duration          : int  3 3 2 1 3 0 1 3 2 2 ...
$ Type.of.services : int  2 2 2 2 2 2 2 2 2 2 ...
$ churning          : int  0 0 1 0 0 0 1 0 1 1 ...
```

Table 3 Set of values of age attribute

Value in dataset	Notation used
17–20 years	0
21–30 years	1
31 years onwards	2

Table 4 Set of values of gender attribute

Value in dataset	Notation used
Male	0
Female	1

Table 5 Set of values of state attribute

Value in dataset	Notation used
Delhi	0
Haryana	1
Odisha	2
Karnataka	3
Uttar Pradesh	4
Uttarakhand	5
Tamil Nadu	6
Rajasthan	7

Table 6 Set of values of no. of connections' attribute

Value in dataset	Notation used
One	0
Two	1
Three or more	2

Table 7 Set of values of provider attribute

Value in dataset	Notation used
Vodafone	0
Airtel	1
Others	2
Airtel and Vodafone	3
Airtel and others	4
Airtel, Vodafone, and others	5
Vodafone and others	6

Table 8 Set of values of payment plan attribute

Value in dataset	Notation used
Prepaid	0
Postpaid	1
Both	2

Table 9 Set of values of duration attribute

Value in dataset	Notation used
Less than 6 months	0
6 months to 1 year	1
1–2 years	2
More than 2 years	3

Table 10 Set of values of type of services' attribute

Value in dataset	Notation used
Calling and Messaging	0
Data	1
Both	2

Table 11 Set of values of churning attribute

Value in dataset	Notation used
Not churn	0
Churn	1

In the above tree, value 0 represents the non-churn and 1 represents the customers who are churners. On the basis of above-generated tree, we can predict the churning behavior of any particular customer by looking at the other details of the person and consequently following the corresponding branch of the tree.

4.6 Cumulative Distribution of Duration Attribute

The following graph shows the cumulative distribution of one of the important attribute from the data file, i.e., duration (time duration of the connection being used by the customer in months or years). From this graph, the conclusion can be drawn

Table 12 Summary of decision tree model

```
n= 210

node), split, n, loss, yval, (yprob)
      * denotes terminal node

 1) root 210 91 0 (0.5666667 0.4333333)
   2) Duration>=2.5 109 32 0 (0.7064220 0.2935780) *
   3) Duration< 2.5 101 42 1 (0.4158416 0.5841584)
     6) Age< 0.5 63 31 0 (0.5079365 0.4920635)
      12) Provider< 0.5 16  4 0 (0.7500000 0.2500000) *
      13) Provider>=0.5 47 20 1 (0.4255319 0.5744681)
        26) Payment.plan>=0.5 14  5 0 (0.6428571 0.3571429) *
        27) Payment.plan< 0.5 33 11 1 (0.3333333 0.6666667) *
     7) Age>=0.5 38 10 1 (0.2631579 0.7368421)
      14) State>=0.5 10  4 0 (0.6000000 0.4000000) *
      15) State< 0.5 28  4 1 (0.1428571 0.8571429) *
Classification tree:
rpart(formula = churning ~ ., data = crs$dataset[crs$train, c(crs$input,
    crs$target)], method = "class", parms = list(split = "information"),
    control = rpart.control(usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] Age          Duration     Payment.plan Provider     State

Root node error: 91/210 = 0.43333

n= 210

        CP nsplit rel error  xerror     xstd
1 0.186813      0   1.00000 1.00000 0.078912
2 0.043956      1   0.81319 0.81319 0.076074
3 0.021978      4   0.68132 0.78022 0.075333
4 0.010000      5   0.65934 0.74725 0.074516

Time taken: 0.09 secs

Rattle timestamp: 2017-05-27 16:33:07 hp
======================================================================
```

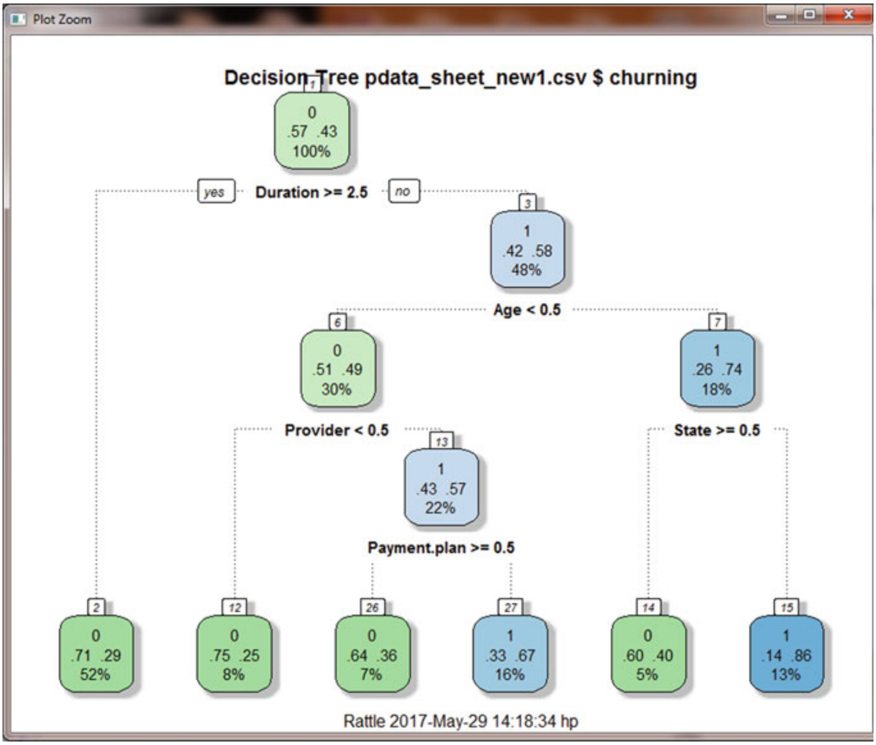


Fig. 1 Decision tree generated by Rattle

that, if the time duration is longer, i.e., if the customer is using the same connection for more than 2 years, he can be labeled as a loyal customer and has less probability to churn, in contrast with the new customers (Fig. 2).

4.7 Accuracy and Error Rate

For the above model, the confusion matrix for the test data is as follows (Fig. 3).

So, according to the above results, the error rate comes out to be 0.25% approx., and hence, the accuracy comes out to be 0.75%.

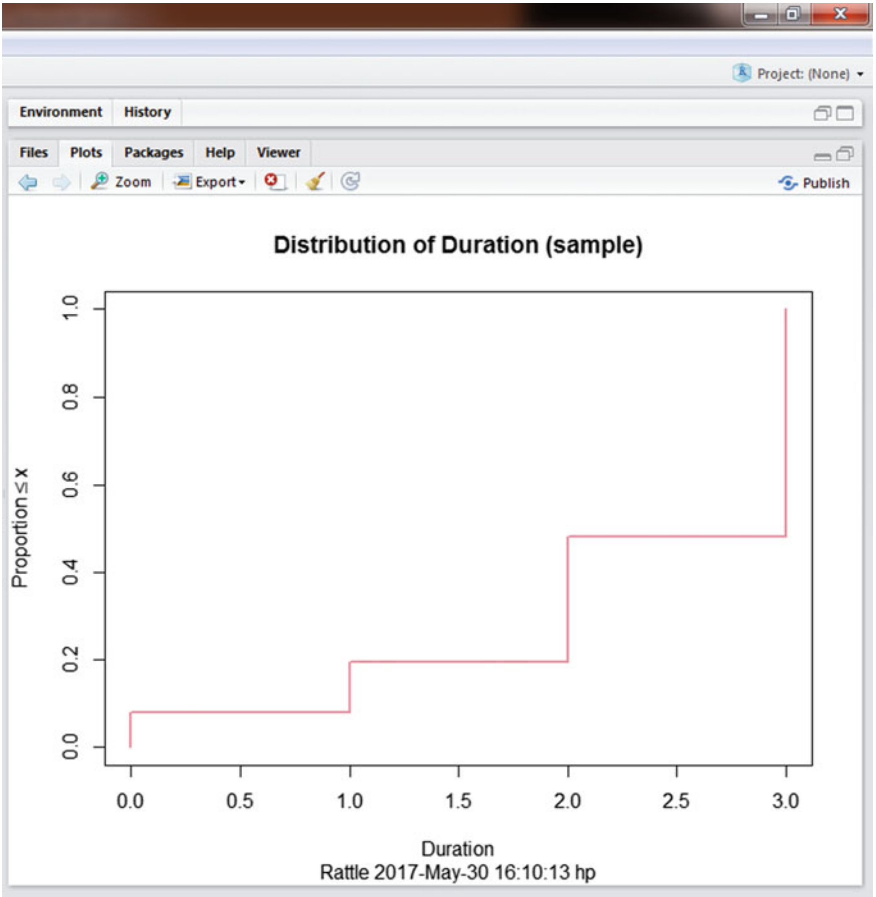


Fig. 2 Cumulative distribution of duration

Fig. 3 Error matrix for the decision tree model on pdata_sheet_new1.csv [test] (counts)

		Predicted 0	Predicted 1
Actual	0	23	4
Actual	1	7	11

5 Conclusion

In this paper, the dataset was prepared by collecting the records from around 300 mobile phone users and the same was used as an input file in RStudio, for generating a decision tree using CART algorithm with the help of Rattle package.

The generated tree reflects the churning behavior of the mobile phone users and hence can be used to predict the churning behavior of the new records of the customers as well.

From the above tree, it is clearly visible that, if the duration of using a particular connection is more than 2 years, i.e., if the person is using the same number for more than 2 years, he is less likely to change it or churn, as compared to the new customers.

6 Recommendations and Future Scope

These results are nothing but the recommendations for the telecom companies to design their corresponding marketing policies, so that they are more focused on the "not so old" customers of the company, which have a more probability to churn and move to the competitors.

This work can be extended in future, and these results can be exploited by the service providers to pay more attention toward the "not so old" customers and give them more discounts and other lucrative benefits. It can help them to stay with the same provider for a longer duration and finally come into the category of loyal customers, reducing the churning rate of the provider or company.

References

1. Gordon, L.: Using classification and regression trees (CART) in SAS® enterprise miner for applications in public health. In: SAS Global Forum 2013, Data Mining and Text Analytics, Paper 089-2013
2. Kaur, M., Mahajan, P.: Churn prediction in telecom industry using R. Int. J. Eng. Tech. Res. (IJETR) **3**(5), May 2015
3. Lakshmi, B.N., Indumathi, T. S., Ravi, N.: An empherical study on decision tree classification algorithms. Int. J. Sci. Eng. Technol. Res. (IJSETR) **4**(11), November 2015
4. Rutkowski, L., Jaworski, M., Pietruczuk, L., Duda, P.: The CART decision tree for mining data streams. J. Inf. Sci. **266**(2014), 1–15 (Elseviers)
5. Santhanam, T., Sundaram, S.: Application of CART algorithm in blood donors classification. J. Comput. Sci. **6**(5), 548–552 (2010). ISSN 1549-3636, © 2010 Science Publications
6. Singh, S., Giri, M.: Comparative study Id3, cart and C4.5 decision tree algorithm: a survey. Int. J. Adv. Inf. Sci. Technol. IJAIST **3**(7). ISSN 2319-2682

Web References

7. http://www.the-data-mine.com/Software/R
8. http://www.kdnuggets.com/2015/05/poll-r-rapidminer-python-big-data-spark.html
9. https://cran.r-project.org/web/packages/rattle/index.html

Traffic Control and Management Over IoT for Clearance of Emergency Vehicle in Smart Cities

Biru Rajak and Dharmender Singh Kushwaha

Abstract With technology permeating every aspect of human life, its adoption for traffic management has now become the norm in most of the developed countries. Many developing countries are now working toward adopting smart and intelligent ways to manage the ever increasing road traffic. The smart traffic management system is envisaged to manage the traffic from a centralized management system. This centralized traffic control center controls the traffic monitoring in real-time traffic conditions and can be used efficiently in the case of a medical emergency. In a medical emergency, the centralized control center may suspend the routine movement, to create a Green Corridor for the ambulance to pass without any delay at traffic junctions. This paper proposes a framework for the centralized traffic control system that is capable of creating a "Green Corridor" for an ambulance in case of a medical emergency. The system uses Internet of Things' (IoT) techniques like WiMAX for the centralized control center for all traffic signals and RFID to identify the vehicle. Our contribution is a framework for creating a Green Corridor route map generation for the ambulance and resuming of the routine traffic after allowing the ambulance to cross the traffic signal.

Keywords IoT · Green Corridor · WiMAX · RFID

B. Rajak (✉) · D. S. Kushwaha
Department of Computer Science and Engineering, Motilal Nehru National Institute of Technology Allahabad, Allahabad 211004, UP, India
e-mail: birurajak@gmail.com

D. S. Kushwaha
e-mail: dsk@mnnit.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_12

1 Introduction

According to a recent proposal, Government of India is planning to develop 100 smart cities [1], with a goal to provide improved quality of life through highly equipped infrastructure. The smart city is a vision to integrate different amenities like smart school, digital library, smart water management, smart waste management, power plant, energy management in a smart and intelligent way to assure an improved quality of life. A smart traffic control and management system is the basic requirement for a smart city. Smart traffic control and management system has the capability to handle the traffic congestion in real time with the help of IoT. A recent survey [2] reveals that every year nearly 18,000 patients who were being transported in an ambulance could have been saved by timely medical help if the ambulance had reached the hospital on time. A prompt medical assistance is a key to saving the life in the case of an emergency. The chance of survival of a patient, in case of a heart attack reduces by 10% every single minute. Traffic control management can be achieved by various techniques. In recent years, wireless communication network has been the most widely used technology for traffic control. Radio-frequency identification (RFID) [3] is a cost-effective technology that can be used for controlling the traffic. RFID technology involves the conversion of electromagnetic signals to radio frequency in order to uniquely identify an object. The emergency vehicle which is fitted with RFID device can be easily detected using this technology. Worldwide Interoperability for Microwave Access (WiMAX) [4, 5] provides an efficient packet radio interface which enables high data transfer rate over the network. WiMAX is one of the most sought-after wireless technologies for centralized traffic control. It represents the interoperability of the devices according to the specifications of IEEE 802.16 or compatible standard.

The rest of the paper is organized into five sections. Section 2 summarized the literature survey. In Sect. 3 proposed model is discussed. Section 4 provides the framework of the proposed model. Section 5 shows the experimental result, and Sect. 6 concludes the paper.

2 Literature Survey

Traffic surveillance and management are the major problems in developing countries. Congestion on the road eventually slows down the moving traffic. Mittal et al. [6] have discussed the use of Green wave technique for clearance of emergency vehicles. “Green wave” is synchronization of the traffic light from red to green for the entire path. The biggest disadvantage of this green wave is that it turns all the red lights into green throughout the entire route. Each traffic light on the entire route is given equal weight. This leads to abnormally high stoppages at the signal that are far from the start. Any disturbance in synchronization can lead to exacerbated traffic situations and increases undue delay for other vehicles. Sharma et al. [7] uses RFID to address the issues related to traditional traffic control system, mainly related to beam interruption and image processing. In this work, the authors use RFID to deal with multiple vehicles, multi-lane and multi-road junction areas. This scheme leads to dynamic scheduling and efficient time management. The authors do not discuss the communication between the traffic signal controller and emergency vehicle. Sunder et al. [8] proposed a traffic control system in which all vehicles are equipped with ZigBee to clear the traffic network. ZigBee transmitter is fitted in the emergency vehicle. On pressing the switch of the transmitter, the signal with unique id and security code is transmitted to the receiver. The receiver compares the security code received to the security code present in its database. If it matches, then the green light is turned on. It matches the security code with database to implement security. This can be implemented by using unique id of ZigBee which is sufficient to uniquely detect a vehicle. The limitation of this work is that ZigBee CC2500 is used for transmission that has a range of 20 m. At traffic signals, an ambulance can be identified from a maximum distance of 20 m. In real-life scenario, the average speed of an ambulance is 60 Kmph. The ambulance will take 1.2 s to cover a distance of 20 m. This is not a viable solution in real life as the ambulance has to stop at the traffic signal. The authors do not discuss the situation in case of traffic jam. In real scenario, a traffic jam can be up to 50 m. long. Therefore in this case, an ambulance which is past 20 m from the traffic signal will not be identified. Sundar et al. [8] consider only a hypothetical situation, making it unsuitable for handling real-life traffic scenario. Thus in real life, all above discussed techniques are not viable. Many solutions have been proposed in literature to solve the traffic management in smart and intelligent ways, but they fail to handle real-time situations. Hence, this paper proposes new technique “Green Corridor” to solve the emergency vehicle clearance in real time.

3 Proposed Work

The models discussed above are not efficient enough to handle the problems related to traffic congestion and emergency vehicle clearance.

Algorithm: Green Corridor setup

```
Input: P[N]; //N Total number of traffic posts;
1:Ambulance requests for route to Centralized Server;
2:Centralized server registers and processes the request;
3:if(Centralized server Grant the Route==true)then
4:Green Corridor Setup:
5:    i ← 0;
6:    while(i<N)  //   Repeat until reached hospital
7:    Centralized server grants the request and provides
      the route to hospital;
8:    Centralized server sends a unique id of the
      Ambulanceto all traffic posts;
9:    Ambulance sendsits current location to nearest
      post(P[i]);
10:   Ambulance starts navigating towards the hospital;
11:   P[i] sets its signal to green ;
12:   Ambulance passes through P[i];
13:   RFID reader at Pi senses the ambulance ;
14:     if(Unique Id of ambulance == Id provided
          byCentralized server ) then
15:       Resume the suspended traffic;
16:       Send timestamp to P[i+1];
17i       i++;
18:     end if;
19:   end while;
20:  end if;
21:  else repeat steps 1 &2;
```

To solve this problem, a novel technique termed as “Green Corridor” is proposed, to handle ambulance clearance especially in case of traffic jam. It consists of mainly three components. The first component is the centralized controller server that monitors all the traffic posts and initially request for a route from the ambulance goes to the centralized controller. Centralized controller grants the request to the ambulance and generates a route to the traffic post. The second component consists of RFID reader, which receives the signal from RFID tag fitted in the ambulance. When the ambulance crosses the traffic signal, the suspended traffic is resumed. The third component consists of RFID device that is installed at each traffic post. When the central controller sends the request to next traffic post, the signal is turned green and remains so till RFID attached in ambulance passes through the post.

4 Framework of the Proposed Work

The proposed model consists of three modules which enable the system to clear the emergency vehicle. The functionalities of each of the three modules are described in the following section.

4.1 Central Controlling Server

The centralized controlling server communicates with WiMAX using IEEE 802.16 wireless network standards that are interoperable, as against the IEEE 802.11 standards that are used by wireless LANs. WiMAX network is so designed that it can be deployed in licensed and unlicensed spectra and can support both IPv4 and IPv6 network in the client–server application. The 802.16 Radio Access Network (RAN) [9] being architecture independent enables WiMAX to support seamless integration and interworking with WiFi, 3rd Generation Partnership Project 2 (3GPP) [10] networks through a set of well-defined interfaces. Given the requirements of the proposed model, we find WiMAX to be the best-suited technology. A WiMAX system consists of two major components: (a) a WiMAX base station and (b) a WiMAX receiver. The centralized server uses WiMAX as the base station which comprises an indoor electronics equipment and a WiMAX tower similar to the mobile tower. WiMAX base station provides a very large coverage area. The maximum radius of the cell is up to 30 miles (50 km). Using the WiMAX technology, the emergency vehicle connects to the centralized server, i.e., the base station through either a separate antenna of stand-alone box installed in the vehicle. The ambulance connects to the centralized server in case of an emergency, and the centralized server provides the route as well as alerts all the traffic posts lying in the generated route as shown in Fig. 1. In the proposed model, each traffic post has prior knowledge of the arrival of the ambulance. In the solution proposed by Sunder et al. [8], the system has no prior knowledge about the arrival of an emergency vehicle at traffic post, resulting in the ceasing of the emergency vehicle at a red signal.

4.2 RFID in Ambulance

The proposed model requires an antenna and an RFID tag on every ambulance. The centralized server provides the route when the emergency vehicle invokes the service "*RouterequestForRoute* (Struct RFID uniqueRFID, String hospitalId, Loc curLoc)." A service call has three input parameters: (a) uniqueRFID by which ambulance can be uniquely identified, (b) hospitalId, which uniquely identifies the destination hospital, and (c) *curLoc* defines the current location of the ambulance. In response, the cental server grants request to the ambulance by "*routeGrantedForAmbulance* (Route hospitalRoute)." The central server uses the service "*notifyAllTrafficPost* (RFID uniqueRFID, Loc curLoc)" to notify all the traffic posts about the arrival of the ambulance, as shown in Figs. 1 and 2.

On receiving the route from the central server, ambulance starts navigating through the provided route. Sunder et al. [8] have 1.2 s which is very less time to set the signal green, but in our proposed system, traffic posts have prior knowledge of ambulance through central server. The signal is set to green to the junction and

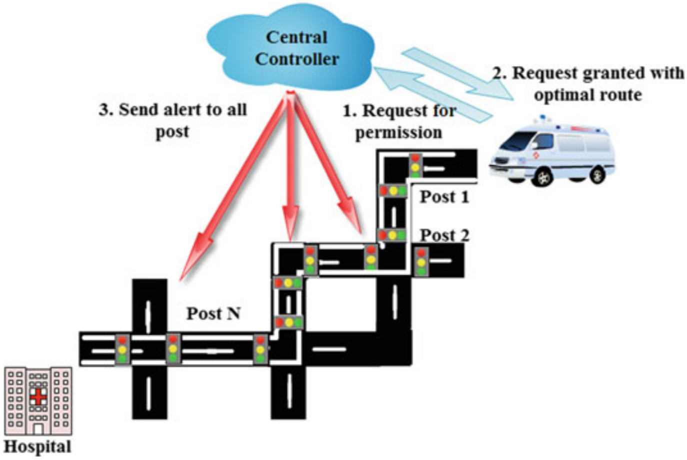


Fig. 1 Request granted to ambulance

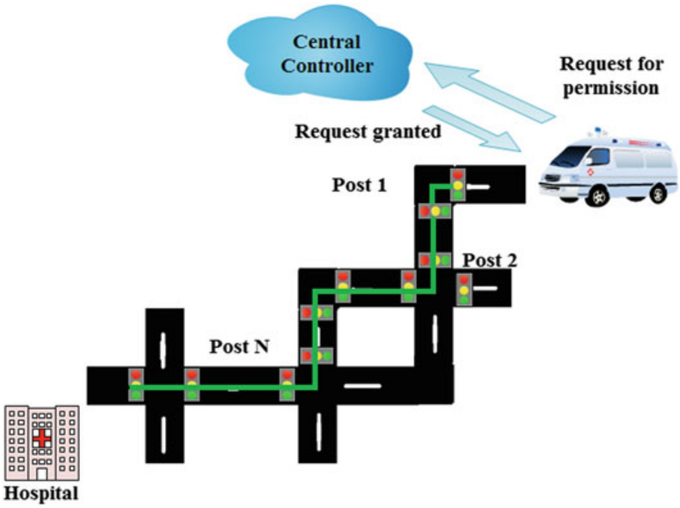


Fig. 2 Green Corridor setup

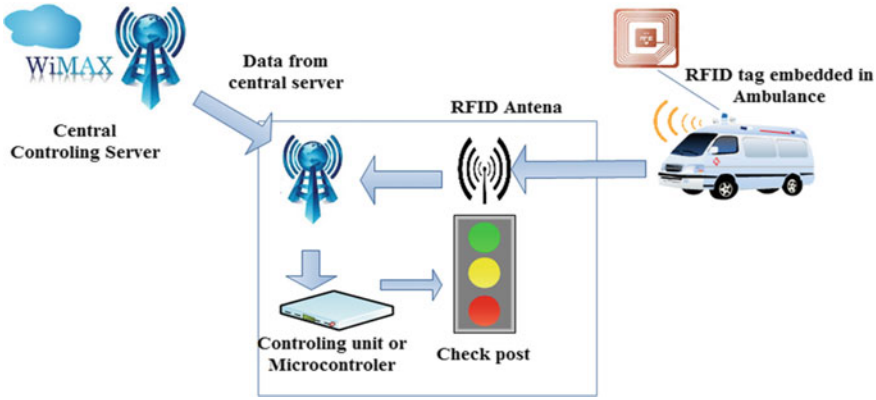


Fig. 3 Working of RFID at traffic signal

when ambulance crosses through the junction, the RFID receiver detects the RFID tag attached in Ambulance subsequently normalize the signal and notify the subsequent junction about the ambulance with a timestamp. On receiving the signal from the previous junction, next junction turns the green signal on. This process of relaying the signal from a traffic post to next is repeated until the ambulance reaches its destination (Fig. 3).

4.3 RFID Receiver at Traffic Junction

After receiving the alert from the central server, the first traffic junction in the generated route turns the signal green. When the ambulance approaches the traffic signal, the antenna which is used as RFID reader senses the RFID tag of the ambulance. RFID reader at each traffic post matches the ID of the ambulance with that provided by the central server.

If RFID of the passing vehicle matches with the previously provided RFID, then normal traffic is resumed as shown in Fig. 4. RFID at traffic post always listening to the event, i.e., always sensing the RFID attached with ambulance. To implement this, an event listener method "*RFIDRederAtPost* (RFID uniqueRFID)" called to sense the RFID of ambulance.

5 Experimental Result

A low-frequency (125 kHz) passive RFID tag and RFID reader is used in our experiment. Arduino Uno (ATmega [11]) is used as a microcontroller. In realistic scenario, RFID reader is installed at every traffic post and RFID tag is attached to

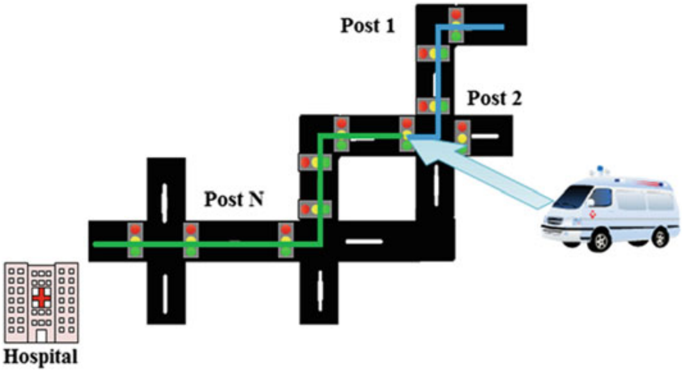


Fig. 4 Normalizing signal after ambulance passed

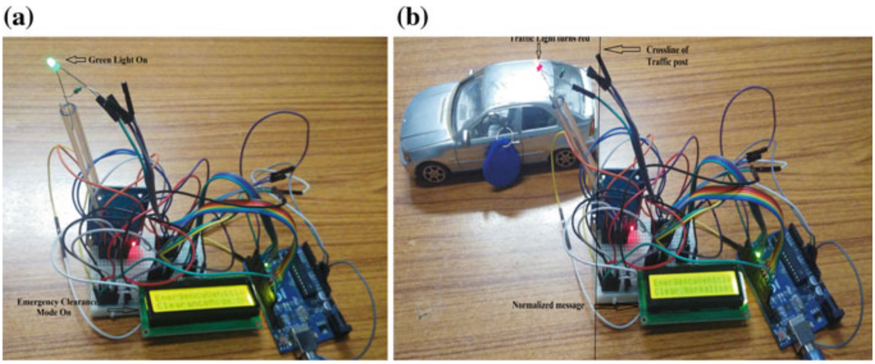


Fig. 5 **a** Emergency clearance mode on. **b** Normalizing the traffic signal after crossing ambulance

every emergency vehicle. The first traffic post turns its signal green after getting the alert from the central server. On the receipt of the signal, the traffic post activates emergency vehicle clearance mode as shown in Fig. 5a. Once the ambulance is sensed at traffic post through the attached RFID, an alert message with the current timestamp is sent to next traffic post and signal became normal. Now, the next traffic post behaves in a similar manner.

6 Conclusion

Smart and intelligent management of traffic is the pressing need for highly populated countries and more so with the rapid need and proposal for development of smart cities. A jam at traffic junctions can cost life in case of an emergency. To tackle such situations, this paper proposes a framework for traffic management for clearance of emergency vehicle by using "Green Corridor" over the IoT. The benefit of using WiMAX is its wide coverage area of approximately 50 km and an operating frequency of 2–11 GHz. The centralized control over the traffic enables sending a prior alert to all the traffic post along the route through which the ambulance is granted clearance. This certainly gives a better performance as compared to the solution proposed by Sunder et al. [8], and also time available to stop all other traffic so as to provide Green Corridor for ambulance is 1.2 s, which certainly is very less and does not guarantee smooth movement of the ambulance; on the contrary, our proposed approach is not plagued by this. No other vehicle shall be stranded for more than 2 min if the distance between two consecutive check-post is 2 km and average speed of ambulance is 60 Kmph. Use of high-range RFID tag and receiver makes possible the detection of the emergency vehicle from a longer distance. The proposed method is also highly efficient and a robust approach as compared to the traditional traffic management which uses GPS for tracking a vehicle or the technique based on image processing [12, 13]. This model can thus help provide better emergency services.

References

1. Smart Cities Mission, Government of India. http://smartcities.gov.in/
2. Lukas Postulka, H.: 'Give way to ambulance' educates people that saving time is saving life. http://www.merinews.com/mobile/article/India/2014/10/17/give-way-to-ambulance-educates-people-that-saving-time-is-saving-life/15902114
3. Chattaraj, A., Bansal, S., Chandra, A.: An intelligent traffic control system using RFID. IEEE Potentials **28**, 40–43 (2009)
4. Pareit, D., Lannoo, B., Moerman, I., Demeester, P.: The history of WiMAX: a complete survey of the evolution in certification and standardization for IEEE 802.16 and WiMAX. IEEE Commun. Surv. Tutor. **14**, 1183–1211 (2012)
5. Tang, S.: Performance modeling of an integrated wireless network using WiMAX as backhaul support for WiFi traffic. Int. J. Wirel. Inf. Netw. **19**, 73–83 (2011)
6. Mittal, A., Bhandari, D.: A novel approach to implement green wave system and detection of stolen vehicles. In: 2013 3rd IEEE International Advance Computing Conference (IACC) (2013)
7. Sharma, S., Pithora, A., Gupta, G., Goel, M., Sinha,. M.: Traffic light priority control for emergency vehicle using RFID. Int. J. Invocations Eng. Technol. 363–366 (2013)
8. Sundar, R., Hebbar, S., Golla, V.: Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection. IEEE Sens. J. **15**, 1109–1113 (2015)
9. Gudipati, A., Perry, D., Li, L., Katti, S.: SoftRAN. In: Proceedings of the second ACM SIGCOMM workshop on Hot topics in software defined networking—HotSDN'13 (2013)
10. Damnjanovic, A., Montojo, J., Wei, Y., Ji, T., Luo, T., Vajapeyam, M., Yoo, T., Song, O., Malladi, D.: A survey on 3GPP heterogeneous networks. IEEE Wirel. Commun. **18**, 10–21 (2011)

11. Microchip Amtel, http://www.atmel.com/Images/Atmel-42735–8-bit-AVR-Microcontroller-ATmega328-328P_Summary.pdf
12. Cucchiara, R., Piccardi, M., Mello, P.: Image analysis and rule-based reasoning for a traffic monitoring system. IEEE Trans. Intell. Transp. Syst. **1**, 119–130 (2000)
13. Wu, Y.-M., Ye X.-Q., Gu W.-K.: A shadow handler in traffic monitoring system. In: IEEE 55th Vehicular Technology Conference VTC, Spring 2002

Review and Analysis of Simulation Results for Consumer Attack Model and Grid Sensor Placement Algorithm

Yuvaraj S. Patil and Swati V. Sankpal

Abstract In smart grid systems, distribution networks and smart meters are more exposed to the public and therefore are at higher risk. Smart meters are the critical elements of smart grid system. Smart meters provide information such as energy usage, utilization pattern, user details to control centers at real time. Smart meters are more exposed to cyber-attacks due to heavy usage of communication devices and communication infrastructure in smart grid system. Researchers across the globe are working on the cyber security of smart meters and have proposed various models and algorithms for network observability, attack detection and prevention. This paper provides a review and analysis of the simulation results of recently proposed consumer attack model and grid sensor placement algorithm for detecting the attacks.

Keywords Advance metering infrastructure · Communication networks
Cyber security · Distribution networks · Energy theft · False data injection
Grid sensors · Smart grid · Smart meter

1 Introduction

Smart grid systems use advanced communication and information technologies, control systems, and computing technologies. Smart grid systems help to modernize and optimize existing electric power systems [1–3]. Due to heavy usage of communication infrastructure, smart grids are more exposed to public and cause a number of security risks. Distribution systems mainly comprise of home and neighborhood networks [4], smart meters, and grid sensors. Smart meters are used

Y. S. Patil (✉) · S. V. Sankpal
D. Y. Patil College of Engineering and Technology, Affiliated to
Shivaji University, Kolhapur, Maharashtra, India
e-mail: yuvaraj.pat@gmail.com

S. V. Sankpal
e-mail: sankpal16@yahoo.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_13

as endpoints in home area networks. Smart meter records energy consumption of the user and provides real-time data to control centers periodically.

Control centers use the data received from the smart meters installed in the field and perform various analysis. Network operators and control centers cannot perform accurate real-time state estimation and analysis of the smart grid system without this data. If the attackers alters the smart meter data and injects the false data in the system by compromising the smart meters, then it may cause serious problems such as energy stealing, electricity outage, smart grid system damage, compromise of customer confidential information, improper functioning of smart grid system, and revenue loss.

Various techniques, secure devices, and secure communication networks are used [5] to prevent such type of attacks. Researchers have used direct current model for state estimation in power systems [6] and formulated an attack model for false data injection at smart meter level. The authors have proposed an intrusion detection system framework to detect attackers at smart meter level and proposed an algorithm for grid sensor placement using graph theory. This paper provides a review and analysis of the simulation results of the consumer attack model and grid sensor placement algorithm [7].

2 Attack Model

2.1 Basic DC Model

Control centers in smart grid system get measurement data from smart meters and sensors installed in the field, and based on this data, the control centers perform various activities like data processing, network state estimation, and observability analysis. The data obtained from the smart meters and sensors may not be always accurate because of measurement errors, equipment and network failures, noise signal introduced in communication network, and false data injection in the network by attackers. Control centers may lead to wrong decision due to inaccurate data.

DC power flow model is usually used for state estimation due to its simple computations and simplicity [6]. A basic DC model that has there state variables as shown in Fig. 1 is considered.

The black node P1 represents generation node, black nodes P2 and P3 represent load nodes (smart meters), white circle represents bus node (electric pole), gray rectangles represent sensors, and lines represent power connectivity. For a balanced network,

$$P1 + P2 + P3 = 0 \quad (1)$$

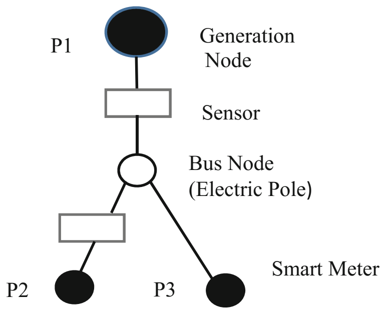


Fig. 1 Basic DC model. Adapted from [7, Fig. 1b]

2.2 Simulated Attack Model

Based on the basic DC model [6], a distribution network is formed using tree topology as shown in Fig. 2.

As shown in Fig. 2, the distribution network consists of main components: A generation node GS at which power is generated or delivered from other sources, and this node supplies the power P_G to the load nodes, set of SM nodes representing smart meters, set of EP nodes representing electric pole, the node R1 representing routing node, and lines representing connectivity. The simulation model also shows control center. The simulation model shown in Fig. 2 is created using NS3 simulation tool.

The distribution network shown in Fig. 2 is characterized by its network topology and a set of observed measurements

$$z = [P_G, P_1, P_2, \ldots, P_i]^T \in \mathbb{Z} \tag{2}$$

where P_G is the total amount of power generated, P_i indicated the energy consumption at smart meter.

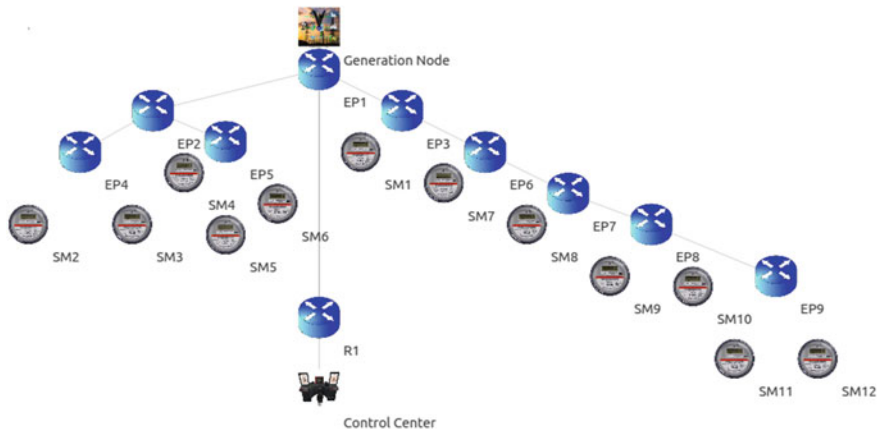


Fig. 2 Simulation model for distribution network

The attacker launches attack by constructing

$$\dot{z} = z + a = [\dot{P}_G, \dot{P}_1, \dot{P}_2, \ldots, \dot{P}_i]^T \neq 0 \tag{3}$$

where $\boldsymbol{a}$ is the attacker vector considered such that

$$a = [a_G,\ a_1 x_1,\ a_2 x_2,\ \ldots,\ a_i x_i]^T \tag{4}$$

where x_i represents that the smart meter of household I is compromised if $x_i = 1$ else $x_i = 0$.

3 Simulation Results and Analysis

3.1 Consumer Attacks in Different Scenarios

In this simulation, the actual amount of power that the attacker consumes is set at 10 kWh, and three different values (9, 8, 7 kWh) are considered to which the attacker aims to reduce his meter readings in order to steal the energy. It means that the difference in energy consumption between real and altered values has to be compensated by attacking neighboring smart meters in order to bypass the attack detection.

Constrained and no constrained conditions are considered for constraint level while the attacker performs such act of energy theft. In no constrained case, there are no boundaries for the attacker to steal the energy; i.e., the attacker can steal the desired amount of energy by compromising one or more smart meters from the neighbors. But in actual practice, control centers set limits/boundaries on fluctuation of energy consumption by consumers. For the simulation purpose, the limit of 1 kWh is set for constraint case.

Simulation results shown in Fig. 3 shows that more number of smart meters needs to be compromised to achieve the desired level of energy theft while bypassing energy theft detection performed by control centers.

The attacker needs to know the energy consumption patterns of his neighboring smart meters and also the upper and lower limits on energy consumption set by control centers. Hence, compromising large number of smart meters may not be possible practically.

3.2 Grid Sensor Placement and Detection Rates

This simulation shows how the detection rate varies with different number of grid sensors placed in the network. Figure 4 shows how the rate of detecting the consumer attacks can be improved by increasing the number of grid sensors.

From Fig. 4, it is noticed that the slope of detection rate becomes steeper when the number of grid sensors is smaller. On the other hand, when the number of grid

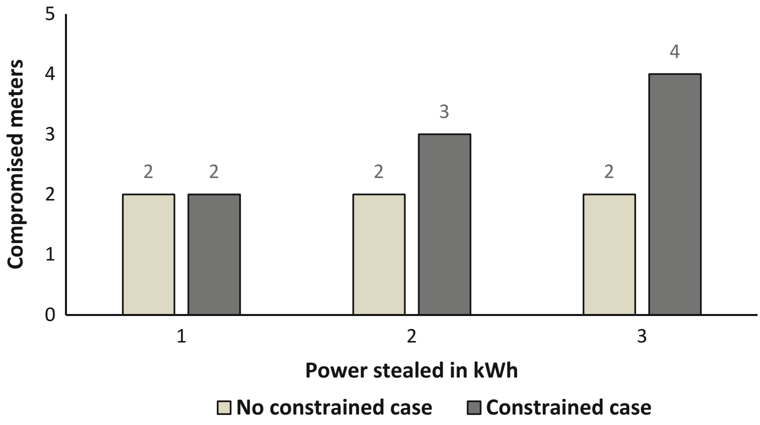


Fig. 3 Attack requirements under different constraints conditions. Adapted from [7, Fig. 6]

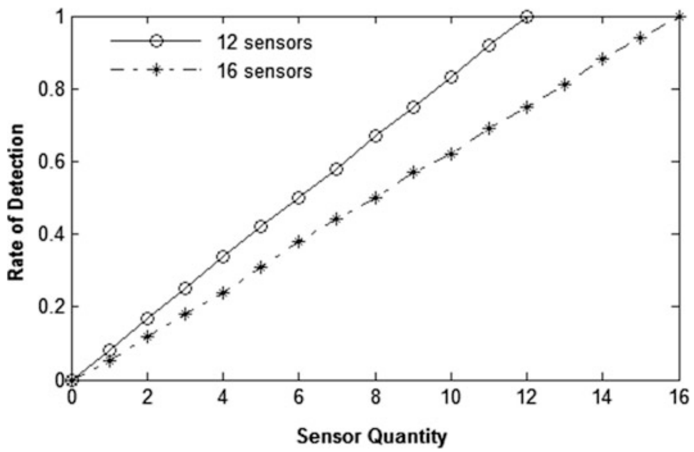


Fig. 4 Sensor quantity versus detection rate. Adapted from [7, Fig. 8]

sensors is increased, the slope of detection rate declines. This means that the network with less number of grid sensors installed becomes more vulnerable than the network with more number of grid sensors installed.

4 Conclusion

This paper provides a review and analysis of the simulation results of the consumer attack model and grid sensor placement algorithm. The attacker can steal the desired amount of energy by compromising one or more smart meters in unconstrained scenarios. In unconstrained scenarios, more number of smart meters needs to be compromised to achieve the desired level of energy theft while bypassing energy theft constraints set by control centers. The consumer attacks detection rate can be improved by increasing the number of grid sensors.

Acknowledgements The review and analysis of simulation results of the consumer attack model and grid sensor placement algorithm presented in this paper is based on the research work proposed by the authors Lo and Ansari [7].

References

1. Santacana, E., Rackliffe, G., Tang, L., Feng, X.: Getting smart—with a clearer vision of the intelligent grid, control emerfrom chaos. IEEE Power Energy Mag. 41–48 (2010)
2. Amin, S.M., Wollenberg, B.F.: Towards a smart grid. IEEE Power Energy Mag. 34–41 (2005)
3. Garrity, T.G.: Getting smart—innovation and trends for future electric power system. IEEE Power Energy Mag. 38–45 (2008)
4. Lo, C.H., Ansari, N.: CONSUMER: a novel hybrid intrusion detection system for distribution networks in smart grid. IEEE Trans. Emerg. Topics Comput. **1**(1), 33–44 (2013)
5. Giani, A., Bitar, E., Garcia, M., McQueen, M., Khargonekar, P., Poolla: Smart grid data integrity attacks. IEEE Trans. Smart Grid **4**(3), 1244–1253 (2013)
6. Xiao, Z., Xiao, Y., Du, D.-C.: Exploring malicious meter inspection in neighborhood area smart grids. IEEE Trans. Smart Grid **4**(1), 214–226 (2013)
7. Van Hertem, D., Verboomen, J., Purchala, K., Belmans, R., Kling, W.: Usefulness of DC power flow for active power flow analysis with flow controlling devices. In: Proceedings. 8th IEEE International Conference on ACDC, pp. 58–62 (2006)

An Approach to Analyze Behavior of Network Events in NS2 and NS3 Using AWK and Xgraph

Radhika Patel, Nehal Patel and Sandip Patel

Abstract In the world of rapidly growing technology, recent update demands robust, reliable, and evaluated network protocols. The behavior of protocols needs to be examined thoroughly. There are numerous methods available for evaluating and analyzing the behavior of protocols. The most famous are cloud computing and network simulation. A simulation tool named network simulator2 is discussed in detail with the help of 6-node simulation example consisting of the details about Awk script (for performance analysis) and Xgraph (for plotting graph). Moreover, comparison of different versions of network simulator and comparison between NS2 and NS3 are discussed which suggests that NS3 has more advantageous features than NS2. Experimental results of Ad hoc Distance Vector Routing protocol (AODV) and Dynamic State Routing protocol (DSR) in both, NS2 and NS3, are also included with respect to throughput and end-to-end delay for 25 nodes.

Keywords Simulation · NS2 · NS3 · OTcl · C++ · NAM Trace · Awk · QoS · Virtualization · Xgraph

1 Introduction

The first simulation is the copy or reflection of some real thing or a process which is extremely important for extending science and technology in this modern world. It can be applied to various areas of science and engineering for fulfilling different purposes [1]. Network simulation is used in various study areas such as academic research, industrial development for analyzing, designing, simulating, and verifying

R. Patel (✉) · N. Patel · S. Patel
Charusat University, Anand, Gujarat, India
e-mail: radhipatel999@gmail.com

N. Patel
e-mail: nehalpatel.it@charusat.ac.in

S. Patel
e-mail: sandippatel872@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_14

the performance of different network protocols. Analyzing the behavior can be done using two techniques, namely cloud computing and network simulation. Cloud computing is one of the incredible technologies which enable the new vision for IT industry [2]. Simulation is a key component of network research which needs debugging, independency from existing hardware or software and parameter exploration and examination [3]. For understanding the behavior and performance of the networks and protocols working in the network, simulation plays a significant role in the research work. It helps the network designers as well as the researchers to gain more knowledge in the particular field.

In the present scenario, wireless sensor network is a vast field of exploration that is attracted more by network researchers. Wireless sensor network is the combination or fusion of several tiny sensors called nodes. Each sensor node is connected to other sensor nodes to communicate via number of links. Performing real experiments for these networks is a time-consuming as well as difficult task. As there is no surety of successful outcome, new untested protocols cannot be simulated on this platform. So, there is a requirement to test the new protocols with the help of simulation.

Network simulator2 is a discrete event network simulator software that predicts the behavior of a computer network, developed by US Berkeley [1]. In networking, it is necessary to know the accurate behavior of any system. Network simulators are used to fulfil this task. The computer network is designed using devices, links, and applications for observing and analyzing the performance in the simulators. They support the most popular networks and technologies.

2 Architecture of NS2

Ns2 is an interpreter which uses Object-Oriented TCL (OTcl) script along with network component object libraries, network module libraries, and network simulation event scheduler [1]. It is widely used for setting and running a network simulation. For simulation, the user of NS2 uses OTcl script for preparing program code [4]. For implementing control code, NS2 uses C++ programming language.

The NS2 architecture is composed of five parts [1]:

- Tclcl
- Event scheduler
- OTcl library
- Network components
- Tcl 8.0 script language (Fig. 1).

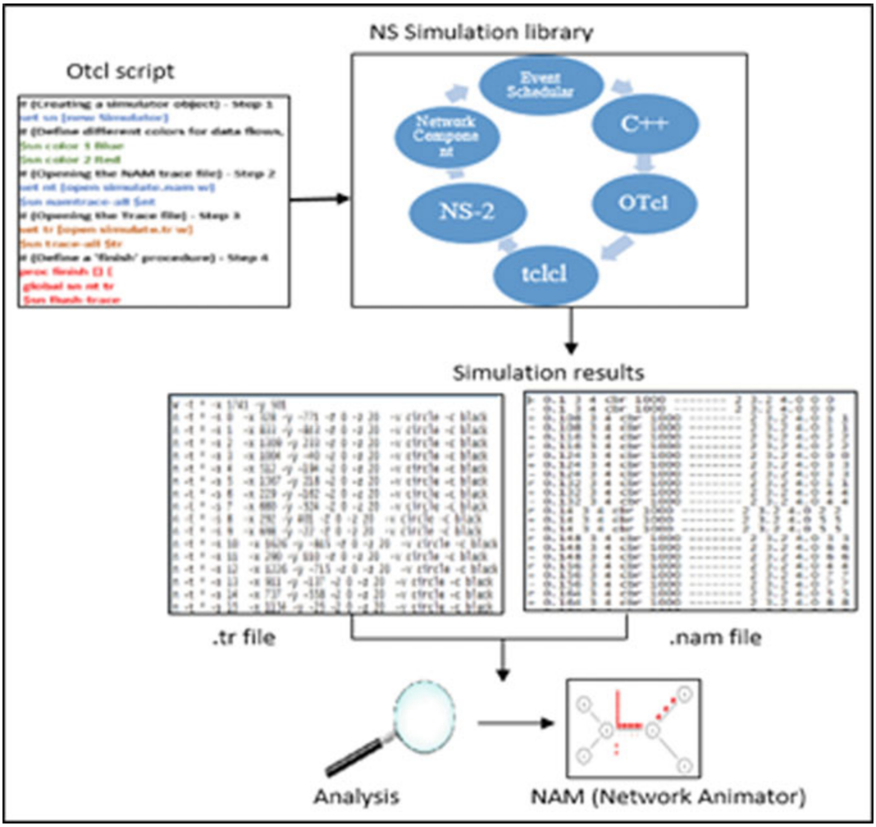


Fig. 1 NS2 architecture model

2.1 NS2 Simulation with Example

Figure 2 shows the architecture of NS2 with 6 node example. It consists of a tcl code which generates two files with ".nam" and ".tr" extension.

To visualize the animation of the script written in ns2, NAM animator is used [5]. To run NAM, the command used is:

nam filename.nam

The code of ns2 is to be written in tcl, and the output file is displayed in the NAM window. While execution of tcl code, it provides with two files named "simulate.nam" and "simulate.tr" as an output. An example of a 6-node network exchanging packets within the network is available on the blog [6]. The syntax for creating a trace file is as follow:

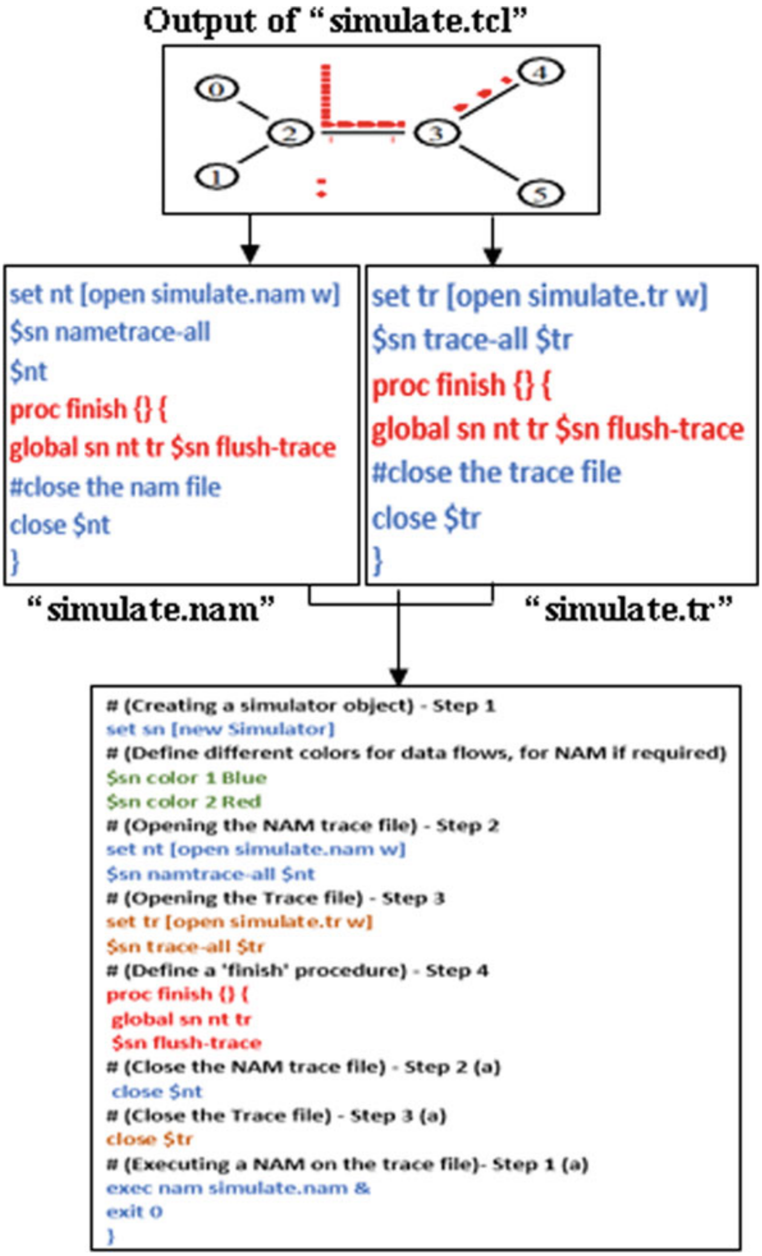


Fig. 2 Complete "tcl" code for 6 nodes in the network

```
#Create simulator object
Set sn [new simulator]
#Open the trace file
Set tr [open simulate.tr w]
$sn trace-all $tr
#close the trace file
Close $tr
```

3 AWK Scripts for Analysis

AWK is a high-level programming language that can be used in processing text files. These scripts are used in network simulation for calculation of parameters such as end-to-end delay, packet delivery ratio, throughput, packet loss ratio [7]. Initially, we need to install gawk in Linux using the following command.

sudo apt-get install gawk

After installation of gawk package, awk script is to be designed. This ".awk" file is in the same location where the ".tcl" file resides. The following snippet shows the "awk" code for obtaining throughput of the network (Fig. 3).

The figure shown below is the screenshot of the output obtained of awk file codes as described above (Fig. 4).

After installing the gawk, following command is to be executed to run the awk file.

gawk -f <awk file name> <trace file name>

i.e. **gawk -f file.awk file.tr**

4 Comparative Analysis of NS2 and NS3

a. **Architecture**: There is a difference of architecture between network simulation tool NS2 and NS3. NS2 requires two languages namely C++ and OTCL. While NS3 simply requires C++ language or python as a scripting language, the following snippet shows the example of script in NS2 and NS3, respectively (Table 1, Fig. 5).
b. **Recompilation**: In NS2, recompilation is a long process, while in NS3, it is very simple and done using a single command, ./waf.

```
BEGIN {
seqno= -1;
dp = 0;
rp = 0;
cnt = 0;
}
{
if ($4==”AGT” && $1==”s” && seqno<$6)
{
seqno=$6;
}
else if (($4==”AGT”) && ($1 ==”r”))
{
rp++;
}
else if ($1==”D” && $7==”tcp”)
{
dp++;
}
```

Fig. 3 Throughput .awk

```
user@user-laptop:~/project/PROGRAM/aodv$ sudo apt-get install gawk
[sudo] password for user:
123
Sorry, try again.
[sudo] password for user:
Reading package lists... Done
Building dependency tree
Reading state information... Done
gawk is already the newest version.
The following packages were automatically installed and are no longer required:
  language-pack-kde-en language-pack-kde-en-base
Use 'apt-get autoremove' to remove them.
0 upgraded, 0 newly installed, 0 to remove and 0 not upgraded.
user@user-laptop:~/project/PROGRAM/aodv$ gawk -f genthroughput.awk aodv.tr
Average Throughput[kbps] = 322.68            StartTime=1.00 StopTime=55.01
user@user-laptop:~/project/PROGRAM/aodv$
```

Fig. 4 Output of executing awk code

Table 1 Comparison of the features of NS2 and NS3

Criteria	NS2	NS3
Released year	1996	2008
Name of simulator	C++, Otcl	C++, Python
Language platform	Windows, Linux	Windows, Linux, Mac OS
Cost and licenses	Free, open source	Free, GNU general public license
Network support type	Wired network, wireless ad hoc mode, wireless managed mode	Wired network, wireless network, wireless sensor network
User interface	Command-line interface	Command-line interface
API	Pure event base	Low-level, users can mix and match between the simpler API
Simulation	Virtual	Real
Execution	Moderate	Excellent
Complexity	More complex	Less complex
Compatibility	More compatible	Less compatible
Memory consumption	Highest	Lowest

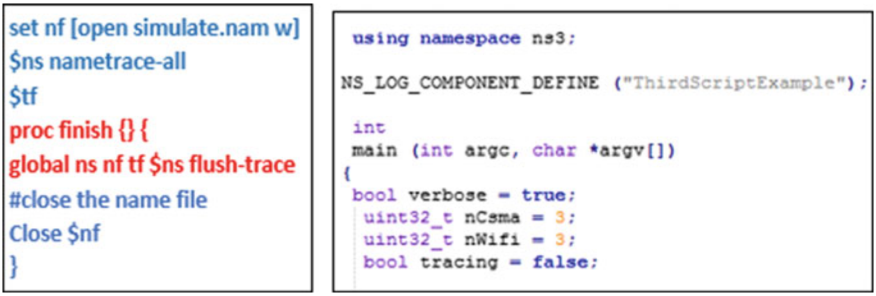


Fig. 5 Script in NS2 and NS3

c. **File formats for post analysis**: A TCL script in NS2 creates two files namely trace file and nam file with extension .tr and .nam, respectively. In addition to trace file and animation file, xgraph can also be created in NS2 for graphical representation.

In NS3, the script creates xml files with .xml extension, .tr file for trace analysis, and gnuplot for graphical representation.

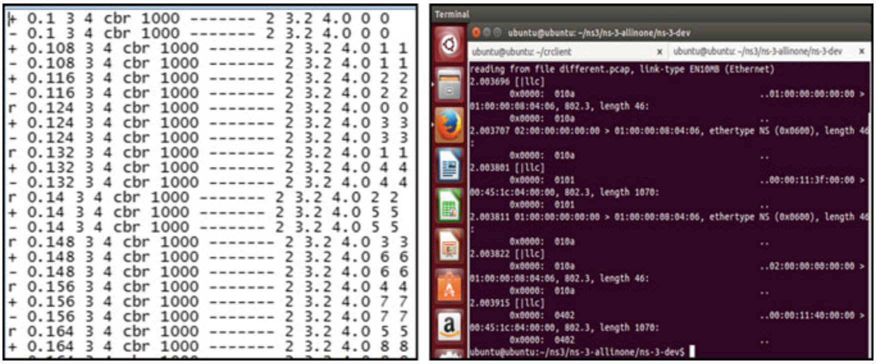


Fig. 6 Trace file in NS2 and NS3

d. **Comparison of trace files in NS2 and NS3**: The file created or written by an application itself in order to store the coverage information or the network information in NS2 is called trace file [8] (Fig. 6).

Comparison of Graphical representation in NS2 and NS3:

Xgraph, a graphical representation in NS2, helps in plotting data from number of files in a single graph. It has the capability of handling unlimited dataset size and data files [9].

To invoke Xgraph, user has to add path in the .login file that directs the location where Xgraph is installed. For example:

'/dell/home/ns2/graph/xgraph'

To plot the graph of any file, following code is to be executed.

xgraph <filename> .dat

Gnuplot is a command-line program that can generate 2D and 3D plots of functions, data fits, and data. It is recurrently used for education as well as publication-quality graphics (Fig. 7).

e. **Visualization of simulation output**: NS2 contains nam animator as a visualization tool while a tool named NetAnim is used for visualizing the output in NS3 (Fig. 8).

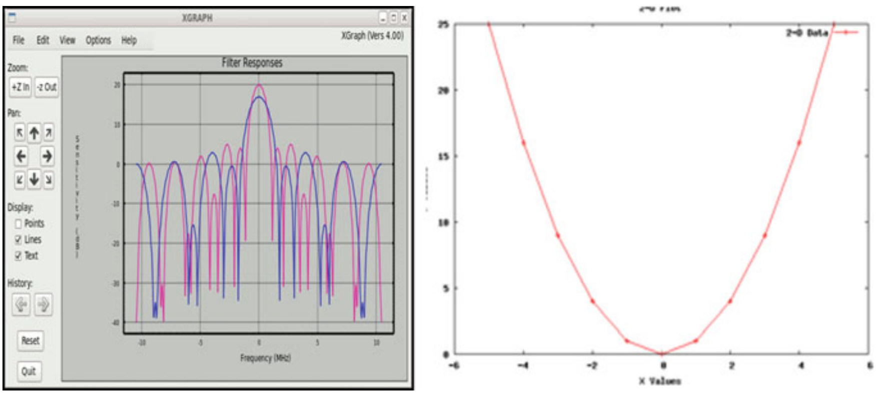


Fig. 7 Xgraph in NS2 and Gnuplot in NS3

5 Experimental Results and Simulation Analysis

Both the network simulators namely NS2 and NS3 execute simulations. We have examined the performance in terms of QoS parameters like throughput and end-to-end delay of MANET routing protocols AODV and DSR in NS2 and NS3 for 25 nodes. Table 2 depicts the variance in the performance parameters of the respective protocols.

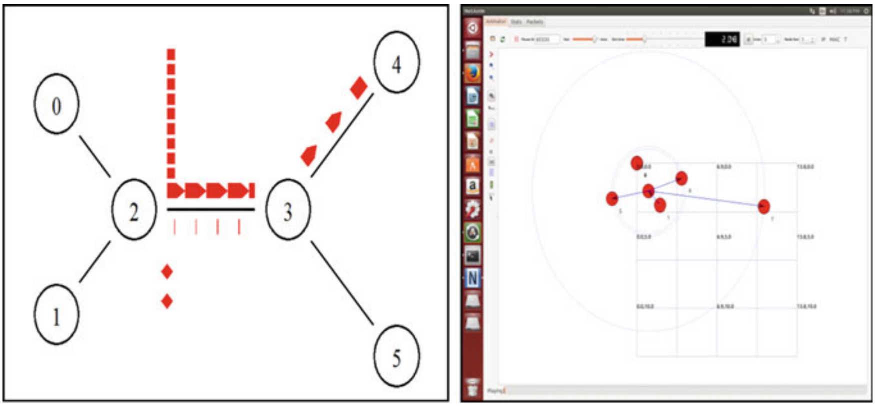


Fig. 8 Output in NAM window (NS2) and output in net animator (NS3)

Table 2 Comparative analysis of performance parameters

Protocols	Throughput (kbps)		End-to-end delay (ms)	
	NS2	NS3	NS2	NS3
AODV	0.9	0.9	16.7	15.5
DSR	0.86	0.9	17.8	10.01

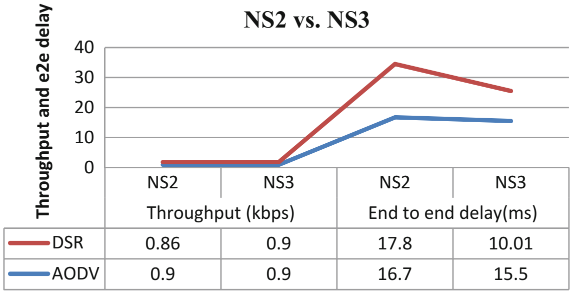


Fig. 9 Comparative analysis of NS2 and NS3

5.1 *Graphical Representation of the Comparative Analysis*

The below shown graph shows the variance in the values obtained of throughput and end-to-end delay for both the protocols, AODV and DSR in NS2 and NS3 (Fig. 9).

6 Conclusion

Network simulator focuses on a wide range of issues rising from limited energy resources, computational power, communication capabilities, and wireless nodes. Currently, NS2 is the best option, as NS2 is an object-oriented event-driven simulator that provides simulation environment for network engineers and researchers to investigate regarding the behavior and performance of various protocols with different topologies and configurations. Network simulator supports multicasting and routing of protocols, simulation of TCP, over wireless as well as wired networks. NS2 provides accurate results of performance parameters, and it is cost-effective and has more memory storage compared to other versions. Moreover, a 6-node example in NS3 is executed to examine the difference between the version 2.35 and version 3.25 of network simulator. The performance comparison shows that NS3 is more advantageous in terms of memory management, animation interface and some performance parameters.

References

1. Singh, N., Dua, R.L., Mathur, V.: Network simulator NS2–2.35. Int. J. Adv. Res. Comput. Sci. Softw. Eng. **2**(5), 224–228 (2012)
2. Patel, S., Patel, R., Patel, H., Vahora, S.: Cloud analyst: a survey of load balancing policies. Int. J. Comput. Appl. **117**(21) (2015)
3. Pan, J., Jain, R.: A survey of network simulation tools: current status and future developments. Email: jp10@cse.wustl.edu **2**(4), 45 (2008)
4. De Melo, T.A., de Oliveira, F.D., Semente, R.S., Benjamim, X.C., Salazar, A.O.: WINSS: a simulation tool of the IEEE 802.15. 4 standard for network simulator 2. In: International Symposium on Instrumentation Systems, Circuits and Transducers (INSCIT), pp. 43–48. IEEE (2016)
5. http://nile.wpi.edu/NS/analysis.html
6. https://sandip9334.blogspot.in/2017/02/ns26.html
7. Gupta, S.G., Ghonge, M.M., Thakare, P.D., Jawandhiya, P.M.: Open-source network simulation tools: an overview. Int. J. Adv. Res. Comput. Eng. Technol. (IJARCET) **2**(4), 1629 (2013)
8. http://www.ns2blogger.in/p/awk-scripting-on-ns2.html
9. http://www.xgraph.org
10. Jha, R.K., Kharga, P.: A comparative performance analysis of routing protocols in MANET using NS3 simulator. Int. J. Comput. Netw. Inf. Secur. **7**(4), 62 (2015)
11. Karl, M.: A comparison of the architecture of network simulators ns-2 and tossim. In: Proceedings of Performance Simulation of Algorithms and Protocols Seminar. Institute for Parallel und Verteilte Systeme, Abteilung Verteilte Systeme, University Stuttgart (2005)

Hybrid Latent Semantic Analysis and Random Indexing Model for Text Summarization

Niladri Chatterjee and Nidhika Yadav

Abstract Latent semantic analysis has been used successfully for extractive text summarization for years, while random indexing-based summarization has been recently proposed in the literature for text summarization. The random indexing-based summarization inherently uses graph-based ranking techniques. In this paper, we propose a hybrid technique of latent semantic analysis and random indexing for text summarization. Further, we have performed experiments to compare the results with several related baseline methods. The effectiveness of the hybrid method so developed is evident from the relative increase in the results over the baseline LSA-based technique.

Keywords Latent semantic analysis · Random indexing · Text summarization SVD · Graph · Extraction · LSA-RI

1 Introduction

With the abundance of text information available, an important problem is to extract the important information for this huge textual data. Text summarization plays its key and important role here. Text summarization [1] selects the most important sentences from a given document collection. Various techniques have been developed for text summarization starting from traditional keywords-based summarization, through heuristics-based approaches to modern sophisticated mathematical techniques, such as latent semantic analysis (LSA) [2], and machine learning techniques, such as deep learning. Below, we briefly discuss some important summarization techniques:

N. Chatterjee · N. Yadav (✉)
Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, Delhi, India
e-mail: nidhikayadav8@gmail.com

N. Chatterjee
e-mail: niladri@maths.iitd.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_15

- Graph-based techniques [3]: A document is represented as a graph with sentences as vertices and their similarity as edge weights. Further, the sentences are ranked in increasing order of importance using PageRank algorithm. Graph-based techniques are elaborately used for summarization.
- LSA-based summarization [2, 4–6]: The techniques are discussed in detail in Sect. 2.
- Lexical chain-based summarization [7]: Lexical chain-based summarization uses semantic knowledge of the document with the help of WordNet to find key topics covered in text in form of lexical chains. Further, the strongest chain is selected. And corresponding sentences are extracted. Various modifications of the method have been proposed in the literature.
- Random indexing: Random indexing-based text summarization was first proposed by Chatterjee and Mohan [8]. Random indexing differs from traditional Word Space Models in the sense that here instead of conventional word vectors one uses ternary vectors of predetermined size to represent words. Chatterjee and Sahoo [9] have enhanced the above model using convolution in their modified random indexing approach.

Gong and Liu [2] used latent semantic analysis (LSA) to extract important sentences for the process of text summarization. LSA uses the insights of the well-known singular value decomposition (SVD) principle to generic summaries. SVD of a matrix is a mathematical approach to partition a matrix into three matrices with unique properties. Given a matrix A of dimension $m \times n$, SVD reduces A to the form $A = USV'$, where U is an $m \times m$ matrix, S is an $m \times n$ diagonal matrix, and V is an $n \times n$ matrix. The diagonal elements of S are nonnegative real numbers determined by square roots of nonzero characteristic values of AA' and are called singular values of A [2]. The columns of U are orthonormal eigenvectors of AA', while columns of V are orthonormal eigenvectors of $A'A$.

LSA uses huge matrix-based computations. Matrix computations are computationally expensive and time-consuming. Random indexing reduces the size of word space model as compared to the size of the vector space model. It encouraged us to consider a hybrid model based on random indexing and LSA. Apart from efficiency in computational time, random indexing represents words using context vectors which describe the knowledge expressed in a more genuine way as compared to the vector space model.

The paper is organized as follows. Sect. 2 discusses latent semantic analysis in detail and its application to text summarization. Some key past works in this area have also been summarized. In Sect. 3, the new LSA-RI hybrid technique is proposed. Section 4 examines the experimental results and their comparisons with baseline methodologies. Finally, conclusions and future works are discussed in Sect. 5.

2 Latent Semantic Analysis for Text Summarization

The extraction of important sentences using LSA is based on the concept that a term is key term if it is present in many sentences which are crucial and a sentence is crucial if it has terms which are crucial. The component values of matrices U and V provide us with degree of significance of a sentence and a term. The process starts with the creation of a word by sentence matrix, TS. There are various techniques to generate the matrix TS. Traditionally, it is computed using the product of local weighing factor, L(s, w), and global weighting factors, G(s, w), where s is the sentence and w is the corresponding word w. Local weighting factor is the local importance of a sentence and word, while global weighing is global importance of a sentence and word. Global weighting scheme for text is typically 1 or is taken as the inverse document frequency, G(w) = log(N/n(w)), where N is the number of sentences in the document and n(w) is the number of sentences that contain the word w. Normalization can be performed everywhere.

Some of the local weighting schemes by Gong and Liu [2] are:

1. Binary weight: L(s, w) = 1 if word appears at least once in the sentence.
2. Weighted frequency weight: L(s, w) = TF(s, w), where TF(s, w) is the number of times word i occurs in the sentence.
3. Logarithmic weighting: L(s, w) = log(1 + TF(s, w)).

Once the matrix TS is generated, SVD is applied subsequently to it. Various algorithms have been used in the literature for extracting the sentences from a document using SVD. Gong and Liu [2] proposed LSA-based technique for text summarization: The idea behind this approach, according to Gong and Liu [2], is that the characteristics vectors and hence the diagonal elements of S are arranged in decreasing values. The singular vector resents the key concepts with least redundancy and arranged in order of increasing importance. The key steps involved in summarization are:

1. TS matrix computations: Computations are performed as above.
2. Application of SVD: SVD is applied on matrix TS.
3. Selection of key concepts: Key concepts are related to the high singular values.
4. Selection of key sentences: From each key concept, an important sentence is selected based on the right singular vector magnitude.

The detailed algorithm is given in Fig. 1. In the following section, we give an introduction to random indexing and the proposed LSA-RI algorithm.

```
Algorithm: Gong and Liu
INPUT: Given document to be summarized
OUTPUT: Key sentences
i)    The matrix TS is generated as discussed above.
ii)   Set NumSent = 1.
iii)  Set Extract = {}
iv)   Apply SVD on matrix TS, TS = U . S .V', where decomposition is as discussed previously.
v)    While NumSent< (ReducedDimension&&SumamryLength)
vi)   Set temp = NumSent^th row of V'.
vii)  Find the index of temp with maximum value.
viii) If index present in Extract. Goto Step vi, with temp = temp – {max}
ix)   Else ,Extract = Extract U { index }
x)    NumSent = NumSent + 1
xi)   End While
xii)  End
```

Fig. 1 Algorithm 1 for LSA-based generation for extracts

3 Proposed Technique

Random indexing was introduced to tackle the problem of high-dimensional matrix generated from vector space model. The LSA techniques discussed above all require to compute the high-dimensional word by sentence matrix. We propose to use the technique of random indexing to reduce the dimension of the problem to be solved by SVD algorithm. Below, we briefly introduce the concepts used in random indexing. Chatterjee and Sahoo [9] have established the use of random indexing for text summarization. Random indexing allows one to represent a document with vectors of a fixed dimension that is lower than the dimension of typical one-hot representation. At the same time, it establishes the relations between the terms that co-occurs up to a window size in the document. There are two important vectors in random indexing. They are:

- Index vectors: All words in the document are enumerated. And each of the words is assigned a unique ternary vector I. The values the elements of I can take are 1, 0, and −1. With 1 restricted to first half of the vector and −1 restricted to second half and the rest of elements are zero. The number of ones and −ones is very small as compared to the number of zeros.
- Context vectors: They represent the syntactic and semantic representations of words in the document represented in terms of index vectors. A context vector is generated by summing the weighted index vectors up to a particular window size. The weights decrease with the distance from the context vector.

The dimension of index and context vectors has been used as $2\sqrt{m}$, where m is the number of unique terms in the given document. This clearly shows that there is a marked reduction in the dimensionality of the representation space. The random indexing-based summarization is based on computation of similarity of between the sentence using cosine and/or Euclidean similarity and applying graph-based

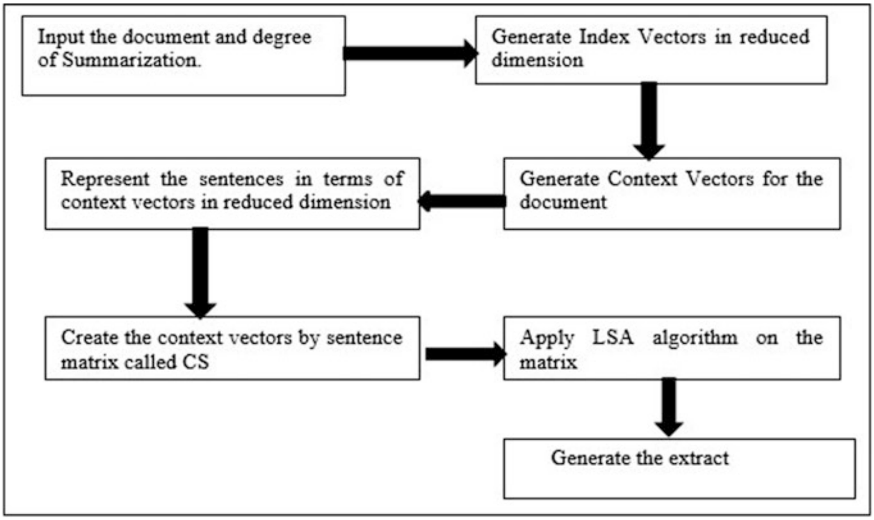


Fig. 2 Flowchart describing the LSA-RI algorithm for generation for extracts

ranking algorithm over it. While Chatterjee and Sahoo [9] have used graph-based techniques to rank the sentences, we propose to use LSA-based technique to rank the sentences of a document. Figure 2 describes the flowchart in detail. The key difference is the generation of matrix TS. Here, the matrix TS is generated using random indexing algorithm used by Chatterjee and Sahoo [9].

4 Results and Observations

The efficiency of the proposed algorithm has been tested with the following three popular extractive summarization algorithms available in the literature:

(1) LSA-based summarization [2],
(2) Graph-based summarization [3],
(3) MRISUM algorithm [9].

We have modified Gong and Liu [2] LSA model to work with random indexing. So it is very natural to compare the results with baseline LSA approach. Other graph-based algorithms are tested since random indexing algorithm has been applied to summarization in combination with graph-based techniques. We have tested the systems with a compression rate of 50–10% summarization. The comparisons are performed with precision and F-measures for various ROUGE scores [10]. The results for 50% summarization are presented in Figs. 3 and 4. Figure 5 depicts the results so obtained for 10% summarization.

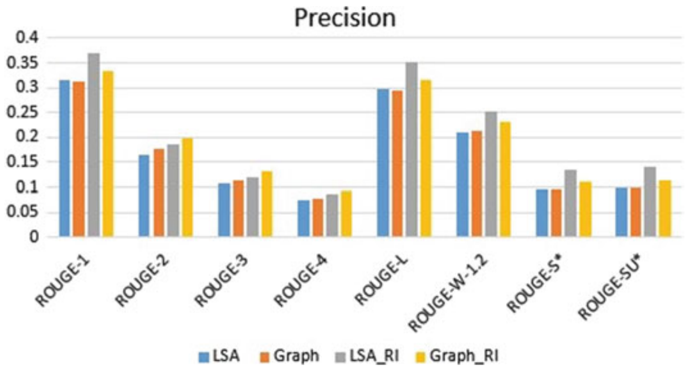


Fig. 3 Scores for 50% summarization

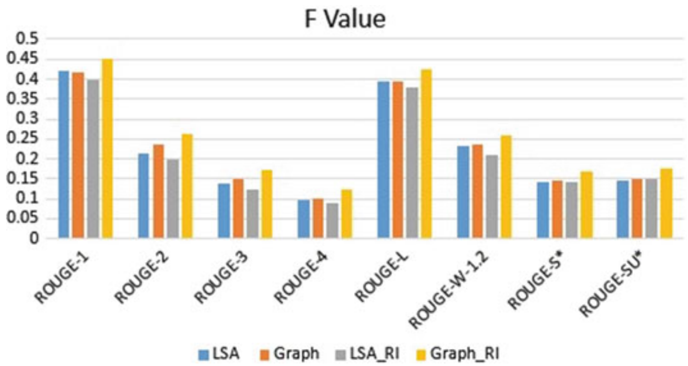


Fig. 4 F-scores for 50% summarization

It is evident in from Fig. 3 that the precision scores of LSA-RI are better than the precision scores of all the baseline algorithms for 50% summarization compression while the F-values for 50% compression are little behind the baseline LSA algorithm. Also, it becomes clear that graph-based random indexing produce the best F-value scores. The results are preliminary measures that the proposed approach of random indexing and LSA can be explored further for application in summarization.

5 Conclusion and Future Work

This paper proposes a new hybrid methodology based on latent semantic analysis and random indexing, namely LSA-RI. LSA-RI aims at improving computational time required by matrix computations of SVD algorithm. Further, this method aims

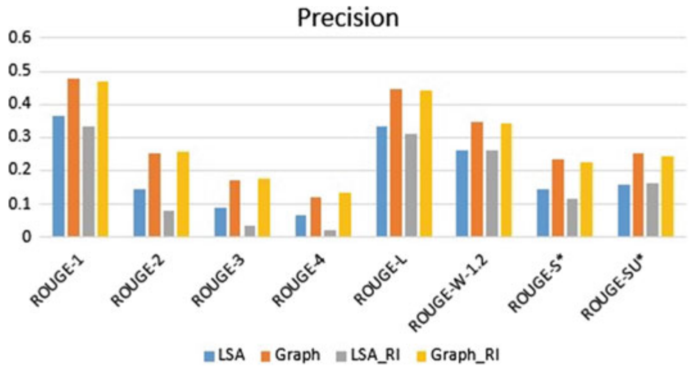


Fig. 5 Precision scores for 10% summarization

to enhance the performance of LSA-based technique. The precision scores for 50% summaries are comparable to LSA baseline summarizer, while for 10% summarization the precision scores are little below the baseline LSA. The current work emphasizes that the efficiency of random indexing with graph-based technique is superior to random indexing with LSA-based technique. These are initial experiments using this approach, and we plan to develop a more robust model which overcomes these disadvantages and at the same time focus on advantages of the proposed methodologies. Further, we have used the base techniques for both LSA and random indexing. In our future work, we are enhancing using the mature models of both LSA and random indexing. We are planning to work on doing comprehensive testing on standardized data sets which are used for summarization to test the validity of the proposed algorithm. Currently, we plan to work on evaluation of other models of random indexing with LSA for their efficacy.

References

1. Mani, I.: Automatic summarization, vol. 3. John Benjamins Publishing (2001)
2. Gong, Y., Liu, X.: Generic text summarization using relevance measure and latent semantic analysis. In: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 19–25 (2001)
3. Mihalcea, R.: Language independent extractive summarization. In: Proceedings of the ACL 2005 on Interactive Poster and Demonstration Sessions, pp. 49–52 (2005)
4. Kireyev, K.: Using latent semantic analysis for extractive summarization. In TAC (2008)
5. Steinberger, J., Jezek, K.: Using latent semantic analysis in text summarization and summary evaluation. In Proceedings of ISIM'04, pp. 93–100 (2004)
6. Yeh, J.Y., Ke, H.R., Yang, W.P., Meng, I.H.: Text summarization using a trainable summarizer and latent semantic analysis. Inf. Process. Manag. **41**(1), 75–95 (2005)
7. Silber, H.G., McCoy, K.F.: Efficient text summarization using lexical chains. In: Proceedings of the 5th International Conference on Intelligent User Interfaces, pp. 252–255 (2000)

8. Chatterjee, N., Mohan, S.: Extraction-based single-document summarization using random indexing. In: 19th IEEE International Conference in Tools with Artificial Intelligence, pp. 448–455 (2007)
9. Chatterjee, N., Sahoo, P.K.: Random indexing and modified random indexing based approach for extractive text summarization. Comput. Speech Lang. **29**(1), 32–44 (2015)
10. Lin, C.Y.: Rouge: a package for automatic evaluation of summaries. In: Proceedings of the ACL-04 Workshop, vol. 8 (2004)

State-of-the-Art Energy-Efficient Thermal-Aware Scheduling in Cloud

Ritu Garg and Rama Rani

Abstract Cloud computing delivers services on demand basis. It provides resource provisioning which includes CPU, memory, and networks. As cloud users are increasing, resource count is also increasing. The cloud resources consume the enormous amount of energy and produce CO_2 emissions. Thus, energy consumption and thermal management are the major challenges for cloud service providers due to increased use of computational resources. Further, the cooling energy efficiency which is affected by thermal environment is the main issue in the recirculation of hot air and it creates hotspots along with inlet temperature distribution. In data centers, electricity can be saved in two areas: computing and cooling. Data center servers need a constant supply of cold air from the cooling mechanisms. In this paper, we present the detailed survey of existing approaches for energy-efficient thermal-aware scheduling in cloud computing environment. We classified the scheduling approaches into energy management mechanisms and thermal-aware techniques for cloud computing system.

Keywords Cloud computing · Energy management · Thermal-aware RC model

1 Introduction

Cloud computing provides on demand, flexible and scalable services through powerful computing servers and enormous data storage on the pay as you use model. It is the dynamic provisioning of services that use virtual machine techniques for consolidation [1]. Cloud-based services are reliable, dynamic, on-demand, scalable,

R. Garg · R. Rani (✉)
Department of Computer Engineering, National Institute of Technology,
Kurukshetra, Kurukshetra, Haryana, India
e-mail: ramachawla27@gmail.com

R. Garg
e-mail: ritu.59@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_16

and device-dependent for storage and processing of data. NIST is given a comprehensive definition of cloud computing as "cloud computing is a framework for enabling the on-demand network access to resources like networks, servers, storage, applications, and services that can be quickly provisioned and released with minimal effort" [2]. With the increased use of computational resources, the expenditure on energy usage and server cooling is estimated to be very high. Due to the scale and complexity of data centers, energy-efficient scheduling is the key challenge in cloud environment. Most of the energy models are based on the hypothesis that the total energy consumption of a data centers depends on the processor utilization [3]. A linear relationship exists between the utilization of processor and energy consumption. This relationship can be utilized to calculate the server total power consumption which is given in Eq. (1):

$$p_u = p_i + (p_m - p_i) * u_c \tag{1}$$

where

$(p_m - p_i) * u_c$	represents the dynamic power consumption,
p_i	static power consumed by the processor in idle state,
p_u	peak power consumed by the processor when the server is 100% used, and
u_c	current CPU utilization

Power consumption in cloud data centers is primary issue for cloud providers. Power costs of a data center lie between 25 and 40%. In August 2014, the Natural Resources Defense Council (NRDC) published a data center efficiency assessment to outline the scale of data centers around the world [4]. This shows that by 2020, US data centers power consumption will be 140 billion KWH electricity that is equal to approximately 50 power plants productivity. Another important point in an analysis by the NRDC shows that 30% of servers are outdated or underutilized or not needed [4].

2 Energy Management in Cloud Environment

According to the World Energy Council [5], energy efficiency refers to decline in energy which can be used for some useful services. In paper [6], the author defines an energy model through static and dynamic power consumption which deals only with energy waste while running. To enhance energy efficiency and cost efficiency, one should use appropriate techniques for power management.

2.1 Classification of Energy Management

In this section, we present the classification of various energy management approaches as shown in Fig. 1.

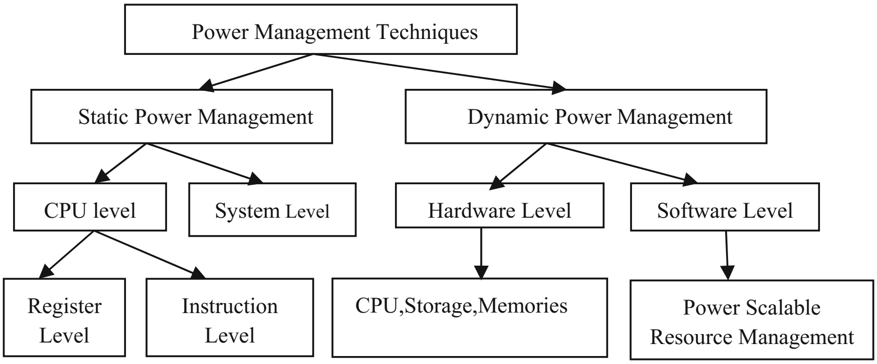


Fig. 1 Classification of power management techniques [6]

Static Power Management

Static power management is known as off-line energy management which deals with circuitry systems. Static power consumption is due to the usage by complementary metal-oxide semiconductor (CMOS) circuits. In any active circuit, the static power consumption is occurred by current leakage which is not dependent on clock cycle. This static power is determined by transistors and process technology [6]. It mainly concerned with the geographical distribution of the processing centers, manipulation of circuits, architecture redesign, instruction sets, sizing of transistor, path balancing. In static power management, the energy consumption is managed at CPU level and system level. Out of all the components, CPU consumes 35–50% of energy and gives the scope for optimizing the energy consumption. Along with CPU, there are other components such as memory devices, network system, and software facilities which are helpful to optimize energy consumption. System-level static power management methods have been used to regulate the energy consumption by such components. At CPU level, the optimization is performed at register level or at instruction set level. The parameters which are used to reduce the energy consumption are considered by optimizing the register transfer level (RTL) activities at register level. Different instruction set architectures (ISAs) have been proposed to reduce the power consumption at ISA level [7]. Static power management (SPM) consists of the optimization techniques which are used at the design time, at the hardware level and also at software level. In the static hardware level, devices can be modified as the low-power CPU server which distributes the workload to the devices. These servers are low in performance but their energy efficiency is more [8].

Dynamic Power Management

Dynamic power consumption depends on some specific usage scenario, clock rate, and input/output activity. It is created by circuit activity like transistor switches, changes of values in registers. These techniques are used to minimize power consumption or used to get better resource usage and application performance.

Dynamic power management (DPM) guidelines the energy consumption by using software rules. Short-circuit current and switched capacitance are the main sources of DPM. Short-circuit current causes only 10–15% of the total power consumption. DPM uses two components, CPU and memory for power regulation [9]. To find the power utilization of processor, power scalable CPUs use the relation between the power supply, operational frequency, and voltage. The dynamic power consumption can be formulated as Eq. (2):

$$p_d = \mathrm{a}fV^2\mathrm{C} \tag{2}$$

whereas

f clock frequency,
a switching activity,
C capacitance, and
V supply voltage

Among all the component used in DPM under hardware level, power scalable memories are the least used resource to minimize energy consumption. Under some specific workload, memory unit can consume 23% of the total power consumption. Mechanisms to reduce energy consumption under software techniques are divided into three categories: proactive, reactive, and hybrid. Reactive management controls the energy consumption according to current situation, feedback, or monitoring. Proactive management is also called predictable management because it uses the average behavior of the system rather than current situation. Hybrid management uses monitoring behavior of reactive and predictive property of proactive [7].

3 Thermal-Aware Scheduling in Cloud Environment

Thermal-aware scheduling is different from the power-aware scheduling. Thermal-aware scheduling performs heat minimization from the active server, while the power-aware scheduling decreases the number of active servers. Power-aware scheduler turns of the idle server and moves all work to active servers which uses maximum power consumption so heat dissipation will be more in small areas which create hotspots if the cooling devices are insufficient and it increases the power usage effectiveness (PUE) of the data centers due to more cooling used at the time of hotspot [10]. The thermal-aware scheduler does not decrease the number of active servers; rather, it minimizes the heat generation so that the chances of hotspot creation are decreased. A thermal-aware scheduler lowers the PUE and Total Cost of Ownership (TCO) as well. The total energy is calculated in two areas: energy consumed on computing and cooling energy in terms of Coefficient of Performance (COP). Thermal-aware techniques were connected to the reduction of heat production, migration of tasks, and heat convection to the neighboring cores. The basic formula for total energy consumption is shown as follows:

Table 1 Comparison of reactive versus proactive scheduling

	Reactive scheduling	Proactive scheduling
Requirement	No need for any training and learning	Training or learning required before start
Performance	Perform better when job having unstable behavior	Less than reactive
Base	Simple	Complex
Time consumption	Less	More
Cost	Low	High
Energy saving	Low	High
Migration of workload possible	Yes	No

$$p_t = p_i + p_c = p_i(1 + 1/\mathrm{COP}(t_s)) \tag{3}$$

where

p_t total energy consumption of data centers,
p_i energy consumed by IT infrastructure,
p_c energy consumed by cooling system,
COP coefficient of performance, and
t_s set temperature of CRAC

3.1 Proactive Scheduling and Reactive Scheduling

Reactive thermal-aware scheduling takes action after problem has occurred and there is no need for any planning or estimation of temperature. But proactive thermal-aware scheduling approach requires prediction and estimation of temperature. In reactive approach, all workload is scheduled according to the current status of thermal statistics [10]. The comparison of both is shown in Table 1.

4 Thermal-Aware Energy Management in Cloud Environment

Thermal awareness is the understanding of temperature distribution inside a data center which is used to increase energy efficiency, cooling and minimize server failure rates. In paper [11], the authors considered the study of different reactive thermal management techniques and provided the trade-off between performance and energy efficiency for high-performance computing workloads. Thermal-aware management is aimed at minimizing the Total Cost of Ownership (TCO) of data centers and

Table 2 Comparison of different heat models

	RC model	Thermal network	Thermodynamic network
Base	It uses resistance and capacitance	Network of devices	Airflow
Difficulty level	Ambient temp is used for cold air temp	Server grading	Heat recirculation matrix
Heat recirculation	No	No	Yes
Is heterogeneity present	Yes	No	No
Is migration possible	No	Yes	No
Power consumption	Direct	Indirect	Direct

environmental impact without any increase in the number of Service Level Agreement (SLA) violations. A typical data center is made with N racks of servers with hot/cold aisle arrangement and floor tiles in the raised floor. All the servers are connected to a high-speed network. The air conditioners referred to as Computer Room Air Conditioning (CRAC) and deliver cold air under elevated floor. From the front side of the rack, cold air enters into the server and heat produced exits from the rear of the racks. The CRAC absorbs hot air from the computer room. The heated air which came from exit place causes hot aisles on the backside of racks [10]. The comparison of different heat models is shown in Table 2.

4.1 RC Model

RC model is used to determine the relationship between the heat transfer and RC circuit. The relationship is given in equation:

$$T_a = R_i C_i \mathrm{dT/dt} + T - \mathrm{R}P_t \tag{4}$$

where

P_t power consumed by processor at time t,
T die temperature,
R_i resistance of the ith system,
C_i capacitance of the ith system,
dT/dt rate of change of temperature of die T in time t. After the expiry of the time period, it goes to the stable state

The value of resistance and capacitance is dependent on the power consumption of the processor and the temperature difference between two areas. The product of R and C is a constant term and remains fixed after the manufacturing of processor packages [10].

4.2 VM Configuration and Migration

VM configuration is used to avoid over-provisioning and under-provisioning by assigning proper amount of resources. Resource over-provisioning means provisioning more resources than required. When resources remain idle, the wastage of resources occur. On the other hand, under-provisioning means provisioning less resources than required and when requirements are not satisfied then performance deterioration occurs [12]. In paper [13], they presented the tenant-based resource allocation model which decides the number of VMs required for a particular workload, and occurrence of performance deterioration due to under-provisioning is minimized and keeps minimum occurrences of over-provisioning resources to avoid resource wastage and to make the model cost-effective. VM migration is used to reduce energy consumption of data centers. If a VM stops their working during or after migration, then efforts for migration are wasted. Migration would be energy-efficient procedure if it performs their work on a capable server. In paper [11], the author focused on minimizing energy efficiency and utilization of data center resources and minimizing undesired thermal behaviors to obtain quality of service. Migration technique is used to reduce hotspot in data centers.

5 Conclusion

The data centers energy consumption is increasing day by day, and it is assumed to be quadruple by 2020. According to data centers viewpoint, two things are of prime concern: (i) minimization of energy consumption and (ii) reduction of CO_2 emission. Therefore, a lot of research work has been carried out to deal with these issues. This survey gives a review of the existing energy-efficient thermal-aware scheduling techniques in cloud system. We classified the existing approaches into energy management and thermal-aware management on the basis of different scheduling and heat model techniques. This survey acts as the groundwork for the future researchers in the area of thermal-aware scheduling in cloud environment.

References

1. Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Warfield, A.: Xen and the art of virtualization. In: ACM SIGOPS Operating Systems Review, vol. 37, no. 5, pp. 164–177. ACM (2003)
2. Grance, T., Mell, P.: The nist definition of cloud computing
3. Mo, C., Dargie, W., Member, S., Schill, A.: Power consumption estimation models for processors. Virtual Mach. Serv. **25**(6), 1600–1614 (2014)
4. Natural Resources Defense Council (NRDC). www.nrdc.org/energy/data-center-efficiency-assessment.as

5. Mastelic, T.: Cloud computing: survey on energy efficiency. ACM Comput. Surv. 47(2) (2015)
6. Beloglazov, A., Buyya, R., Lee, Y.C., Zomaya, A.: A taxonomy and survey of energy-efficient data centers and cloud computing systems, vol. 82 (2011)
7. Sharma, Y., Javadi, B., Si, W., Sun, D.: Reliability and energy efficiency in cloud computing systems: survey and taxonomy. J. Netw. Comput. Appl. **74**, 66–85 (2016)
8. Raj, V.K.M., Shriram, R.: Power management in virtualized datacenter: a survey. J. Netw. Comput. Appl. **69**, 117–133 (2015)
9. Beloglazov, A., Buyya, R., Lee, Y.C., Zomaya, A.: A taxonomy and survey of energy-efficient data centers and cloud computing systems, vol. 82 (2011)
10. Chaudhry, M.T., Ling, T.C., Manzoor, A., Hussain, S.A., Kim, J.: Thermal aware scheduling in green data centers. ACM Comput. Surv. **47**(3), 1–48 (2015)
11. Rodero, I., Viswanathan, H., Lee, E.K., Gamell, M., Pompili, D., Parashar, M.: Energy-efficient thermal-aware autonomic management of virtualized HPC cloud infrastructure. J. Grid Comput. **10**(3) 447–473 (2012)
12. Pietri, I., Sakellariou, R.: Mapping virtual machines onto physical machines in cloud computing: a survey **49**(3) (2016)
13. Espadas, J., Molina, A., Jim´enez, G., Molina, M., Ram´ırez, R., Concha, D.: A tenant-based resource allocation model for scaling software-as-a service applications over cloud computing infrastructures. Future Gener. Comput. Syst. **29**(1), 273–286 (2013)

Eye Gaze Tracking-Based Adaptive E-learning for Enhancing Teaching and Learning in Virtual Classrooms

I. Joe Louis Paul, S. Sasirekha, S. Uma Maheswari, K. A. M. Ajith, S. M. Arjun and S. Athesh Kumar

Abstract Today, the educational system has completely moved to e-learning and virtual classroom-type learning where biometric analysis can be done to put an estimate on the concentration levels of the student. This can be implemented by using an EyeTribe Tracker Pro which uses a camera to track the user's eye movement. The tracking is done so delicately that it tracks even the most minuscule movements of the users' pupils. This data in the form of images is run through a data analysis tool Open Gaze and Mouse Analyzer (OGAMA). The tool reads the "on-screen gaze coordinates" and determines the exact location on the screen where the user is looking. The gaze data is then fed into analysis tools. Various kinds of analysis such as Area of Interest (AOI) and fixation ratio are done. The resulting parameters such as gaze: fixations, fixations duration mean, complete fixation times at AOI, etc., are studied. In this way, the nature of the students' concentration on the on-screen material can be analyzed and the corresponding changes can be introduced.

Keywords Adaptive learning · E-learning · Virtual classroom
Eye gaze tracking

1 Introduction

Conventionally, biometrics has always been associated with security industry for authentication purposes using our fingerprints, facial recognition, voice patterns, etc. However, eye tracking can be used for providing invaluable feedback for teachers to understand if his/her students are paying attention to their class or they need to change their course content. This would be very helpful in particularly when education system has undergone a huge shift to virtual classroom-based

I. Joe Louis Paul (✉) · S. Sasirekha · S. Uma Maheswari · K. A. M. Ajith
S. M. Arjun · S. Athesh Kumar
Department of Information Technology, SSN College of Engineering, Chennai, India
e-mail: joelouisi@ssn.edu.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_17

learning. In fact, companies have been using eye-tracking technology for advertising research for a while now to see how target customers respond to their ads and to determine what captures their attention.

Wedel and Peters [1] reviewed eye-tracking applications in advertising (print, TV, banner), health, and search behaviors. In this work, authors suggest that eye movements are tightly coupled with visual attention which makes them eminent indicators of the covert visual attention process. Research reveals that visual attention is not a gate but reflects higher order cognitive processes and is closer to actual behavior than intuition informs us as suggested by Steinman in [2].

Finally, measuring visual attention is now easy with modern eye-tracking equipment. The EyeTribe Tracker is an eye-tracking system that can calculate the location where a person is looking by means of information extracted from person's face and eyes. The eye gaze coordinates are calculated with respect to a screen to which the person is looking at, and are represented by a pair of (x, y) coordinates given on the screen coordinate system. The user needs to be located within the Tracker's trackbox. The EyeTribe Software Development Kit (SDK) includes a trackbox sample that illustrates how to indicate users their location with respect to the Tracker so that they can adjust their position accordingly. When the system is calibrated, the eye-tracking software calculates the user's eye gaze coordinates with an average accuracy of around 0.5–1° of visual angle. Assuming the user sits approximately 60 cm away from the screen/tracker, this accuracy corresponds to an on-screen average error of 0.5–1 cm.

2 Related Works

This section briefly describes the related works carried out for our proposed study. The summaries of Peters and Wedel [1] provide the foundations of visual attention and eye tracking and a conceptual framework for eye-tracking research in marketing and also reviewing the marketing literature within that framework. Fixations can be used to calculate time spent looking at a particular location, which in turn is thought to reflect engagement of attention and the time needed to process the stimulus at that location. This metric has been used to gain insights into what to remember, how to perform mental computations, how to read, how to solve problems, and how to learn [2, 3]. Miyoshi and Murata illustrated the need for the design and development of an eye-gazing input system by using eye-tracking system. This tracking system consists of two subsystems where the first one detects the eye movement of the users and the second one is used to calculate the mouse pointer coordinates from the eye movement patterns, respectively [4].

Lupu et al. presented their research project in [5] for achieving communication with people with neuromuscular disabilities by using a cost-effective, reliable mobile system based on eye-tracking mouse where the movement of the eye is detected by a head-mounted device and the mouse cursor is moved on the screen accordingly. Majaranta and Bulling addressed the growing interest toward the pervasive attention-aware systems and the need to have good interfaces for human–computer interaction. Here, authors provided the state-of-the-art technologies for eye tracking and gaze estimation. Real-world applications were discussed and detailed how to identify/choose the best tracking system for each application based on the user's requirements [6]. The effectiveness of the gaze interaction within the near-to-eye display was well investigated and developed as the open platform by combining the multisensory electronic glasses with the eye-tracking module by Kocejko et al. in their research work. Here, the advantages and limitations are described for their proposed work [7]. Similar to that, Cheung and Peng addressed eye gaze tracking problem in a desktop environment using a low-cost and more flexible camera. Here, the human face was tracked in a real-time video sequence and the eye regions were extracted. Finally, the eye gaze tracking is accomplished by integrating the eye vector and the extracted head movement information [8]. The various key challenges and opportunities of eye movement technology and guidelines for developing an eye tracking application (i.e., a man to machine interfacing) was discussed in their work by Chandra et al in [9].

In general, the human–computer interface (HCI) uses the conventional input devices such as mouse and keyboard. Salunkhe and Patil proposed a hand-free interface between the human and computer. Authors presented a novel idea to control the mouse cursor using human eyes movement. Here, the mouse movement is controlled based on the eyesight position [10]. Due to the advancements in the communication technologies, there are many number of computing systems available in the market for our day-to-day use. As the number of persons using these devices increases, the demand for faster and non-intrusive methods for accessing these devices also increases. Fatima et al. discussed the human–computer interaction as the promising input medium for communication. Here, authors analyzed the various eye-tracking techniques to find the line of the gaze of the user [11].

Saccades, the rapid eye movements that allow us to shift between fixations, can reflect shifts in attention that are either controlled (e.g., a voluntary eye movement or saccade toward a target) or automatic and stimulus-driven (a reflexive saccade toward a sudden stimulus) [12]. In this regard, Dalmaijer found the spatial precision and accuracy good enough in this work for fixation checking and point of regard analysis after having tested and compared device with higher priced alternatives [13]. Similarly, the same form of analysis can be used to ascertain courses and monitoring individual learning styles. In this work, it has been decided to use an affordable eye tracker called EyeTribe for monitoring and a tool called Open Gaze and Mouse Analyzer (OGAMA) for analysis [14]. OGAMA is an open-source software designed to analyze eye and mouse movements in slideshow study designs recently [15, 16].

3 Proposed Work

Students nowadays prefer to mask their doubts and do not approach the teachers. Staff has no way of knowing whether his material is tuned to the capability of his students, especially in virtual classrooms' scenario. Hence, an EyeTribe device was used and tested for this purpose which monitors the eyes of the viewer. EyeTribe SDK and EyeTribe User Interface (UI) are installed. Actually, the EyeTribe Tracker is set up on a tripod stand which comes with the EyeTribe package. It is then connected to the computer's USB 3.0 port using the USB 3.0 cable. The EyeTribe software is installed, and the server is run. The student eye's calibrations have to be noted down first; that is, his eye movements have to be recorded. Recorded readings are processed and fed into a tool called OGAMA using which the concentration level of students is measured.

A chin rest is used to make sure the EyeTribe readings are not deviated because of the students' head movements. The tutorial or the material that the student is to learn is played on the computer. During the length of the video, the student's eye movements are consistently tracked at every instance. If at any instance the student's eye movements go beyond the initially calibrated eye tribe, it is taken as a lapse of concentration. This gaze data gets stored in the database, and further analysis is done on the same to find out which portions of the video got least attention and if a particular student was not paying attention for more than 80% of the duration of the lecture. This recorded gaze data is then analyzed using a tool called OGAMA with the help of which we put an estimate on the concentration

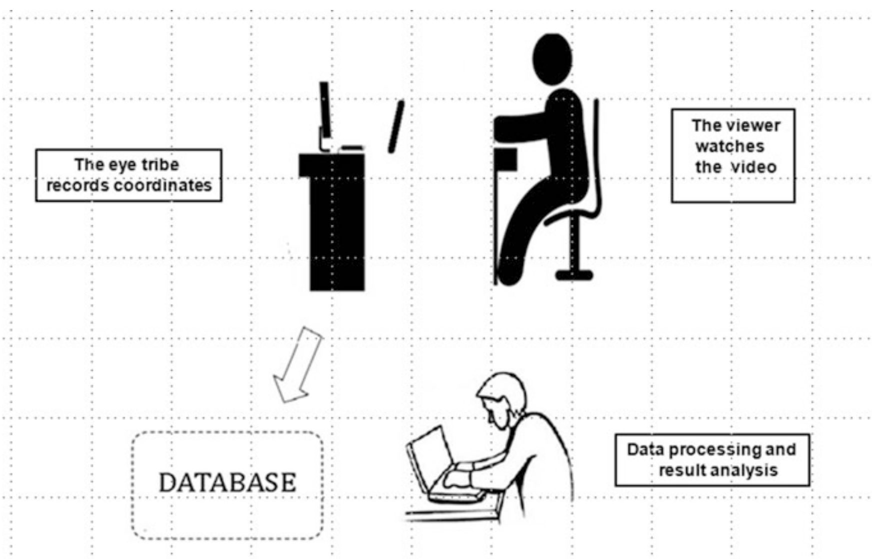


Fig. 1 System model

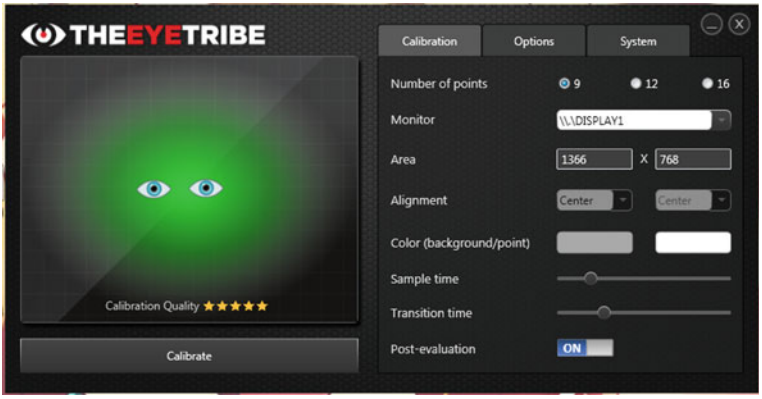
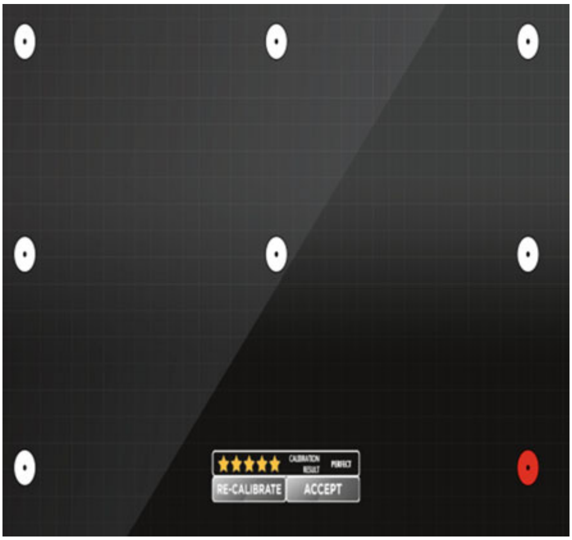


Fig. 2 EyeTribe UI client

detail of the student. After testing many students from various categories and then analyzing their concentration levels, we can come to a conclusion on the effectiveness of the content. Figure 1 illustrates the system model of the proposed work.

The user interface (UI) of the EyeTribe Pro is shown in Fig. 2. Once the EyeTribe server was launched, the initial calibrations are noted as depicted in Fig. 3. Gaze data is taken and recorded in a log file. The Application Programming Interface (API) Console is launched to receive and store the data received from the EyeTribe Server. Various kinds of analysis such as fixation analysis, Area of Interest analysis, statistical analysis, and heat map analysis are done on the recorded data. Over the course of the video after a particular topic, questions are set up to act as a check to ensure the accuracy of the gaze results.

Fig. 3 Device calibration



4 Experimental Results

The first step in this experimental setup is data acquisition during which the lecture video on a topic called Dynamic Host Configuration Protocol (DHCP) is loaded into the OGAMA tool which is used as an analysis tool when interfaced with an eye tracker. This tool can be used to study from the recorded gaze data. The Area of Interest has to be defined before the start of recording the gaze data.

Figure 4 represents the Area of Interest (AOI) analysis, where the target AOI is selected as shown in rectangular box in pink color. Figure 5 displays the fixations made on the targeted area which we have selected. That is, Fig. 5 shows the various

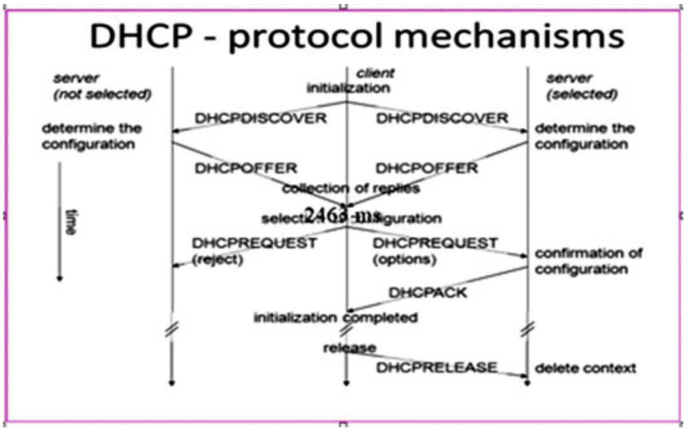


Fig. 4 AOI analysis

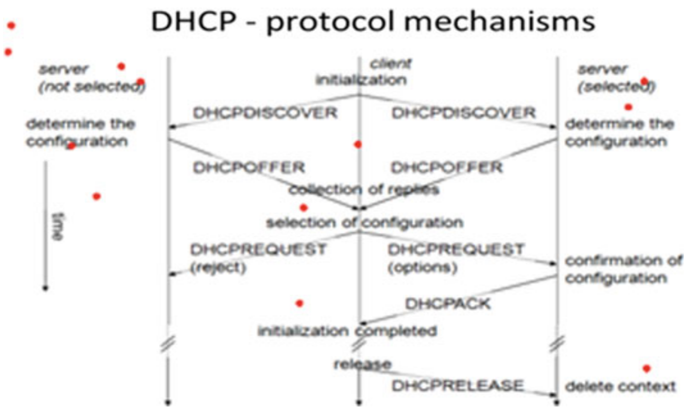


Fig. 5 Fixation map analysis

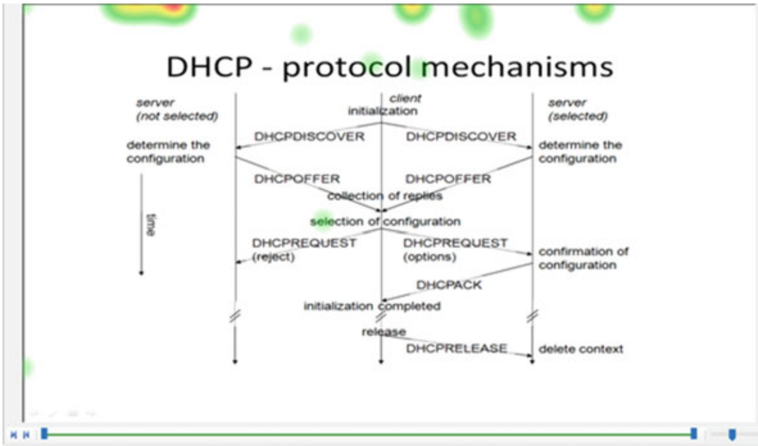


Fig. 6 Attention map analysis

fixation points on the slide, duration of each fixation point, and their coordinates on the screen.

After termination conditions, AOI is set up. Subject information is obtained, and initial calibrations are done. The lecture video is played, and the student is asked to watch the same. After the recording is done, different kinds of analysis are done on the recorded gaze data. Figure 6 shows the attention map analysis shows the different levels of concentration exhibited by each portion of the slide by different persons. Darker shades represent high concentration, lighter shades represent less concentration, and un-shaded areas represent no concentration, respectively.

The gaze statistics are exported to a .csv file. The parameters considered are trial ID, trial duration, fixation duration mean, and complete fixation time. The concentration level of each student is calculated by dividing complete fixation time by trial duration. Figure 7 shows the gaze statistics.

Subject: Name	Subject: Age	Subject: Category	Trial: ID	Trial: Duration (ms)	Gaze: Fixation Duration Mean (ms)	Gaze: Complete fixation time	Concentration in %
Ajith	21	IT student	1	30342	420.2045455	16956	55.88293455
ambigai	21	IT student	2	30376	281.45	4496	14.80115881
Arjun	21	student	3	30442	276.5	2765	9.082846068
arjuns	21	IT student	4	30341	266.3125	2465	8.124320227
ashok	20	it stud	5	30475	688.3333333	24780	81.31255127
Athesh	21	student	6	30276	1179.857143	8259	27.2790329
durgaharika	21	IT student	7	30508	606.5	8291	27.1764783
Ganesh	21	IT student	8	30241	758.9583333	12988	42.9483152
Jayakanthan	21	student	9	30342	311.2941176	765	2.521257663
Logith	22	student	10	30376	289.3846154	1297	4.269818278
Rajadurai	21	it stud	11	30208	399.3333333	3027	10.02052436
Sathish	21	eee stud	12	30376	504.68	12617	41.53608112
Thamil	21		13	30376	563.4054054	20680	68.08006321

Fig. 7 Gaze statistics

5 Interpretation

We have a focus group of around ten students and asked them to learn a topic DHCP. The material has five slides, and we have observed that uniformly all students experienced concentration lapses in slide #4. We can see clearly that concentration in the second slide is uniformly low for both samples from the attention map shown in Fig. 8. Similarly, many other samples also showed the same results.

Figure 9 shows the respective concentration levels of the students.

So, we have decided to modify the lecture content and add certain prerequisite material to help the students understand the slide/diagram better. The attention map for the same is shown in Fig. 10.

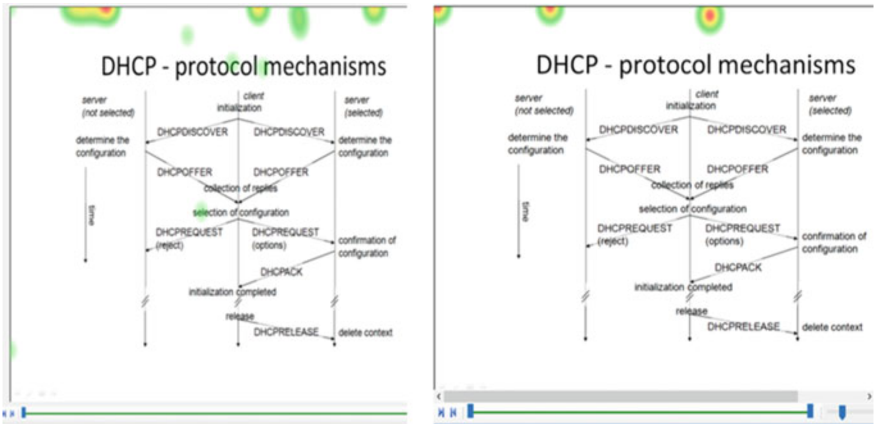


Fig. 8 Attention map of slide #4 on DHCP

Subject: Name	Trial: ID	Trial: Duration (ms)	Gaze: Fixations (count)	Gaze: Number of fixations at AOI:SAMPLE	Gaze: Complete fixation time at AOI:SAMPLE (ms)	Concentration
Ajith	1	20117	17	4	965	4.80%
Aravind	1	20082	13	12	3095	15.41%
Arjun	1	20082	29	8	1567	7.80%
Aasif	1	20051	13	5	1301	6.49%
Ligith	1	20051	18	10	2834	14.13%
Ashok	1	20117	20	20	4935	24.53%
Ganesh	1	20083	7	2	966	4.81%
Jeevana	1	20117	15	14	3396	16.88%

Fig. 9 Initial concentration detail

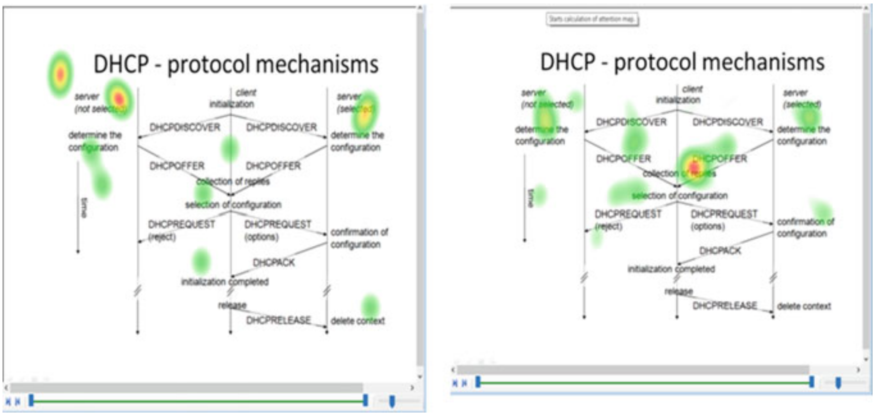


Fig. 10 Attention map of slide #4 on DHCP after including prerequisite

Subject: Name	Trial: ID	Trial: Duration (ms)	Gaze: Fixations (count)	Gaze: Complete fixation time at AOI:Sample (ms)	Gaze: Number of fixations at AOI:Sample	Concentration
Aasif2	3	20118	13	18717	13	93.04%
Ajith2	3	20084	20	17085	20	85.07%
Arjun1	3	20083	29	17419	29	86.74%
Jeevana1	3	20083	29	16524	29	82.28%
Ligith	3	20116	20	16255	20	80.81%

Fig. 11 Final concentration detail

We did the study again, and these times, the concentration levels improved by a huge margin. Figure 11 shows the concentration levels after adding prerequisite, respectively.

We have used three types of videos to study the concentration levels of the samples—a lecture video on HTML, a sports video, and a dance video. The interest of each person varies and that was reflected in the amount of their concentration levels. Generally, figures, highlighted parts, and keynotes attract more attention and we could see that from the fixation points. In order to confirm the findings of the tool, we used a quiz at the end of the lecture to see if the students were really paying attention.

To check the accuracy of the results obtained, a test was designed and added to the slideshow after the lecture video to confirm the findings. This is the significance of our proposed work. Figures 12 and 13 represent the test provided to the subjects and the corresponding results obtained, respectively.

The responses of the subjects were analyzed and sent to the faculty along with the concentration details. Hence, using a combination of EyeTribe and OGAMA,

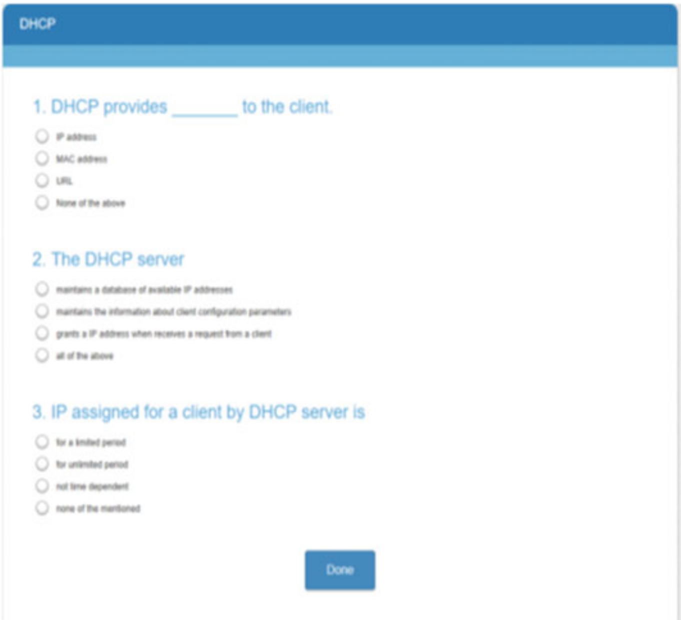


Fig. 12 Setting up a test

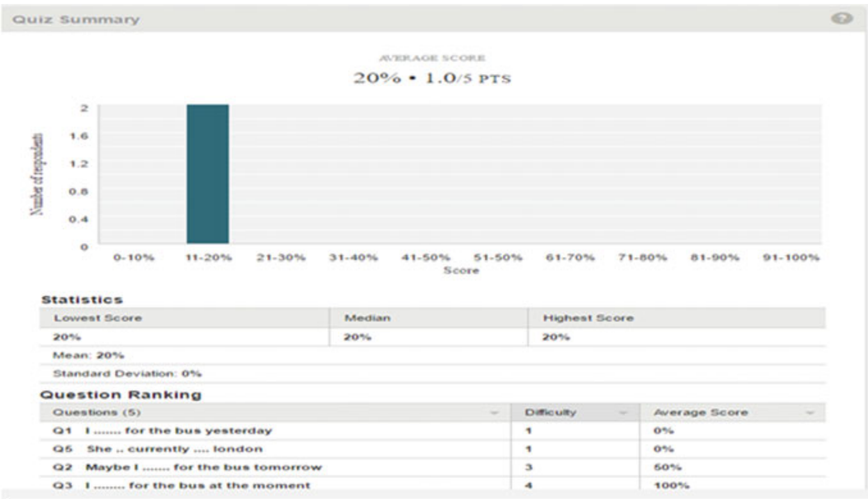


Fig. 13 Analysis of the test

we could clearly identify the concentration lapses of the subjects and which portions of the video got the least attention. The quiz at the end also helped to verify the same.

The experiments were undertaken with the understanding and appropriate informed consent of each individual involved in this study.

6 Conclusion

When used in tandem with traditional teaching, the proposed eye gaze tracking-based adaptive e-learning system can make the work of teachers and students easier, thus making learning a lot smoother in virtual classrooms. The efficiency of learning and reach of the course material to the target audience can be improved. Our system can also be used in a business perspective helping companies measure the attractiveness of advertisements to potential customers. Eye tracking as a research field is still in its nascent stage, and EyeTribe is the first affordable tracker of its kind, but it suffers from slight precision issues when compared to other high-end alternatives.

References

1. Wedel, M., Pieter, R.: A review of eye-tracking research in marketing. Rev. Market. Res. **4**, 123–147 (2008)
2. Steinman, L.: Elaborate interactions between the immune and nervous systems. Nat. Immunol. **5**, 575–581 (2004)
3. Hannula, D.E., Althoff, R.R., Warren, D.E., Riggs, L., Cohen, N.J., Ryan, J.D.: Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. Front. Human Neurosci. **4**(166), 1–16 (2010)
4. Miyoshi, T., Murata, A.: Input device using eye tracker in human-computer interaction. In: 10th IEEE International Workshop on Robot and Human Interactive Communication 2001, pp. 18–21, IEEE (2001)
5. Lupu, R.G., Ungureanu, F., Siriteanu, V.: Eye tracking mouse for human computer interaction. In: E-Health and Bioengineering Conference (EHB) (2013)
6. Majaranta, P., Bulling, A.: Eye tracking and eye-based human–computer interaction. In: Advances in Physiological Computing, pp. 39–65 (2014)
7. Kocejko, T., Ruminski, J., Wtorek, J., Martin, B.: Eye tracking within near-to-eye display. In: 8th International Conference on Human System Interactions (HSI) 2015, pp. 166–172 (2015)
8. Cheung, Y., Peng, O.: Eye gaze tracking with a web camera in a desktop environment. IEEE Trans. Human-Mach. Syst. **45**(4), 419–430 (2015)
9. Chandra, S., Sharma, G., Malhotra, S., Jha, D., Mittal, A.P.: Eye tracking based human computer interaction: applications and their uses. In: International Conference on Man and Machine Interfacing (MAMI) (2015)
10. Salunkhe, P., Patil, A.R.: A device controlled using eye movement. In: International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) 2016, pp. 732–735 (2016)

11. Fatima, R., Usmani, A., Zaheer, Z.: Eye movement based human computer interaction. In: 3rd International Conference on Recent Advances in Information Technology (RAIT) 2016, pp. 489–494 (2016)
12. Luna, B., Velanova, K., Geier, C.F.: Development of eye-movement control. Brain Cogn. **68** (3), 293–308 (2008)
13. Dalmaijer, E.: Is the low-cost EyeTribe eye tracker any good for research? PeerJ Pre-Prints 2: e585v1, https://doi.org/10.7287/peerj.preprints.585v1 (2014)
14. http://www.hongkiat.com/blog/future-classroom-technologies/
15. Open Gaze and Mouse Analyzer (OGAMA), Adrian Vobkuhler
16. http://www.ogama.net/sites/default/files/pdf/OGAMA-DescriptionV25.pdf

A Novel Text-Based User Authentication Scheme Using Pseudo-dynamic Password

Ramsha Fatima, Nadia Siddiqui, M. Sarosh Umar and M. H. Khan

Abstract With the advancement in technology, the number of attacks taking place on a user's identity is increasing. Thus, there is a need for an increase in the security of the user's identity. Authentication of a user is the most important element of security. In this paper, we have proposed a novel method for user authentication. This method is a challenge–response-based user authentication technique that will provide a higher level of security than conventional text-based passwords. In this method, the password entered at the time of authentication will be different for each session, even when the actual user password is same. Thus, it is a pseudo-dynamic scheme of password generation. In this paper, we have also analyzed the factors on which the password computation time depends. We have also analyzed its susceptibility to the various attacks that are common on the text-based passwords.

Keywords Authentication · Text-based password · Attacks
Pseudo-dynamic password · Password strength · Random number
Round robin method

R. Fatima (✉) · N. Siddiqui · M. Sarosh Umar · M. H. Khan
Department of Computer Engineering, Zakir Husain College of Engineering
and Technology, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India
e-mail: ramsha.fatima94@gmail.com

N. Siddiqui
e-mail: nadiasid2134@gmail.com

M. Sarosh Umar
e-mail: saroshumar@zhcet.ac.in

M. H. Khan
e-mail: muneebhkhan@zhcet.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_18

1 Introduction

As the Internet is expanding, the number of services provided to the user online is also increasing. Thus, there is a need for a system that will provide users with high security, privacy, and reliability to carry out all the transactions. The oldest method of securing anything is through the use of passwords. A password is a string of characters (including special characters and numerals) or a word used for user authentication. It is used to prove identity of a user or to gain access to a resource, which is accessible to only specific users [1]. This password-based scheme has many vulnerabilities. The most common of these vulnerabilities is the difficulty in remembering the password, if a password is easy to remember, that means it will generally be easier for the attacker to guess [1, 2]. However, the use of difficult passwords may not increase the security of the system as the user may need to write down or store the password or may reuse the same password leading to an ease for the attacker to crack it.

In the method proposed in this paper, we have tried to eliminate some of the disadvantages inherent in the text-based user authentication. We have developed a challenge-based password authentication scheme that generates pseudo-passwords and will provide the user an enhanced security even with the use of easier passwords. This scheme provides a higher security even when there is a password reuse.

2 Background

2.1 Authentication Techniques

For authenticating a user for access to a resource, many techniques can be used. These techniques are divided into three basic types. These are as follows.

Token-Based Authentication. A token-based authentication methodology uses small hardware devices like key cards, bank cards, and smart cards [3] to provide authorization to the user. This authentication technique has an additional level of security using two-factor authentication [4]: A personal identification number (PIN) authorizes the user as the owner of particular device, and then the unique device identifies the user to the service and allows them to log in.

Biometric-Based Authentication. In biometric authentication, it is assumed that every individual has unique physical or behavioral characteristics, which can be used for identification. There are mainly two classes of biometric identifiers: physiological characteristics, which include fingerprints, DNA, face, hand, retina [2, 3], while the second one is behavioral characteristics which include the behavioral pattern of a human being like gestures, typing rhythm, stepping, and voice.

Knowledge-Based Authentication. Knowledge-based authentication (KBA) is a security measure that asks answer to specific security questions [5] to provide accurate authorization. There are two main types of KBA security measures: static KBA [5] in which the users themselves specify the security questions and their answers at the time of registration, and dynamic KBA [5] that gathers data about a user and then uses data mining to ask questions that the systems already know the answer to.

2.2 Text-Based Used Authentication

The text-based password is a subtype of KBA. In this, the user has to enter a string, which is a sequence of alphabets, numbers, and special characters. This string is to be entered every time the user wants to access this resource. This type of user authentication is susceptible to different types of attacks [3, 6]; these are as follows.

Brute Force Attack. In this type of attack, an algorithm is used that checks every combination of characters and numbers to determine the password. The smaller the password length, the easier it is to break the password [7, 8].

Dictionary Attack. **In** this type of attack, every possible word in the dictionary is checked for password. If the password is a valid English word, then it can be cracked by using the dictionary attack [7, 8].

Spyware Attack. In this type of attack, a spyware is installed in the device. This spyware records the sequence of the keys pressed on the keyboard. Thus, it is able to crack the password [8].

Shoulder Surfing Attack. Shoulder surfing can also be carried out to determine a user's password. In this, the attacker can encounter the password of a person by looking over the person's shoulder when they are entering their password [8].

3 Novel Method of Text-Based User Authentication

In this section, we will describe the user authentication method that we have developed. It is a challenge–response-type text-based user authentication technique. Here, user selects a challenge at the time of registration according to which the password is provided during login.

3.1 Registration

During registration, i.e., when the user is making a new account (Fig. 1), the user will have to select a user id. The user will have to enter a password, and the password need not be a difficult combination of words. Along with the username and password, the user will also have to select an operator (like '+', '−', '*', '/') that will be used to calculate the result of all the random numbers displayed at the time of authentication. The user will also have to enter a method from the list of

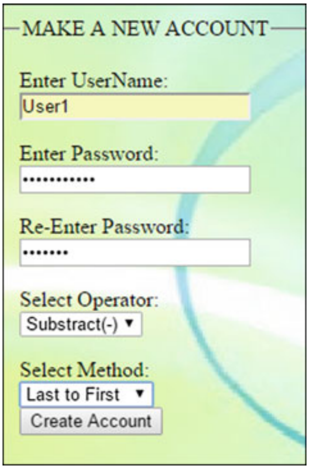


Fig. 1 Making a new account

methods displayed that will determine how the password selected will be entered at the time of authentication.

During the time of registration, the user will also be shown the list of methods available to make the password dynamic. These methods will be available with their descriptions and an example to help the user in understanding how the method is to be used. The methods used in the implementation of this paper will be described in the next section.

3.2 Login

During the time of authentication, user will have to enter their username (Fig. 2). This username will be checked with the entries present in the database; if the username is present in the database, the user will be directed to the second page that will display a list of random numbers (Fig. 3). The user will apply the operator selected during the time of registration on the list of random numbers displayed to get the result. This result along with the list of random numbers generated will be added to the password selected by the user according to the method specified at the time of registration. The

Fig. 2 Enter username during authentication

Fig. 3 Enter password using the random numbers generated and the method selected

result generated with this calculation will be entered in the space provided for the password. If the password matches, the user will be granted access.

Some methods used to make the password dynamic are as follows.

Round Robin Method. In this method, the result calculated by applying the operator selected on the random numbers is added to the first letter/digit of the password, and the random numbers are then added to the rest of the password in a circular fashion.

Add Result Method. In this method, the result is calculated by applying the operator selected during registration on the set of random numbers generated. This result is then added to every letter/digit of the password.

Add First Random Number. In this method, there is no need for the calculation of result. Here, first random number displayed during the time of login will be added to each letter/digit of the password to make it dynamic.

Add Last Random Number. This method also does not require the calculation of the result by applying the operator. In this method, the last random number shown during the time of login will be added to every letter/digit of the password.

3.3 An Example

Taking an example, suppose during the registration process, the user enters 'User1' as the username and 'password' as the password. Also, suppose that the operator selected is '+' and the method selected is 'round robin.'

Now, during the time of login, let the random numbers generated be '1, 2, 3, 4'. Then, on applying the addition operator selected during registration, on these random numbers, we will get the result as: 1 + 2 + 3 + 4 = 10.

So, using the round robin method of dynamic password generation, the password to be entered will be:

$$(\mathrm{p}+10)(\mathrm{a}+1)(\mathrm{s}+2)(\mathrm{s}+3)(\mathrm{w}+4)(\mathrm{o}+1)(\mathrm{r}+2)(\mathrm{d}+3) = \text{`zbuvaptg}'$$

4 Method Analysis

In this section, we will analyze the method we have proposed for its effectiveness and usability.

4.1 Password Strength

'Password strength is a measure of the effectiveness of the password in resisting guessing and brute force attacks. The strength of a password is a function of length, complexity, and unpredictability' [9]. The password space is also a measure of the effectiveness of the password. Calculating the password space of the proposed scheme [10]:

A: number of alphabet symbols (i.e., the total number of symbols—including alphabets, numbers, and special characters—that can be used in the password).

M: password length (length of the password entered by the user).

S: password space.

Then,

$$S = A^M$$

A = 26 (small alphabets) + 26 (capital alphabets) + 10 (digits) + 34 (special characters)

$$A = 96$$

Taking the example given in the previous section, we have password as 'password', so the password length is 8, i.e.,

$$M = 8$$
$$\Rightarrow S = 96^8$$
$$S = 7.2 * 10^{15}$$

So, by the use of brute force approach the probability of getting a correct result will be

$$P = 1/(7.2 * 10^{15})$$

4.2 Password Computation Time

The time required to compute the password in this method depends on many factors [11]. These are as follows.

Method of Making Password Dynamic. The computation time for password depends on the method selected at the time of registration for making the password dynamic. This can be concluded from the table shown (Table 1). In this table, we have taken the password and the operator as same for each instance.

From this table, it can be shown that those methods in which result has to be calculated take more time for password evaluation as compared to the methods in which the result is not needed (like first random number and last random number method).

Password Length. The computation time for password also depends on the length of the password selected at the time of registration. This can be concluded from the table shown (Table 2). In this table, we have taken the method and the operator as same for each instance and varied the password length.

From this table, we can determine that the evaluation time for password increases with the increase in password length.

Operator. The computation time for password is also dependent on the operator selected for the calculation of result at the time of registration. The evaluation time for calculation of the result with operators '+' and '−' is less as compared to evaluation time for the calculation of the result with operator '*' and '/'. Even in '*' and '/', the evaluation time for '/' is greater than that for '*'. Because of these conditions, the evaluation time for password is dependent on the operator when the method selected during the time of registration requires the use of result.

Table 1 Time taken with different methods for entering the same password

Method	Password	Computation time
Round robin with result	Abc	14
Add result of random numbers	Abc	15
Add first random number	Abc	7
Add last random number	Abc	8

Table 2 Dependency of the password computation time on password length

Method	Password	Computation time
Add first random number	Abc	4
Add first random number	Nadia	22
Add first random number	Ramshaf	37
Add first random number	Ramshanadia	60

Training. The computation time for the password also depends on the training of the user with the particular technique. If the user continues to use a particular method for any length of time, then time required by him/her for the evaluation of the password will get decreased as the user will become accustomed with that method (Table 3).

Result to be Added. The time taken to compute the final password will also depend on the final result that has to be added to the password. If the number to be added is small, then the final password will be written in a lesser time as compared to when the number to be added is large (Table 4).

4.3 *Susceptibility to Attack*

The pseudo-dynamic method of password generation is less susceptible to the attacks that plague the text-based password authentication methods.

Brute Force Attack. The susceptibility to the brute force type of attack can be reduced by limiting the number of wrong passwords that can be entered.

Dictionary Attack. The pseudo-dynamic method of password generation cannot be broken through the use of dictionary attack. This is because the final password that will be entered will be a meaningless combination of numbers and words, even when the original password is a word present in the dictionary.

Spyware Attack. As the numbers generated are random, so the password that will be entered will be different for every session. Thus, even knowing the sequence of key presses is not helpful for breaking the password.

Table 3 Computation time dependency on user training

Method	Result to be added	Computation time
Add first random number	3	33
Add first random number	3	26
Add first random number	3	22
Add first random number	3	23

Table 4 Dependence of computation time on the result to be added

Method	Result to be added	Computation time
Add first random number	9	13
Add first random number	7	10
Add first random number	5	8
Add first random number	3	6
Add first random number	2	5

Shoulder Surfing Attack. A user can choose from a number of methods for making the password dynamic; thus even if the attacker sees the random numbers generated and the password entered, it will not be easy to find out the original password.

5 Future Work

The implementation of this technique will provide a higher security to any resource than the conventional text-based password. It is meant for protecting the confidential information like emails, data, accounts, files. Some enhancements that can be carried out in the future in this technique are:

- The current methods to make the password dynamic are round robin, add result, add first random number, and add last random number. A new method of making the password dynamic can also be developed in the future.
- The security of this scheme can be enhanced by providing the random numbers generated as CAPTCHA so that it is not possible for an attacker to attack with the help of any system.
- A database can be maintained with each user that will store the time they have taken to login in previous attempts. If the current time is similar to the previous times, the user is provided access, but if the current time deviates exceptionally with the previous times, that access is not provided to the user.

6 Conclusion

In this paper, we have implemented a novel method for text-based user authentication. As the number of services provided online increases, so do the attacks on the user's identity. Thus, to safeguard the user against identity theft, we have developed a method that uses the pseudo-dynamic text-based passwords to provide the users with a higher security than the conventional text-based passwords. The time taken by this method for logging into a system is more than the time taken by the conventional passwords, but it decreases as the user becomes accustomed to the method. In this paper, only a few methods are shown to make the passwords dynamic, but more methods can be added. The larger the number of methods, the more difficult it will be to hack into the system.

References

1. Password. https://en.wikipedia.org/wiki/Password. Accessed on 26 Nov 2016
2. Kushwaha, B.K.: An approach for user authentication one time password (numeric and graphical) scheme. J. Global Res. Comput. Sci. **3**(11), 54–57 (2012)
3. Bhanushali, A., Mange, B., Vyas, H., Bhanushali, H., Bhogle, P.: Comparison of graphical password authentication techniques. Int. J. Comput. Appl. (0975–8887) **116**(1), 11–14 (2015)
4. Vaithyasubramanian, S., Christy, A., Saravanan, D.: Two factor authentications for secured login in support of effective information preservation and network security. ARPN J. Eng. Appl. Sci. **10**(5), 313–318 (2015)
5. Knowledge Based Authentication. http://searchsecurity.techtarget.com/definition/knowledge-based-authentication. Accessed on 12 Nov 2017
6. Giri, M., Singh, D.: Theoretical analysis of user authentication systems. Int. J. Innov. Res. Dev. **2**(12 [Special Issue]), 1825–1840 (2013)
7. Towhidi, F., Manaf, A.A., Salwani, M.D., Lashkari, A.H.: Knowledge based authentication attacks. In: Proceedings of the 2011 International Conference on Security & Management. CSREA Press, pp. 1–5, USA (2011)
8. Charathsandran, G.: Text Password Survey: Transition from First Generation to Second Generation (Unpublished)
9. Password strength. https://en.m.wikipedia.org/wiki/Password_strength. Accessed on 12 Nov 2017
10. Saeed, S., Umar, M.S.: A hybrid graphical user authentication scheme. In: 2015 Communication, Control and Intelligent Systems (CCIS), Mathura, pp. 411–415 (2015)
11. Saeed, S., Sarosh Umar, M.: PassNeighbor: a shoulder surfing resistant scheme. In: 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 797–802 (2016)

Concurrency Control Algorithms for Translation Lookaside Buffer

Manisha Agarwal and Manisha Jailia

Abstract A multiprocessor which shares the memory among processors and uses multiples translation lookaside buffers (TLBs) can face various problems. One such problem is the problem of inconsistency which may occur when the page table entry (PTE) is updated because of the multiple copies of same page table entry in various TLBs. Commonly, the inconsistency problem exists in virtually tagged caches, which keep page table entry information, like reference bit, dirty bit and protection bit, in every cache line (Agarwal et al Inconsistency in translation lookaside buffer, 2016 [1]). This paper presents concurrency control algorithms for translation lookaside buffer. We focus on algorithms that reduce the access rights to a page without causing the TLB inconsistency problem.

Keywords Concurrency control · Translation lookaside buffer
Page table entry · Cache · Inconsistency

1 Introduction

The shared memory multiprocessor system is a system which contains multiple (N) processors, and the main memory is shared among those N processors. Any such processors can access any location of main memory, and all N processor can run in parallel.

In such types of multiprocessing systems which shares memory, a page table entry, at the same time, can be copied to multiple translation lookaside buffers. Whenever any of the copied entry is changed, the problem of inconsistency may occur [2]. Inconsistency means the entries in the page table, and TLB may differ.

M. Agarwal (✉) · M. Jailia
Banasthali Vidyapith, Vanasthali, Rajasthan, India
e-mail: manishaagarwal18@yahoo.co.in

M. Jailia
e-mail: manishajailia@yahoo.co.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_19

Whenever the page table entry is not consistent with translation lookaside buffer entry, following problems may occur:

(a) Faulty or wrong results may be produced by memory references. For example, the page table entry allows a read-only access to a page, but the translation lookaside buffer allows read-write access to the same page.
(b) The status bit (reference bit and dirty bit) may not set properly by the processor. For example, in page table, the dirty bit value is set to one, but in translation lookaside buffer, it is zero.
(c) The new page table entries can be lost. That is, if status bits are not set atomically, then a recently updated page table entry may be overwritten with a previous TLB entry by the processor.

All these problems can be handled by flushing the translation lookaside buffer entries after every page table update. But it is simple when the page table entries are cached in single translation lookaside buffer at a time. Because of process synchronization, it is hard to flush multiple translation lookaside buffers [3]. The updating of page table is further complicated by overwrite problem (problem c).

Now, to handle all these inconsistencies between PTE and TLB and to update both (TLB and PTE) concurrently, various types of concurrency control algorithms can be used.

These algorithms may vary in nature of TLBs, i.e. for what types of TLBs they can be used and how consistency implemented by algorithms.

1.1 Consistency Implementation

Consistency can be implemented in two ways,

(a) **While execution**
(b) **After execution**

For concurrency control, the discussed algorithms reduce the rights to access a page table entry. This reduction is done in a consistent way, and this consistency can be implementing either of two ways.

First, become consistent while execution, that is the page will be accessed only with reduced rights as soon as the algorithm changes access rights (that is anywhere before the completion of algorithm).

Second, become consistent only after the completion of the algorithm, that means a page access rights are reduced only after the completion of algorithm not after the rights actually reduced by algorithm (that is the right reduction is not effective between reduction done by algorithm and the completion of algorithm)

1.2 TLB Types

Whenever a processor wants to access a memory location, then that memory reference can be carried out as follows. It first loads the corresponding PTE into its TLB, if it is not already in the TLB [4]. It then translates the virtual address using a TLB entry. The reference is rejected if the virtual page is not mapped to a physical page, or the type of access is not allowed. A rejected memory reference generates a fault, a software exception handled by the operating system. If the reference is accepted, the processor accesses the data and, when necessary, adjusts the referenced and dirty status in the TLB entry and the PTE [5].

Now, how TLB is loaded will depend upon the type of TLB. The various types can be [6]:

1. A PTE may be loaded into a TLB by hardware,
2. A PTE may be loaded into a TLB by software (via a trap),
3. The status bit is to be set atomically,
4. The status bit is not being set atomically.

2 Concurrency Control Algorithms

Whenever a processor wants to update a page table entry, two types of operations are performed; first, the page table entry will be changed, and second, the corresponding TLB entries are flushed. The order of execution of these operations is an important issue. If PTE is changed first, the newly updated PTE may be overwritten, and if TLB flushing is done first, a processor may reload the old PTE into its TLB after having flushed the TLB [7].

This paper discusses four types of concurrency control algorithms. These algorithms have different order to execute above-mentioned operations.

(1) Change and Flush (CF): This algorithm changes the PTE first and then flushes the TLB; it works only for machines that set status bits atomically.
(2) Flush and Change (FC): This algorithm flushes the TLB entries first and then changes PTE; it works only for machines that load TLB entries via software.
(3) 2-Phase (2P): This algorithm works for machines that have neither of the above properties.
(4) Optimistic (Opt): This algorithm works for machines that have neither of the above properties.

All four algorithms assume that TLBs are initially consistent. That is when an algorithm starts, TLBs are not in a state that may cause non-recoverable errors. This assumption holds if the page table is always updated using these algorithms (Table 1).

Table 1 Summary of concurrency control algorithms

Algorithms	TLB characteristics	Consistency implementation
FC	Status bits set atomically	After execution
CF	TLBs loaded by software	While execution
2P	Any	While execution
Opt	Any	After execution

Following notations are followed in the description of algorithms:

PTE: The PTE to be updated

new _ PTE: The target value to be written to PTE.

old _ PTE: The old value of PTE.

hot _ TLBEs: TLB entries corresponding to PTE.

Requester: The processor that starts the algorithm. Usually, it sends interprocessor requests to other processors.

Replier: A processor that responds to an interprocessor request.

target _ set: The set of processors that contain hot _ TLBEs.

2.1 CF (Change and Flush) Algorithm

CF algorithm updates the PTE first and then flushes the corresponding TLB entries. It works only if the newly updated PTE can never be overwritten as a result of setting status bits. This condition holds if the status bits are set atomically. This algorithm implements after execution consistency, because TLB entries may still contain non-reduced access rights after the PTE has been changed.

```
Requester ( )
{
        PTE = new _ PTE ;
        if (requester in target _ set )
            flush hot _ TLBs locally ;
        for each processor in (target _ set - requester)
            send an interprocessor request ;
        for each processor in (target _ set - requester)
            wait for an acknowledgement ;
}
Replier ( )
{
        flush hot _ TLBs locally ;
        acknowledge ;
}
```

2.2 FC (Flush and Change) Algorithm

FC Algorithm flushes TLB entries before updating the PTE. It works only if TLB entries are loaded via software traps. Otherwise, the hardware may load old _ PTE into a TLB after the algorithm has flushed it. It implements while execution consistency, because the PTE is changed after all old TLB entries have been flushed.

```
Requester ( )
{
        changing _ PTE _ flag =true ;
        tentative _ PTE = new _ PTE ;
        if (requester in target _ set )
                flush hot _ TLBs locally ;
        for each processor in (target _ set - requester)
                send an interprocessor request ;
        for each processor in (target _ set - requester)
                wait for an acknowledgement ;
        PTE - new PTE ;
        changing _ PTE _ flag = false ;
}
Replier ( )
{
        flush hot _ TLBs locally ;
        acknowledge ;
}
Load _TLB _ trap_ handler ( )
{
        if (changing _PTE_flag)
                load tentative _ PTE into TLB ;
        else
                load PTE into TLB ;
}
```

2.3 2P (2-Phase) Algorithm

2P algorithm works for any combination of TLB parameters and Implements while execution consistency [8]. To ensure that the newly updated PTE will not be overwritten as a result of setting status bits, the algorithm stalls all processor that may access the page. In the first phase, the algorithm flushes all hot _ TLBEs and stops all active processors except the requester itself. In the second phase, it updates the PTE and then resumes all stopped processors. Both the requester and replier have to block. Furthermore, they have to wait for the slowest processor. The blocking time could be long, especially when there is a replier that disables interrupts.

```
Requester ( )
{
        if (requester in target _ set)
                flush hot _ TLBEs locally ;
        for each processor in (target _ set - requester)
                send an interprocessor request ;
for each processor in (target _set - starter)
        wait for an acknowledgement ;
PTE = new _PTE ;
for each processor in (target _set - starter)
        send a continuing signal ;
}
Replier ( )
{
        acknowledge ;
        wait for the contnuing signal ;
        flush hot _ TLBEs locally ;
}
```

2.4 Opt (Optimistic) Algorithm

Opt algorithm is intended to avoid the blocking of replier. It proceeds in the same way as CF algorithm, even when status bits are not set atomically. Therefore, the newly updated PTE may be overwritten as a result of setting status bits. However, the algorithm can detect this and, when necessary, repeat the updating. It implements after execution consistency, which allows the PTE to be overwritten while the algorithm is in progress.

The idea of this algorithm is similar to that of optimistic currency control in database systems [9]. It does not employ a complex mechanism, as in 2P algorithm, for a problem that happens only rarely. Instead, it proceeds as if the problem does not exist and fixes it when it does happen. Its performance is close to that of CF algorithm if the probability that a PTE is overwritten is low.

The requester code shown below contains a loop, which may make the worst-case time unbounded. This loop can be easily removed, if desirable, by executing the 2P algorithm when the test fails. However, this has little effect on overall performance if the test fails rarely.

```
Requester ( )
{
   do {
 PTE = new _PTE ;
 if (requester in target _ set)
            flush hot _ TLBEs locally ;
 for each processor in (target _ set - requester)
            send an interprocessor request ;
 for each processor in (target _ set - requester)
            wait for an acknowledgement ;
} while (PTE. access _ rights ! = new _PTE. access _ rights) ;
}
Repliers ( )
{
            flush hot _ TLBEs locally ;
            acknowledge ;
}
```

3 Discussion

The section discusses some details that we omitted in presenting the above algorithms.

3.1 Disabling Interrupts

Interrupts, when not disabled, may affect the correctness of a TLB concurrency control algorithm. An interrupt handler may access the target page in the middle of TLB concurrency control algorithm. This may affect the PTE and the corresponding TLB synchronization entries, which are either being updated or being flushed [10]. Interrupts, if not disabled, affect the correctness only of the 2P algorithm among the four algorithms described above. In addition to correctness, handling an interrupt slows down a processor [11], possibly delaying every processor involved in a TLB concurrency control algorithm.

The key idea of algorithm 2P is to stop accessing the target page on all processors until the TLB entries have been flushed on that processor, and the PTE has been updated. Therefore, the algorithm must disable any interrupt that may cause the target page to be accessed.

CF Algorithm is used when status bits are set atomically. Therefore, accessing the target page cannot destroy the newly updated PTE. Furthermore, accessing the page may load only new _ PTE into a TLB because the PTE has been updated at the beginning. Hence, interrupts are not a problem for them.

FC Algorithm is used when TLB entries are loaded by software. Hence, accessing the target page may load only new _ PTE into a TLB. The PTE is updated when no TLB's contains old _ PTE, so the new PTE cannot be overwritten by old _ PTE due to accessing the target page. Again, accessing the target page in the middle of these algorithms causes no problems.

In Opt algorithm, servicing an interrupt may overwrite the newly updated PTE, or load old _ PTE into a TLB if the PTE has already been overwritten. However, these do not violate the after execution consistency and can be detected and fixed before the algorithm terminates.

3.2 Interprocessor Request and Batching

An interprocessor request is carried out in two steps: preparing the request in main memory and optionally notifying the replier. The first step is typically done by enquiring the request to a per-processor queue. The second step is done by issuing an interprocessor interrupt. Acknowledging a request is easier. The replier sets a shared variable, and the requester spins on it.

An interprocessor interrupt is necessary only when the response time of the request is important, for example, when the requester blocks for the acknowledgement of the request. It is potentially expensive because the interrupt may not be handled immediately if the replier has interrupts masked. This will make the requester wait longer for an acknowledgement and increase the total overhead [12].

On the other hand, a processor does not have to be notified for every request. It can check its request queue and execute the requests at a time that is convenient to it, such as on a context switch or on a clock interrupt. This saves the overhead of servicing an interrupt for each request. Additionally, this has the effect of batching requests together and can further reduce the average overhead per request.

Some machines, such as MIPS [13], allow the TLB to contain entries for multiple virtual address space (VAS) at the same time. However, a TLB entry may never be used until the processor switches to the VAS containing it [14]. Therefore, all algorithms that issue interprocessor requests should be modified for these machines as follow:

(1) Still enqueue an interprocessor request to every processor in target _ set.
(2) Interrupt only processors running in the VAS that contains PTE.
(3) Before switching to a VAS, finish processing all requests for that VAS.

3.3 Deadlocking

Deadlocking is a problem only for 2P algorithm, which must disable interrupts during the algorithm. In this algorithm, the requester sends an interprocessor request to all repliers and waits for an acknowledgement from each one. Since the algorithm disables interrupts, deadlocks may happen. For example, a deadlock happens when two processors both enter the requester code, disable interrupts, interrupt the other one and wait for an acknowledgement.

The deadlocking problem can be solved by not waiting for the acknowledgement of a request, but waiting for the replier to start executing the algorithm (possibly due to another request). The requests for a processor are put in a queue in main memory, and the processor will finish all requests before leaving the algorithm. The problem can also be solved by checking the request queue and servicing requests while waiting for acknowledgements [15].

4 Conclusion and Future Work

Multiprocessor system with shared memory is a very common approach these days. And to improve the memory access time or to improve the system performance, the use of translation lookaside buffer is also very common [16]. But it may produce an inconsistency problem between the main memory page table entry and the translation lookaside buffer entry. This inconsistency causes so many faults and anomalies. This paper discusses various methods to handle inconsistency and to maintain consistency between TLB and PTE. Four algorithms were discussed; each of them can be used in different way and can be used in different situation. Interrupt handling, interprocessor communication and deadlock-related issues were also discussed and can be handle in separate ways in the algorithms.

The problem opens up the various research area in computer architecture, high-performance computing, operating system, parallel processing and database management systems. One can study and compare many other existing concurrency control algorithms for translation lookaside buffer consistency and can check their performances. One can also think about inconsistency tolerance, i.e. to work with inconsistency and then find out the resulting effects.

References

1. Agarwal, M., et al.: Inconsistency in Translation Lookaside Buffer. IEEE Xplore, 19 Nov 2016
2. McCurdy, C., et al.: Investigating the TLB behaviour of high-end scientific applications on commodity microprocessors. In: ISPASS'08, Washington, DC, USA, pp. 95–104. IEEE Computer Society (2008)

3. Rosenhurg, B.: Low-synchronization translation lookaside buffer consistency in large-scale shared memory multiprocessors. In: Proceedings of 13th Symposium Operating Systems Principles, pp. 137–146 (1989)
4. Kitchens, K.: Predicting and prefetching TLB entries from irregular access streams, 8 Dec 2013
5. Jaleel, A., Jacob, B.: In-line interrupt handling for software managed TLBs. In: 2001 IEEE International Conference on Computer Design (ICCD 2001), Austin TX, September 2001
6. Agarwal, M., et al.: A Survey on translation look aside buffer. In: National Conference on "Information & Communication Technology" organized by Sir Padampat Singhania University, Udaipur dated 5–6 March 2010
7. Jacob, B.L., Mudge, T.N.: A look at several memory-management units, TLB-refill mechanisms, and page table organizations. In: Eighth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS'98), San Jose CA, pp. 295–306, October 1998
8. Black, O.D., et al.: Translation lookaside buffer consistency: a software approach. In: Proceedings of Third International Conference Architectural Support for Programming languages and Operating Systems, pp. 113–122 (1989)
9. Adya, A., et al.: Efficient optimistic concurrency control using loosely synchronized clocks. In: ACM SIGMOD International Conference on Management of Data, San Jose, CA, May 1995
10. Sohi, G.S., Vajapeyam, S.: Instruction issue logic for high-performance, interruptable pipelined processors. In: 14th Annual International Symposium on Computer Architecture (ISCA'87), June 1987
11. Smith, J.E., Pleszkun, A.R.: Implementation of precise interrupts in pipelined processors. In: 12th Annual International Symposium on Computer Architecture (ISCA'85), Boston MA, pp. 36–44, June 1985
12. Moudgill, M., Vassiliadis, S.: Precise interrupts. IEEE Micro **16**(1), 58–67, February 1996
13. Zilles, C.B., Emer, J.S., Sohi, G.S.: Concurrent event-handling through multithreading. IEEE Trans. Comp. **48**(9), pp 903–916, September 1999
14. Nagle, D., et al.: Design tradeoffs for soft-ware-managed TLBs. In: 20th Annual International Symposium on Computer Architecture (ISCA'93), May 1993
15. Abdoos, M.: Improved deadlock prevention algorithms in distributed systems. IJEACS **02**(02), February 2017
16. Agarwal, M., et al.: Effect of TLB on system performance. In: ACM International Conference Proceeding Series, 05 March 2016

Digitization of Disaster Management: A Multimedia Ontological Approach

Vishal Lama and Sarika Jain

Abstract In the era of World Wide Web, dependency has been increasing on real-time response, communication, disaster preparedness, and analysis by our society, experts from different areas including paramedics, risk management team, police, and firefighters. To ensure real-time response, appropriate information retrieval, expediting undertaking and response preparation and handling communiqué with all involved parties, there is a requirement of digitizing the resources, concepts semantically along with exploiting the information embedded in multimedia content. We also know that ontology has been proven the excellent mean of digitization. So through this paper, we propose a multimedia-driven disaster management ontology, featuring a different representation scheme integrating the multimedia content. This will offer semantic analysis and interoperability across heterogeneous multimedia data sources, hence facilitating in real-time disaster management.

Keywords Disaster management ontology · Semantic modeling
Multimedia OWL (MOWL) · Semantic Web

1 Introduction

Event like management of disaster made us heavily reliant on a variety of sources and infrastructures like emergency responses and services during emergency situations. Variety of infrastructures evolves demanding inter-dependency and

V. Lama · S. Jain (✉)
National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India
e-mail: jasarika@nitkkr.ac.in

V. Lama
e-mail: vishal_51610080@nitkkr.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_20

inter-disciplinary approaches to disaster management [1]. There need active experts from different domain of knowledge and need to integrate information across domains not only confined to people but between machines as well. The integration and interoperability of exchange of information in a quick response situation with ease are still significantly challenging and consequently may cause misunderstanding and even the end of existence. The solution recognized is that mutual understanding is must so that machines too could read and understand and to encode this human knowledge of a domain for machines to understand we have ontologies.

Ontology is a formal and explicit demonstration of knowledge of an abstraction of a domain that provides a common understanding of information [2]. Using ontology, machine would able to understand semantics which could apply reasoning and deduce like human brain, and hence, semantic interoperability could be achieved [3]. Another significant aspect of using ontology is digitization of information in a manner that the representation and reasoning are certain.

To represent ontology, there are several techniques for knowledge representation ranging from RDF [4] to advanced description logic (DL). The World Wide Web Consortium (W3C) has defined and standardized Web Ontology Language (OWL) [5] as the language for data demonstration intended for semantic hetero-operability of data on the Web in 2004. The OWL supports text-symbolic representation of abstract model of linguistic constraints. Other things include text retrieval, information extraction and reasoning with the discovered facts [6].

There are many ontologies available related to disaster management but not of those support the multimedia content integration which somehow leads to a semantic gap [7]. So, there is a requirement of extension of ontology with symbolic media properties of the concepts. It requires formalize the meaning of the multimedia content descriptors along with handling uncertainty.

Due to the availability and abundant use of multimedia content in the digital space in previous many decades, it has motivated to develop ontology on disaster management integrating multimedia contents using the constructs Multimedia Web Ontology Language (MOWL).

The Multimedia Web Ontology Language (MOWL) [8] is a new ontology representation language with different syntax and semantics from traditional language, for knowledge-based multimedia applications. It supports two kinds of entities, namely the concepts, representing the nonconcrete real-world entities, and the media objects, representing the exhibition of concepts in the media world. Alike other ontology language, the concepts and media objects can be organized in a taxonomical hierarchy. The concepts and media objects can have properties.

A special class of properties that associates media objects with concepts represents the causal relations in the domain. The uncertainties in such causal relations are captured through a set of conditional probability tables.

2 Literature

Many projects concern with the issue of improving the efficiency of crisis management. Out of those, AktiveSA [9] has also presented an ontology-driven approach for the integration of information. The developed ontology includes many things, and few of them covers transportation, human aid, military, equipment, organization, and weapons. This ontology is not aligned with a foundational ontology, and as a consequence, it makes insignificant use of criteria for categorization. In [10], SokNOS, a prototype featuring ontology-driven system extensibility lets new system component to integrate with negligible sort of effort which could be used in dealing with an emergency situations. It has used DOLCE [11] ontology, which has vocabulary related to emergency management domain. Another methodology used was creating ontology from glossaries and vocabularies available on the Web [12]. It cited one from Canadian sources: the Emergency and Crisis Communication vocabulary and the Emergency Management Ontario (EMO) glossary, other from American source: National Incident Management System (NIMS) glossary, The Institute for Crisis, Disaster, and Risk Management (ICDRM) glossary, and Incident Command System (ICS) glossary [13].

Mallik et al. [8] define that media properties of concept unveil certain exclusive features that cannot be dealt using conceptual modeling schemes followed in the current ontology representation and reasoning scheme—OWL. Therefore, suggested a new perceptual modeling procedure—Multimedia Web Ontology Language (MOWL).

Ghosh et al. [7] outline the significance of ontology for multimedia applications. It focuses on the need and critical overview of the ontologies for multimedia applications. Few of the application include—NrityaKosha [14]: conserving the intangible heritage of Indian classical dance. In [15], also multimedia ontological approach is used to preserve the heritage including tangibles like monuments, sculpture and non-tangibles like music, dance.

From this survey, we found that we already have apt methodologies for the development of domain knowledge in disaster management. Current aspects show that the formal explicit specification used to indicate the knowledge of realm and to provide semantic over this domain knowledge is Web Ontology Language shortened as OWL, but none of those approaches looked into realm of multimedia data collections. We require semantic interpretation of multimedia contents and a semantic interoperability across heterogeneous data sources so that we could exploit information embedded in multimedia assets. In order to achieve the semantic gap [7], we find and will use some of the features of Multimedia Web Ontology Language (MOWL) [8], and an extension of OWL provides multimedia content integration.

Putting the light on the applications of multimedia ontology, one of its features addresses the specific needs of knowledge representation for multimedia content. With the utilization of MOWL, we can able to extend the ontology with symbolic media properties of the concepts. This has made us propose a new methodology for ontology development in support with MOWL for the Digitization of Disaster Management.

3 Drafting of a Multimedia Ontology for Disaster Management

The created multimedia has gone through to many phases as follows

3.1 Construction

Editor: Traditional ontologies can be developed using an open-source ontology editor named Protégé or can code using any text editor in any of the supporting formats. The multimedia disaster management ontology created is totally hand-coded in any text editor and saved with extension as .mowl.

Formats and Construct: Various formats are being used in writing ontology, and few include RDF/XML, TURTLE, XML/OWL, or NTriple. In order to write mowl file, the supporting format is XML. Concepts and the media objects are the elementary language constructs of the language. Real-world objects or events are being represented by concepts, whereas manifestation of the concepts is being represented by the media objects in different media forms. Once the detection of the media objects has been done, it leads to the concept recognition. Along with the elementary constructs of RDF/OWL, we have used constructs like <mowl:Concept>, <mowl:hasMediaExample>, and <mowl:MediaExample> of mowl to incorporate the concepts and media properties. <mowl:Concept> depicts the real-world instance or event, whereas media objects depict the appearance of concepts in various diverse forms.

Following is a snippet of disaster management ontology

```
<owl:Class rdf:ID="Event">
  <rdfs:subClassOf rdf:resource="&mowl;Concept"/>
</owl:Class>
<owl:Class rdf:ID="DisasterManagement">
  <rdfs:subClassOf rdf:resource="#Event"/>
  <mowl:hasMediaExample rdf:resource="#DM"/>
</owl:Class>
<owl:Class rdf:ID="EmergencySituation">
  <rdfs:subClassOf rdf:resource="&mowl;Concept"/>
</owl:Class>
<owl:Class rdf:ID="EmergengyResponse">
  <rdfs:subClassOf rdf:resource="#Requirements"/>
</owl:Class>
<owl:ObjectProperty rdf:ID="callsForAn">
  <rdfs:domain rdf:resource="#Disaster"/>
  <rdfs:range rdf:resource="#EmergencySituation"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:ID="callsFor">
  <rdfs:domain rdf:resource="#DisasterManagement"/>
  <rdfs:range rdf:resource="#EmergencySituation"/>
</owl:ObjectProperty>
<mowl:MediaExamples rdf:ID="DM">
  <mowl:hasURI>C:\Users\VISHAL\NIT\MediaExamples\DM2.jpg</mowl:hasURI>
</mowl:MediaExample>
```

3.2 Visualization

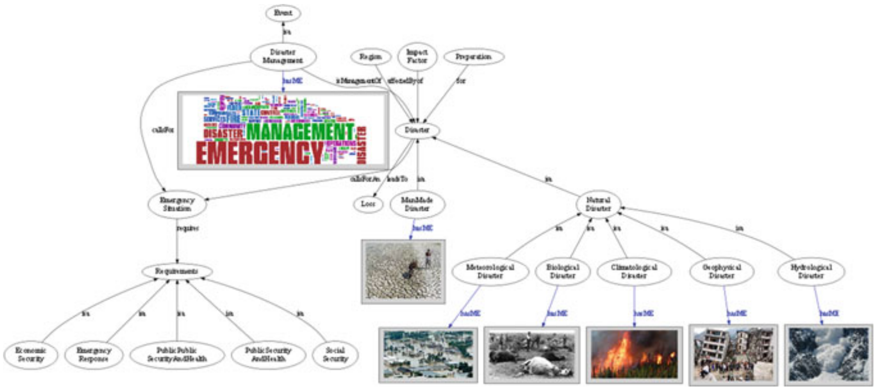
Following requirements are needed to carry out the visualization:

MOWL parser and APIs: Mowl parser is one of the structural specified components which is responsible for deploying mowl file into Bayesian network. We used Java for creating mowl parser. The APIs used are NanoXML and Netica. NanoXML itself provides a non-validating parser for Java. NanoXML further has many classifications, but we used standard and recommended out of it, i.e. NanoXML/Java. To work with belief network, commonly called Bayesian network or Bayes network or causal probabilistic networks, we used NeticaJ, a Java version of the Netica API. It is used to represent the relationship between concepts, even if the factors like uncertainty, unpredictability, or impression are involved. It may be learned automatically from ontology created and captures the knowledge that can be transported from one situation to another. It also allows a clear visualization of ontology.

Visualization Tool: For visualization, we used GVEdit tool by Graphviz—graph visualization software, an open source.

Working: In mandate to carry out the visualization of created multimedia ontology, various steps are considered. The multimedia ontology developed using text editor with extension .mowl has to be fed into the mowl parser as an input file and as a result generates another file. Generated file with extension .dot contains the Bayesian network of the multimedia ontology developed. This .dot file now can be visualized through suitable visualization software. In order to visualization of Bayesian network, we used GVEdit as a visualization tool.

The Bayesian net for the ontology is as follows



4 Conclusion

With this we conclude, digitization of critical events like disaster management is very crucial as it aids in finding the appropriate measures that are taken during such emergency situations. With digitization real-time situation, appropriate data retrieval, analysis, and suitable actions could be taken. It helps in training personal purposes, public benefits, data integration. We also find among many, the proposed ontology on disaster management is the only one which supports multimedia integration helping in exploiting the semantics embedded with multimedia facets. As a part of extension, we try to incorporate Bayesian probabilistic reasoning, another important feature of MOWL. We will also try to map and represent this multimedia ontology as an E-MOWL (event representation) [16] extending MOWL.

References

1. Jain, S., Mehla, S., Mishra, S.: An ontology of natural disasters with exceptions. In: International Conference IEEE on System Modeling & Advancement in Research Trends (SMART), pp. 232–237, Nov 2016
2. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5** (2), 199–220 (1993)
3. Mishra, S., Jain, S.: A study of various approaches and tools on ontology. In: IEEE Conference CICT 2015 in ABES College of Engineering, Ghaziabad, pp. 57–61 (2015)
4. Ontology Representation Standard by W3C. Resource Description Framework (RDF). https://www.w3.org/RDF/ (2014)
5. Ontology Representation Standard by W3C. Web ontology Language (OWL) https://www.w3.org/OWL/ (2012)
6. Jain, N.K., Jain, S.: Live multilingual thinking machine. J. Exp. Theor. Artif. Intell. (Taylor and Francis) **25**(4), 575–587 (2013)
7. Ghosh, H., Chaudhury, S., Mallik, A.: Ontology for multimedia applications. IEEE Intell. Inf. Bull. **4**(1), 21–30 (2013)
8. Mallik, A., Ghosh, H., Chaudhury, S., Harit, G.: MOWL: an ontology representation language for web-based multimedia applications. ACM Trans. Multimed. Comput. Commun. Appl. (TOMM) **10**(1), 8 (2013)
9. Smart, P.R., Russell, A., Shadbolt, N.R., Schraefel, M.C., Carr, L.A.: Aktivesa: a technical demonstrator system for enhanced situation awareness. The Comput. J. **50**(6), 703–716 (2007)
10. Babitski, G., Bergweiler, S., Grebner, O., Oberle, D., Paulheim, H., Probst, F.: SoKNOS–using semantic technologies in disaster management software. In: Extended Semantic Web Conference, pp. 183–197. Springer, Berlin, Heidelberg, 29 May 2011
11. Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A.: Wonderweb deliverable d18, ontology library (final). ICT project 33052, 31 Dec 2003
12. Grolinger, K., Brown, K.P., Capretz, M.A.: From Glossaries to Ontologies: Disaster Management Domain
13. Han, J.: Terror Preparedness of Municipal Departments in Illinois in 2013. Western Illinois University (2014)
14. Mallik, A., Chaudhury, S., Nrityakosha, G.H.: Preserving the intangible heritage of Indian classical dance. J. Comput. C. Herit. (JOCCH) **4**(3), 11 (2011)
15. Mallik, A., Chaudhury, S., Dinesh, T.B.: An intellectual journey in history: preserving Indian cultural heritage. In: International Conference on Image Analysis and Processing, pp. 298–307. Springer, Berlin, Heidelberg, 9 Sept 2013
16. Pahal, N., Chaudhury, S., Lall, B.: Extending mowl for event representation (e-mowl). In: Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 03, pp. 171–174. IEEE Computer Society, 17 Nov 2013

Prevalent Study of Cooperative Biometric Identification System Using Multimodal Biometrics

Shaleen Bhatnagar and Sushmita Kumari

Abstract Multimodal biometric systems are gaining popularity as they have greater precision compared to unimodal biometric systems. To develop a system that ensures greater reliability, robustness, foolproof identification, it is essential to incorporate biometric traits in the system. This research work proposes a cooperative biometric multimodal for fingerprint identification based on minutiae matching by addressing above-mentioned problems. This research paper proposes a design of such fingerprint identification modal which is expected to provide better accuracy and acceptance rate. We discuss here a multimodal biometric system, its utility, and benefits over unimodal biometric systems.

Keywords Cooperative biometric system · Multimodal system
Security · Performance evaluation · Integrating multiple features

1 Introduction

Prevalent security systems use passwords, PIN, or tokens, and all can be easily duplicated or misplaced. To develop a system that ensures greater reliability, robustness, foolproof identification, it is essential to incorporate biometric traits in the system. A biometric technique of identification and verification uses a person's physical features or behavioral feature to verify their identity. Biometric samples are very much unambiguous to every individual, effortlessly acquired non-intrusively, and remain mostly unchanged for entire life span and detectable without any specialized guidance. It has one of a kind impression, which is acceptable everywhere [1].

S. Bhatnagar (✉) · S. Kumari
Presidency University, Bengaluru 562106, India
e-mail: shaleenbhatnagar@gmail.com

S. Kumari
e-mail: sushmitakumari@presidencyuniversity.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_21

2 Biometric System

Biometric scheme uses a person's physical features or behavioral feature to verify their identity. Biometric samples are very much unambiguous to every individual, effortlessly acquired non-intrusively, and remain mostly unchanged for entire life span and detectable without any specialized guidance. It has one of a kind impression, which is acceptable everywhere. There are two important phases in a biometric authentication technique: data registration phase and data confirmation phase. During data registration phase, biometric data samples are collected and after applying appropriate processing on the samples, they get stored in the database in form of template. During data confirmation phase, same processing will apply on the new data sample and testify the template characteristics to confirm the authenticity [2] (Figs. 1 and 2).

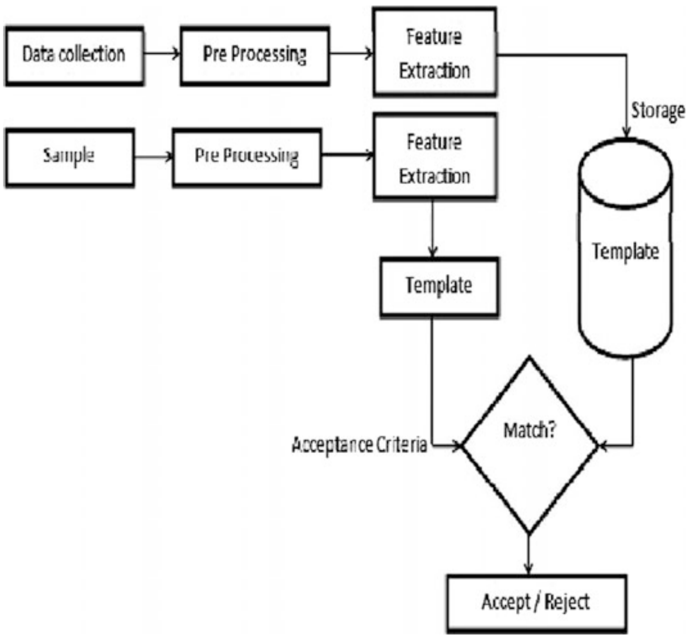


Fig. 1 General biometric model

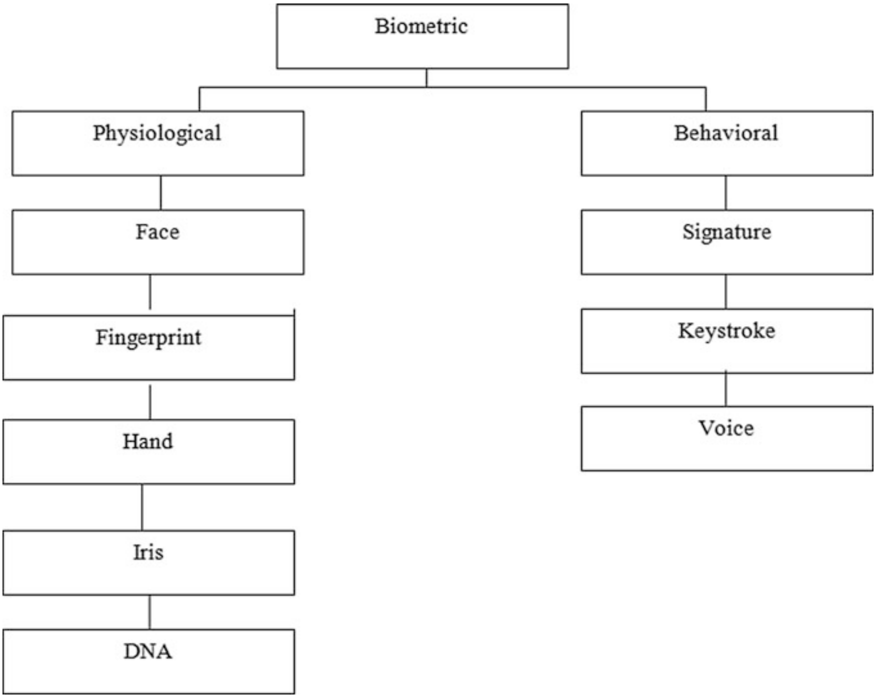


Fig. 2 Various body attributes that can be used for biometric identification

2.1 Physical Versus Behavioral Characteristics

Thus, there are different biometric attributes that can be used for assured authenticity of a person, and selection of biometric attribute depends according to the necessity of application. Attributes according to which we can select suitable biometric are given below:

1. Universality: Without exception, every human should possess these characteristics.
2. Uniqueness: Characteristics possessed by only one.
3. Permanence: Those characteristics which are persistent and should not change with time.
4. Acceptability: It extends to which humans are agreeable to acquire a specific biometrics in their day-to-day life.
5. Performance: It means that the attainable identification accuracy, robustness, speed, and the resources needed to obtain accuracy and speed.
6. Collectability: It means how easy is to capture the biometric characteristics [3].

3 Challenges in Unibiometric Systems

As compared to the traditional methods of person recognition, biometrics provides greater security, but there are many challenges in biometric technology [4].

False Match (or False Accept): It means at the time of identification, system inaccurately giving successful matching. In a biometric system, it should be minimized as possible and ideally it should be zero.

False Non-match (or False Reject): It means at the time of identification, system inaccurately giving failure of matching. In a biometric system, it should be minimized as possible and ideally it should be zero.

Failure to Capture (FTC): It means that when biometric system is not able to capture a good quality sample.

Failure to Enroll (FTE): It means that when biometric system is not able to store appropriate template in the database.

4 Literature Survey

Most of the available literature tends to categorize that the biometric systems are effective for human identification and authorization over various levels of implementation, for small to a large population, such systems are difficult to forge and can be made for secure by combining more than one biometric traits, that is multimodal biometric systems. Biometric system proposed by Yang and Ma [5] used fingerprint, palm print, and hand geometry to implement personal identity verification, which can be extracted from one image. In this system matching score fusion is done on different levels. First fusion of the fingerprint and palm print features and then, successive fusion between the multimodal system and the palm-geometry unimodal system. Hybrid approach using fingerprint and iris features is proposed by Besbes et al. [6]. This approach is based on fingerprint minutiae extraction and iris template encoding. In this system, two independent recognition modalities providing its own decision are used. Both the unimodal decisions are combined using an AND operator to provide a final decision. Another multibiometric method is proposed by Aguilar et al. [7]. This system enhances fingerprint image using combination of fast Fourier transform (FFT) and Gabor filter. The proposed system uses the fingerprints of both thumbs. Here fingerprints of both thumbs are processed separately, thus generating unimodal results for each thumb. These unimodal results are compared, and final fused result is provided.

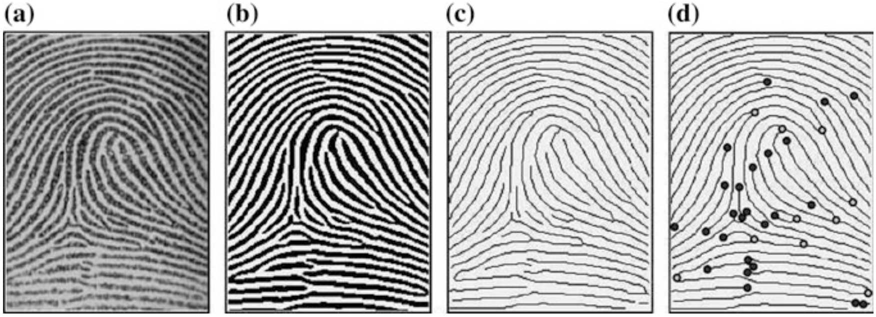


Fig. 3 Fingerprint recognition

5 Fingerprint

Fingerprint identification or fingerprint verification systems deal with a systematic procedure of comparing two person's finger impressions. It is the most widely used biometric characteristic that is used to give genuine and assured authenticity of a person. It is used widely due to its uniqueness and persistency over time [8].

We know those fingerprints are made of different ridges and valleys on the surface of the finger. Upper skin layer segments of the finger patterns are known as ridges, and the lower segments are termed as valleys. These ridges and valleys combined to form a minutia point. There are many types of minutiae present in fingerprint, including dots, islands, ponds or lakes, spurs, bridges, and crossovers. By these minutiae points, it is determined that how unique the fingerprint is [1].

Generally, fingerprint identification systems have two main steps [9] (Fig. 3)

(a) Extraction of Minutiae.

(b) Matching of Minutiae.

6 Cooperative and Non-cooperative Biometric

Non-cooperative biometric systems are those systems when an individual does not know that his/her biometric information is taken. That means without the person's knowledge, sample is acquired. For example, a person drinking coffee in a shop and without his knowledge, his fingerprint and DNA sample can be taken from the coffee glass, whereas in cooperative biometric, the person knows about the enrollment and verification, and the sample cannot be collected without his/her permission, for example, iris scan, fingerprint scan.

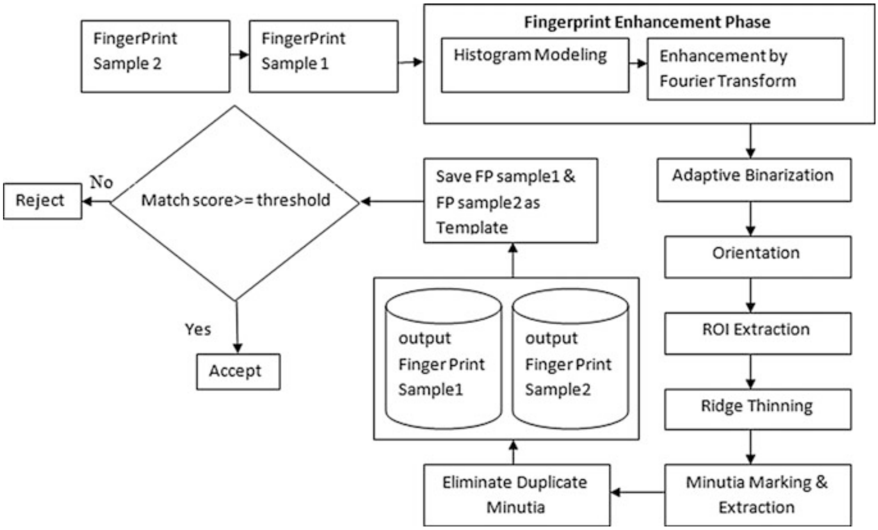


Fig. 4 Integrating multiple traits in biometrics system

7 Proposed Method for Integrating Multiple Traits in Biometric

In the above Fig. 4, we are integrating two fingerprints and making a cooperative multimodal biometric system, where we are applying preprocessing to obtain results from preprocessed images providing certainty of accurate matches. Similarly using different traits we can make multimodal biometric system for other biometric traits such as face and iris, face, iris and fingerprint etc. These systems will give better performance and accurate results as compared to unimodal biometric systems.

8 Conclusion

Few problems noticed in unimodal system are reduced by multimodal biometric systems. Besides correcting matching performance of systems, they also point out the issues of non-universality and spoofing. Multimodal biometric systems can combine information at different stages, extremely well-known one is integrating the individual processing score and checking with the match score. Using multi-modal biometric increases the probability of greater and assured authenticity of the user and decreases the issues like false acceptance and false rejection. We are getting more robust and accurate verification and identification systems by using multimodal biometric systems; the only drawback is cost overhead as compared to unimodal biometric systems, and thus, it is necessary to evaluate the cost against performance on prior basis.

References

1. Bhatnagar, S.: Cooperative biometric multimodal approach for identification. In: Satapathy, S., Joshi, A. (eds.) Information and Communication Technology for Intelligent Systems (ICTIS 2017)—Volume 1. ICTIS 2017. Smart Innovation, Systems and Technologies, vol. 83. Springer, Cham (2018)
2. Bhatnagar, S.: A reconsideration of schemes for fingerprint identification. Int. J. Advanc. Res. Comput. Sci. **4**(3) (2013)
3. Jain, L.C.: Intelligent Biometric Techniques in Fingerprint and Face Recognition. CRC Press (1999)
4. Jain, A.K., Ross, A.: Learning user specific parameters in a multibiometric system. In: Proceedings of International Conference on Image Processing
5. Yang, F., Ma, B.: A new mixed-mode biometrics information fusion based-on fingerprint, hand-geometry and palm-print. In: Proceedings of 4th International Conference on Image Graph, pp. 689–693 (2007) https://doi.org/10.1109/icig.2007.39; Donaldson, W.R., Gan, C.K.: Adaptive silence detection for speech storage and voice mail applications. In: lEEE Transactions on ASSP, vol. 13, pp 924–925 (1988)
6. Besbes, F., Trichili, H., Solaiman, B.: Multimodal biometric system based on Fingerprint identification and Iris recognition. In: 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA 2008), pp. 1–5. https://doi.org/10.1109/ICTTA.2008.4530129
7. Aguilar, G., Sánchez, G., Toscano, K., Nakano, M., Perez, H.: Multimodal biometric system using fingerprint. In: Proceedings of International Conference on Intelligent and Advanced Systems, pp. 145–150 (2007). https://doi.org/10.1109/ICIAS.2007.4658364
8. Jain, A., Pankanti, S.: Fingerprint classification and matching (Department of Computer Science & Engineering, Michigan State University) & (Exploratory Computer Vision Group. IBM T. J. Watson Research Centre)
9. Wang, X., Li, J., Niu, Y.: Fingerprint matching using orientation codes and polylines. Pattern Recogn. **40**(11), 3164–77 (2007)

Most Persistent Feature Selection Method for Opinion Mining of Social Media Reviews

Savita Sangam and Subhash Shinde

Abstract Many business organizations use social media data in order to understand their customer on an individual level. Consumers are keen to share their views on certain products or commodities. This leads to the generation of large amount of unstructured social media data. Thus, text data is being formed gradually in many areas like automated business, education, health care, show business. Opinion mining, the subfield of text mining, deals with mining of review text and classifying the opinions or the sentiments of that text as positive or negative. The work in this paper develops a framework for opinion mining. It includes a novel feature selection method called Most Persistent Feature Selection (MPFS). MPFS method uses information gain of the features in the review documents. The performance of the three different classifiers, namely Naïve Bayes, Maximum Entropy, and Support Vector Machine, with the proposed feature selection method is evaluated on movie reviews using the parameters accuracy, precision, recall, and F-score. The different classifier models generated show the acceptable performance in comparison with the other existing models.

Keywords Feature selection · Opinion mining · Sentiment classification

S. Sangam (✉)
Shivajirao S. Jondhale College of Engineering, Dombivli(E), Thane, Maharashtra, India
e-mail: savita.sangam@gmail.com

S. Shinde
Lokmanya Tilak College of Engineering, Kopar Khairane, Navi Mumbai, Maharashtra, India
e-mail: skshinde@rediffmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_22

1 Introduction

The review data produced by the social media applications may not be in a proper structure and may require lot of processing in order to make it usable. In order to process these reviews, data models need to be constructed. The focus of this research work is to process and analyze the opinions or the sentiments of the social media reviews by applying some data mining techniques. According to Liu [1], the study of analyzing opinions in written language is termed as opinion mining. According to Pang and Lee [2, 3] "the task of analyzing the opinion, sentiment, and subjectivity computationally is known as opinion mining," and it is also called as Sentiment Analysis (SA). Decision makers rely on SA for making their decisions. For example, various shopping sites like Amazon, Flipkart take feedback from the customers which will help them to take proper decisions for improving the quality of their services. SA techniques have been applied widely in many areas like business, entertainment, medicine, politics. Sentiment Classification (SC) process classifies the sentiments of the text reviews into negative or positive or sometimes neutral. The two main approaches for sentiment classification are lexicon-based approach and machine learning approach. In lexicon-based approach, a sentiment score is calculated using a dictionary of positive and negative words with a positive or negative sentiment value assigned to each of the words. The overall sentiment of the entire text passage is sum or average (or any other function) of all the words. This approach is domain specific and gives low recall.

The machine learning algorithm uses labeled data sets in order to perform the classification task. The classifier gets trained on training data in the form of features which are the words or phrases in the text. It then classifies the unseen test data based on its training. There are three types of machine learning techniques, namely supervised, unsupervised, and semi-supervised. Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Maximum Entropy (MAXENT) are some of the machine learning systems mentioned by the researchers for the sentiment classification work. Different feature selection mechanisms to select the features in the text and deeper analysis of the sentences as a whole are the points that are to be considered for the accurate sentiment classification.

In this paper, we have proposed a novel feature selection method and proved that the popular machine learning techniques when trained using the features produced by our method generate good enough classification accuracy of up to 96%. The classifiers are also trained using simple unigrams, bigrams, and trigrams features, and their performance is compared with our models. The results produced with the proposed method are found to be satisfactory and are discussed in detail in result analysis section.

The rest of the paper is arranged as follows: Sect. 2 includes related work, proposed framework for the opinion mining is explained in Sect. 3, experimental results and result analysis are given in Sect. 4, and conclusion and future scope of the work is presented in Sect. 5.

2 Related Work

Lot of research work is being undertaken in opinion mining in recent times. Researchers are working on classifying the sentiments of the reviewers for different domains like restaurant reviews, product reviews, and movie reviews. Sentiment classification task has been done using machine learning approach, lexicon approach, or the combination of both the approaches which will produce a hybrid approach. It can be performed at three levels, document level, sentence level, and feature level [4]. NB and SVM models are normally used as baselines for other systems in text labeling and sentiment analysis research [5]. Pang et al. [6] first used these classification methods in their experiments to classify movie reviews. A lexicon-based approach is based on the overall sentiment score of the sentiment words in the passage of text [7]. This approach is mentioned by Hu and Liu for the first time for aspect level and sentence level sentiment classification. Sentiment classification at the sentence level is analogous to document sentiment classification as sentences are part of the documents [8]. But this task is difficult as sentences are less informative when compared to the entire document. There are different types of sentences like direct sentence (e.g., the movie is superb) and indirect sentences (e.g., Golmal 2 is almost like its previous version) which require more understanding of the problem. Feature level classification tries to determine the sentiment on certain aspects in the text reviews.

The words, terms, or the phrases present in the text passage which contribute in finding the polarity of the sentiment of the text passage are called as features. The machine learning systems first get trained on these features and then classify the unseen text. Selection of the best features ensures better accuracy of the classifier by reducing the dimensionality of the training data set. There are several approaches mentioned in the literature for finding out the finest features [9–11].

Opinions can be expressed in any language. Many researchers have worked on multilingual data. The work usually translates data from one language to another and then finds the sentiments of the original data. Cross-language sentiment classifiers are built for various languages like Chinese, Spanish, and Arabic by many researchers achieving comparable results with the monolingual ones [12, 13]. The authors of the paper [14] have combined two different machine learning algorithms, SVM and Artificial Neural Network (ANN) for sentiment classification of movie reviews' data. Zainuddin and Selamat [15] have mentioned three different weighting schemes to generate the word vectors which are Term Frequency-Inverse Document Frequency (TFIDF), Binary Occurrence (BO), and Term Occurrence (TO). In [16], the authors showed that Naive Bayes with binarized features seems to work better for several text classification tasks. In [17], the authors proposed a statistical method using weight by Gini Index method for selecting the features. Chunping et al. [18] introduced word embedding features based on deep learning technology for optimizing the accuracy of their proposed model to carry out attribute-level sentiment analysis.

Many researchers have developed feature selection algorithms which lack in finding the most informative features that are needed for the machine learning algorithms to produce accurate results. Single words or unigrams are considered as best features but they require more space and time for processing. Hence in this study, most persistent bigrams and trigrams are selected as informative features and are considered for training the sentiment classifiers.

3 Framework for Opinion Mining

The machine learning approaches require a set of useful features for sentiment classification. So a novel feature selection method called Most Persistent Feature Selection (MPFS) that makes use of information gain of the features in the text is proposed here. MPFS method is applied on bigram and trigram features in the documents. Most of the existing work shows that NB and SVM are perfect methods in sentiment classification. So these classifiers and also the MAXENT are used as the base classifiers in our approach. The brief description of the framework is given in Algorithm 1.

Algorithm 1: Algorithm for the proposed framework

Input: movie reviews documents
Output: review documents classified as positive or negative

```
Step 1: Preprocessing of the review documents is carried out to filter out
        punctuations, stop words and special characters.
Step 2: The preprocessed documents are tokenized into bigrams (trigrams)
        features.
Step 3: Features with minimum occurrence of three or more times are selected.
Step 4: for each feature fi in document di feature score is calculated
        using chi square scoring function.
Step 5: for each document di
            for each feature fi
                if feature score is not less than 0.5
                    select fi as most persistent feature
                else reject the feature.
                end if
            end for (//feature)
        end for (//document)
Step 6: The selected most persistent feature set is then given as input to sentiment
        classifier for classification.
```

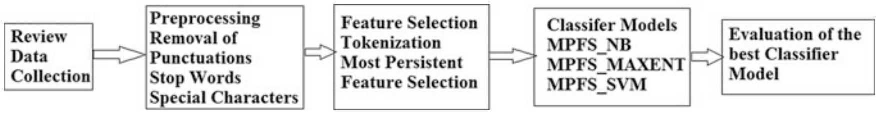


Fig. 1 Proposed system architecture

The framework is diagrammatically shown in Fig. 1.

The detailed description of the above figure is explained as follows.

3.1 Review Data Collection

The review data can be collected from web which contains the social media data like Facebook, Twitter, and blogs. Several review datasets of movies, products, restaurants are available for sentiment classification task. The dataset used in this work is movie reviews dataset developed by Pang and Lee [3] and Zhou et al. [12]. It contains 2000 processed positive and negative text files.

The reviews on the movies are considered because they contain range of emotions or sentiments.

3.2 Preprocessing

The activities involved here are:

- Removal of punctuations marks (“.”, “:”, “?”, etc.)
- Filtering out natural language specific stop words (in, on, an, etc.)
- Elimination of special characters (“@”, “$”, “#”, etc.).
- Discarding repetitive characters like in okkkk, gooo, noooo, etc.

3.3 Feature Selection

Feature selection techniques can be used to identify and remove irrelevant and redundant features that do not contribute in increasing the accuracy of the model. Several combinations can be made for selecting features which involves lots of effort. Therefore, sophisticated methods are required to perform feature selection in practice. The MPFS method is applied here to find the most relevant features in the review documents.

3.4 MPFS Method

The MPFS method proposed here tries to find out the most persistent features in the documents. Initially, the feature set consists of all the bigrams like “movie is”, "is very", "very beautiful". Instead of considering all the bigrams, only useful bigrams like "very beautiful" which contributes mainly in finding the sentiment can be considered. This requires the information gain or the bigram score of the features. This score is calculated using the chi-square statistic. The bigram features with the score of 0.5 or above are taken as the most persistent features. Similar to bigrams, trigrams (e.g., "not so good") are also considered here for the experimentation purpose.

4 Results

The experiments are conducted on Intel core i3 processor with 32/64 bit operating system. Python 3.5.2 with NLTK 3.2.1 version is used for programming purpose. The experiments are performed on various size of movie reviews dataset. The movie reviews polarity dataset used in this work contains a folder named movie_reviews. In this folder, there are subdirectories called "pos" and "neg" which contain 1000 positive and 1000 negative processed text files, respectively. The performance is evaluated using fivefold and tenfold cross-validation (CV) methods. Based on the feature set selected (MPBFS—Most Persistent Bigram Feature Selection, MPTFS—Most Persistent Trigram Feature Selection) and the classifier used, different classifier models are generated. The results produced by the various models are presented in Tables 1 and 2.

Figure 2 is the graphical illustration for the comparison of the performances of various models in terms of accuracy.

Figure 3 shows the pictorial representation for performance comparison of classifier models with fivefold CV and tenfold CV.

4.1 Result Analysis

In Table 3, the accuracies of the three classifiers with unigram, bigram, and trigram features are compared with the accuracies produced using MPBF and MPTF.

Table 1 Fivefold cross-validation result

S. No	Classifier	Accuracy	Precision	Recall	F-measure
1	MPBFS_NB	0.814	0.836	0.817	0.814
2	MPTFS_NB	0.764	0.810	0.763	0.754
3	MPBFS_MAXENT	0.79	0.821	0.794	0.785
4	MPTFS_MAXENT	0.748	0.809	0.749	0.734
5	MPBFS_SVM	0.981	0.983	0.980	0.980
6	MPTFS_SVM	0.976	0.981	0.975	0.974

Table 2 Tenfold cross-validation result

S. no	Classifier	Accuracy	Precision	Recall	F-measure
1	MPBFS_NB	0.800	0.823	0.800	0.796
2	MPTFS_NB	0.763	0.812	0.763	0.752
3	MPBFS_MAXENT	0.768	0.811	0.772	0.758
4	MPTFS_MAXENT	0.741	0.806	0.743	0.723
5	MPBFS_SVM	0.964	0.964	0.963	0.963
6	MPTFS_SVM	0.964	0.967	0.963	0.963

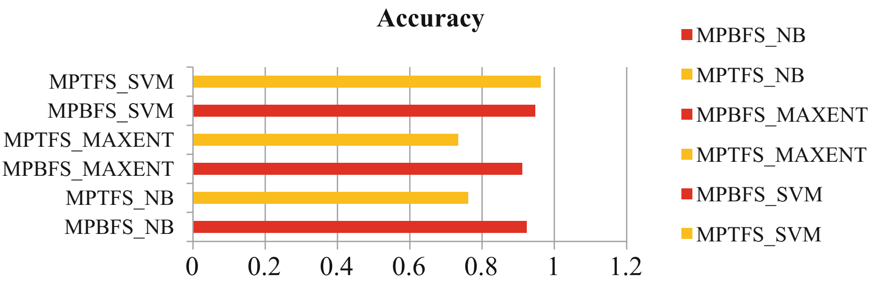


Fig. 2 Performance of NB, MAXENT, and SVM using MPBFS and MPTFS

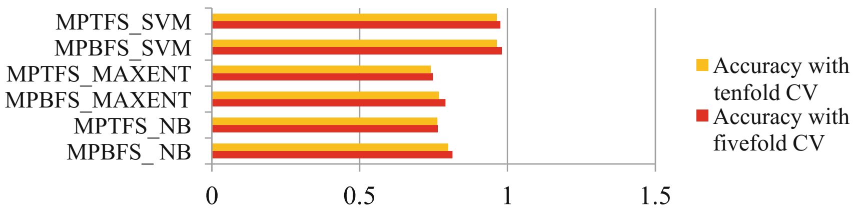


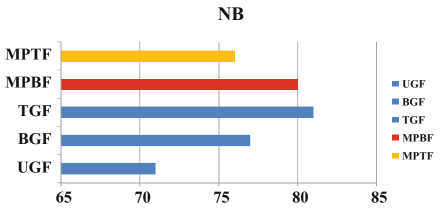
Fig. 3 Performance comparison with fivefold CV and tenfold CV

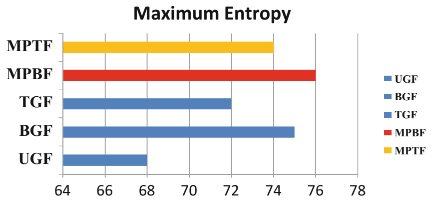
Table 3 Performance evaluation of classifiers with different features

S. no	Classifier	Unigrams features	Bigram features	Trigram features	MPB features	MPT features
1	NB	0.71	0.776	0.785	**0.800**	0.763
2	MAXENT	0.687	0.752	0.723	**0.768**	0.741
3	SVM	0.846	0.944	0.956	**0.964**	0.964

All the classifiers performed notably better with the features selected using MPBFS method when compared to other features. But in case of MPT features, SVM and MAXENT displayed better accuracy when compared to their accuracy with trigram features, but the accuracy of NB decreased. So it can be concluded that

Feature selection method	NB	Maximum Entropy	SVM
UGF	71	68	84
BGF	77	75	94
TGF	78	72	95
MPBF	80	76	96
MPTF	76	74	96





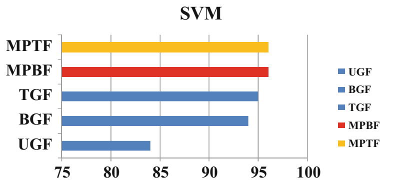


Fig. 4 Performance of NB, MAXENT, and SVM using UGF, BGF, TGF, MPBF, and MPTF

MPBFS method works well with all the classifiers used for the experiments. Figure 4 shows graphically the performance of NB, MAXENT, and SVM using unigram features (UGF), bigram features (BGF), trigram features (TGF), MPBF, and MPTF.

5 Conclusion

A novel feature selection method proposed here uses only the most persistent features of the review documents for sentiment classification purpose. The classifiers are trained on those selected features, and testing is done with five- and tenfold cross-validation methods. The reduction in the dimensionality of the feature set and also the use of most important features make the classifiers perform better in terms of accuracy which is the main objective of this work. The results show that SVM with the most persistent features shows better performance when compared to other models. The use of most persistent bigrams works well with all the three classifiers used. The results can be further improved by applying optimization technique using genetic algorithm for feature selection. The different combination of the classifiers to generate a hybrid classifier model may be another idea to improve the performance of sentiment classification.

References

1. Liu, B.: Sentiment Analysis and Opinion Mining. vol. 5, No. 1. Morgan & Claypool Publishers (2012)
2. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis (2008)
3. Pang, B., Lee, L.: A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: Proceedings of ACL (2004)
4. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings ACM SIGKDD, pp. 168–177 (2004)
5. Wang, S., Manning, C.D.: Baselines and bigrams: simple, good sentiment and topic classification. In: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, 8–14 July 2012, pp. 90–94, Jeju, Republic of Korea (2012)
6. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In: Proceedings of EMNLP, pp. 79–86 (2002)
7. Jurek, A., Mulvenna, M.D., Bi, Y.: Improved lexicon–based sentiment analysis for social media analytics. Secur. Informat. (2015)
8. Fu, G. Wang, X.: Chinese sentence-level sentiment classification based on fuzzy sets. In: 2010 Proceedings of the 23rd International Conference on Computational Linguistics, August 2010, Poster Volume, pp. 312–319, Beijing (2010)
9. Xing, F., Zhan, J.: Sentiment analysis using product review data. J Big Data (2015)
10. Tripathy, A., Anand, A., Rath, S.K.: Classification of sentiment reviews using n-gram machine learning approach. Exp. Syst. Appl. **57**, 117–126 (2016)
11. Sohail, S.S., Siddiqui, J., Ali, R.: Feature extraction and analysis of online reviews for the recommendation of books using opinion mining technique. Perspect. Sci. **8**, 754–756 (2016)
12. Zhou, X., Wan, X., Xiao, J.: CL opinion miner: opinion target extraction in a cross-language scenario. IEEE/ACM Trans. Audio Speech Lang. Process. **23**(4) (2015)
13. Tartir, S., Abdul-Nabi, I.: Semantic sentiment analysis in Arabic social media. J. King Saud Univ. Comput. Informat. Sci. 29, 229–233 (2017)
14. Tripathy, A., Anand, A., Rath, S.K.: Document-level Sentiment Classification Using Hybrid Machine Learning Approach. Springer, London (2017)
15. Zainuddin, N., Selamat, A.: Sentiment Analysis Using Support Vector Machine, pp. 333–337. IEEE I4CT, Langkawi, Kedah, Malaysia (2014)
16. Jurafsky, D., Martin, J.H.: Naive Bayes and sentiment classification. Speech Lang. Process. (2016)
17. Manek, A.S., Shenoy, P.D., Mohan M.C., Venugopal, K.: Aspect Term Extraction for Sentiment Analysis in Large Movie Reviews Using Gini Index Feature Selection Method and SVM Classifier. Springer Science + Business Media, New York (2016)
18. Chunping, O., Yongbin, L., Shuqing, Z., Xiaohua, Y.: Opinion objects identification and sentiment analysis. Int. J. Database Theor. Appl. 8(6), 1–12 (2015)

Physical Layer Security with Energy Harvesting Techniques: A Study

Poonam Jindal and Rahul Kumar

Abstract In wireless communication, the message signal transmitted by a source is received by all the nodes present in the range of the transmitter; thus, reliable and secure links are desired. Moreover, a battery of the relay nodes present in a wireless network needs to be recharged on regular basis. The combination of physical layer security and energy harvesting offers reliable and secure transmission of information plus prevents draining out the battery of relays that helps in cooperative communication. Therefore, it has become a hotspot area for research work. In this paper, we present a study on research progresses in the area of physical layer security with energy harvesting techniques.

Keywords Physical layer security · Energy harvesting · Cooperative communication · Simultaneous wireless information and energy transfer

1 Introduction

Wireless communication has witnessed most technological revolutions. Each generation of this technology has brought improved characteristics in terms of speed, size of device, applications, device lifetime, etc. It is being used in every sector and getting better with each passing day. So, it is one of the most thrust areas for research which needs to be explored extensively. Due to the open nature of wireless medium, the transmitted signal is overheard by both legitimate as well as illegitimate receivers. Therefore, the transmitted data is vulnerable to a number of security attacks so providing security is a critical issue in wireless networks and required to be explored to the great extent. A new paradigm that has gained lot of importance among research community is physical layer security. The principle behind the physical layer security is [1]: if the capacity of the legitimate channel is

P. Jindal · R. Kumar (✉)
Department of Electronics & Communication, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India
e-mail: mailiirahulkumar@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_23

higher than the capacity of the eavesdropper channel, then the transmission can be done at a rate closer to that of the channel capacity of the intended destination. In this way, only the authorized receiver decodes the signal.

One important feature of the wireless technology is that it allows two remotely located nodes to communicate. For this, the range of the network should be wide enough and cooperative communication [2] is a promising way to achieve large coverage area. In cooperative schemes, an intermediate node called as relay is used between the sender and the receiver to complete the transmission. The relay node provides a replica of the received signal, and this replica is combined with the direct signal at the destination [3]. The key aspect of this technique is how the processing of signal is done at the relay. For this, different protocols have been described in the literature among which amplify-and-forward (AF) relaying and decode-and-forward (DF) relaying are the most common [3]. The cooperation provided by the relays reduces the path loss between the nodes, increases the reliability of the links, improves system capacity and provides cost-effective coverage extension of the wireless network [3] (Fig. 1).

Another issue in wireless network is that the nodes have limited lifetime and need to be recharged periodically to remain active. Because of the energy constrained nodes, the overall performance of the system gets limited. There has been a phenomenal growth in the area of RF Energy Harvesting (EH) where power constrained devices are either inaccessible or recharging and replacing them is not convenient, expensive and unsafe. EH is based on the fact that the same RF signal can be used for carrying the information as well as the energy. This technique is referred to as simultaneous wireless information and power transfer (SWIPT) [4]. In [5], authors proposed an optimal secrecy policy (OSP), depending on channel state information (CSI) available to investigate the secrecy rate that can be achieved in an EH communication system by deciding when to transmit and how much power should be transmitted.

In this paper, a study on physical layer security along with EH and cooperative communication techniques has been presented. It is found that high secrecy rate,

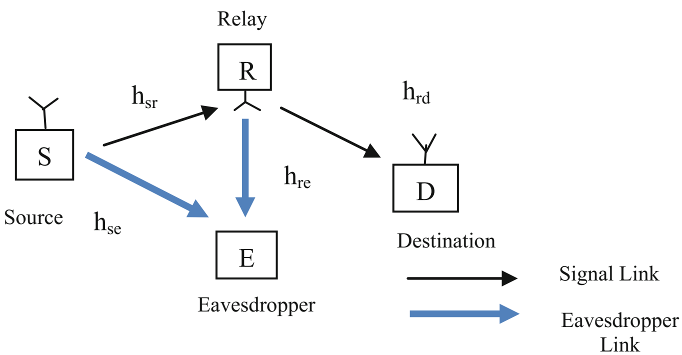


Fig. 1 Secrecy relay network having a source, a destination and an eavesdropper

better coverage and energy efficiency can be achieved using physical layer security with cooperative communication using EH techniques. The rest of the paper is organized as follows: Sect. 2 describes the various aspects of physical layer security using cooperation schemes; Sect. 3 explains about different EH techniques; Sect. 4 concludes the paper.

2 Physical Layer Security

Wireless technology is being increasingly used for a wide range of applications including cellular phones, banking, so it is important to secure wireless networks. In traditional methods, security had been implemented at higher logical layers of communication network instead of applying it to the physical layer. Data encryption is the simplest method for ensuring data confidentiality. However, it is not convenient in some emerging network architectures, e.g. ad hoc networks. So, the picture of ensuring secrecy using the physical properties of the radio channel or the physical layer came into existence. This idea was initiated by Shanon in and was extended by Wyner [2]. It emphasized the fact that the communication channel can contribute secrecy without sharing secret keys. In wireless communication, the signal is received by the intended user as well as by the eavesdropper. But if the signal transmission rate is approximately same as the capacity of the legitimate channel, then only the legitimate user can retrieve the message signal [6]. This is the principle behind the concept of physical layer security, where the secrecy capacity is determined by the difference of the channel capacities of the legitimate channel and the eavesdropper channel [1]. Secrecy rate is the maximum rate at which a source can transmit a message signal to a receiver through a reliable and secure link. It is generally expressed in terms of secrecy capacity which is given by the following expression [6]

$$C_s = [C_t - C_e]^+ \tag{1}$$

where $[x]^+ = \max(x, 0)$, C_t and C_e are the data transmission capacity and eavesdropper capacity, respectively.

A better secrecy capacity is obtained if the CSI is known to all the network nodes. But it is not possible to have perfect knowledge about the eavesdropper channel. In that case, where only receiver CSI is known, transmit jamming or artificial noise (AN) is preferred [7]. However, the transmit jamming utilizes transmit power which results in lower SNR at the receiver. Therefore, proper allocation of the power between information signal and jamming signal is significant. In general, to allocate power optimally between these two is quite difficult, so the best way is to allocate more power to the jamming signal provided: (i) perfect knowledge about receiver's CSI; (ii) more number of eavesdroppers. Transmit jamming schemes make use of multiple antennas, if there is only one transmit antenna then it is better to use receive jamming [8] in which the receiver sends the jamming signal to the eavesdropper. Another way to improve secrecy performance is to use beamforming techniques, in which the sender transmits the signal in the direction of receiver [9] and SNR at receiver is higher than that at eavesdropper. However, using above techniques either result in wastage of additional resources or complex computations. These impairments can be overcome by using diversity techniques [10]. As discussed in [10], different diversity techniques include MIMO diversity, multi-user diversity, and cooperative diversity. While MIMO and multi-user diversity have certain drawbacks [10], cooperative diversity is a better option where a relay is used to complete the transmission between the sender and the receiver. At destination, signal from direct transmission and relayed signal are added in constructive manner to decode [10] the message and simultaneously achieving security from the eavesdropper. The protocols used for processing can be categorized as fixed protocol and adaptive protocol [3]. Fixed protocol allots resources between source and relay in a fixed manner. Adaptive relay protocols comprise selective and incremental relaying. Among all these protocols, AF and DF are the most common. The various techniques for physical layer security is summarised in Fig. 2.

Traditionally, most wireless nodes operate in half-duplex (HD) mode due to easy implementation. This leads to 50% decline in the spectral efficiency because to transmit a data packet, two time slots are required [10]. Full-duplex (FD) transmission has gained much attention due to advancements in the self-interference (SI) cancellation techniques [11, 12]. In FD transmission, a wireless node transmits and receives simultaneously and has potential to improve physical layer security. The receiver node receives the message signal and simultaneously sends a jamming signal to the eavesdropper. The jamming signal affects signal both at the receiver and at the eavesdropper, but because of SI cancellation techniques used at the receiver it gets cancelled and a secure transmission takes place [11, 12]. The secrecy

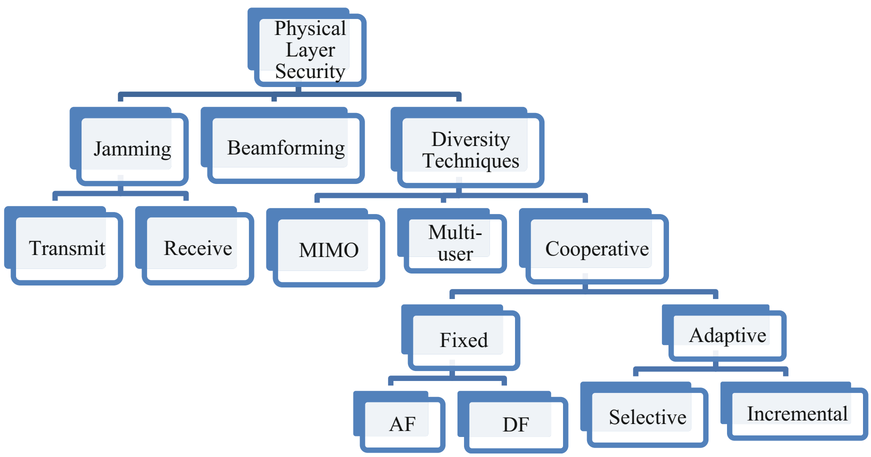


Fig. 2 Various techniques for physical layer security

performance of AF and DF cooperative schemes for a single-hop system operating under FD mode is compared in [13], and it was analysed that AF outperforms DF by 71.88% when relay and eavesdropper were 70 m apart.

3 Energy Harvesting

The combination of physical layer security and EH has been gaining considerable attention. On the one hand, physical layer security ensures reliable transmission of data; while on the other hand, EH techniques keep the energy constrained nodes active for relaying. SWIPT was first introduced in [4] for enhancing the efficiency of the wireless networks to utilize energy. The relays used for cooperative communication can process the received signal using two simple methods [14], i.e. EH with AF and EH with DF schemes. In both methods, the best relay is chosen to achieve the maximum secrecy rate. The evaluation results [14] show that the EH–DF protocol performs better than the EH–AF protocol. Practically, having the complete knowledge about the CSI of eavesdropper is not possible. The authors in [15] have considered an AF relay network consisting of colluding eavesdropper with imperfect CSI and proposed a joint cooperative beamforming (CB) and energy signal (CB–ES) to achieve secure and energy efficient communication. The results show that CB–ES outperforms other non-robust schemes and approaches the ideal scheme.

The secrecy capacity using SWIPT in FD mode is studied in [16]. It employs an FD Base Station (FD-BS) capable of simultaneous transmission and reception of information. An EH node is also used to scavenge energy from the RF signal. The energy profile is the main factor which determines how efficient the EH node is in the cooperative communication. In [17], it is found that the average rate at which each wireless node is scavenging energy from the power beacon affects the system performance. To decide whether a node is energy constrained, consider the following factors [17]: (i) average rate of harvesting energy, (ii) transmit power, (iii) total number of relays in the system.

There are two key challenges for the practical implementation of SWIPT. First, in [4, 18], it is assumed that both decoding and EH are performed simultaneously from the received signal, which is impractical. Second, the traditional information receiver architecture designed for wireless information transfer may not be suitable for SWIPT. However, two possible practical EH techniques are time switching (EH-TS) and power splitting (EH-PS) [20]. In EH-TS, the operation of receiver switches between the EH and information decoding. While in EH-PS, the received signal is splitted into two portions, one is for EH while other for information decoding. Authors in [19] demonstrate that a harvest-and-forward strategy-based antenna selection (AS) and PS techniques help the relay to harvest energy and receive the information signal simultaneously. The processed signal is relayed to the destination using the harvested energy. In [20], a self-back interference technique is used in which an artificial noise (AN) is transmitted by the destination to confuse the eavesdropper. This is a two-phase process:

- In the first phase, the source transmits the message signal while the destination transmits an AN signal, at the same time the relay harvests energy from these two received signals.
- In the second phase, these two signals are combined and the harvested energy is used to forward it, employing the EH–AF protocol.

Since the AN signal is known at the intended destination, so it gets cancelled out and a secure communication takes place. This two-phase communication protocol for single-antenna jammer and multi-antenna jammer case has been studied in [21], and results show that the throughput for multi-antenna jammer increases unbounded with the increase in transmit power. For proper operation of a system, the resources should be allocated optimally. The authors in [22] discussed the optimal allocation of resources in frequency domain AN-aided OFDM system with SWIPT to maximize the secrecy capacity of information receivers (IRs) by optimizing the transmit power splitting ratios. What if the AN cannot be cancelled out by the destination? In [23] the authors compared two cases where AN is used and other one where AN is used but not cancelled at the destination. The authors [23] proposed optimization of transmit power and power splitting ratios to minimize the secrecy outage probability (SOP) and maximize the average rate. However, with single-antenna transmitter, this technique is invalid. Cooperative jamming (CJ) [24] solves this issue by imitating the effects of multi-antennas where single/multiple helping nodes cooperatively generate artificial noise to confound the eavesdropper. Accumulate and Jam (AnJ), a new cooperative jamming protocol is presented in [25]. AnJ improves the physical layer security using a full-duplex friendly jammer using imperfect CSI. The jammer in [25] relies only on EH and works either in dedicated EH (DEH) mode or opportunistic EH (OEH) mode. The results show that AnJ protocol outperforms the conventional half-duplex schemes. Similarly, another two-phase protocol, namely Harvest and Jam (HJ), was proposed in [26] which employs cooperative jamming of the eavesdropper. In the first transmission phase, the information signal is transmitted from source-to- relay and at the same time sends power to a group of idle multi-antenna users which supports cooperative jamming in the network. During the second phase, the received signal is relayed to the legitimate receiver and the helping nodes provides secure transmission by sending jamming signals using the harvested energy. In [27], the authors compared the throughput of TS and PS Relaying protocols with delay limited transmission modes in a wireless network employing EH–AF scheme. The results are summarized in Table 1.

According to [20], the authors have also considered an ideal relay receiver (IRR) case and concluded that the secrecy capacity in EH-TS, EH-PS and IRR relaying protocols can be improved by (i) increase in number of relay antennas, (ii) increasing AN power, (iii) increasing source-to-relay distance and

Table 1 Comparison of EH-TS versus EH-PS

System parameter	Throughput of EH-TS versus EH-PS relaying
Noise variance	At low values, PS is better and vice versa at high values
Source-to-relay distance	PS performs better for small as well as large values
Transmission rate	For small values, PS is better and vice versa for large values
EH efficiency	For low efficiency, TS performs better than PS and vice versa for high efficiency

(iv) increasing source-to-eavesdropper distance. The comparison of various techniques used for EH is described in Table 2.

3.1 EH-TS Protocol

In TS relay protocol, say the time T is required for transmitting a certain block of message signal from source-to-destination [20]. For a period of αT, where $0 \leq \alpha \leq 1$, EH is done at the relay. Out of the remaining time period, half is used for Phase I process while the other half is used for the Phase II process as described in Fig. 3. An optimal value of EH time factor α should be selected to ensure high secrecy capacity. The secrecy rate of EH–AF and EH–DF protocol using TSR scheme for a single-hop relay network is compared with the traditional techniques in [28] and shows that the system with EH outperforms the latter. In [29], a block-wise TS protocol for EH–AF and EH–DF scheme has been proposed which performs better than the traditional fixed time duration protocols because by tracking the level of scavenged energy it smartly switches between the EH and IT phase. To deal with a system in which multiple relays are present between the sender and the receiver, two relay and antenna selection protocols have been proposed in [30]. The proposed optimal relay selection (ORS) and partial relay selection (PRS) schemes ensure the secrecy performance in terms of SOP. The results show that: (i) ORS performs better than PRS, (ii) ORS performs better with the increase in number of relays and power beacons.

3.2 EH-PS Protocol

As described in Fig. 4, in EH-PS the time duration T required for transmitting a certain block is divided into two halves [20]. During the first half, the EH and signal processing have been carried out by the relay. Specifically, a portion of power from the received signal, ρP, is utilized for EH and the remaining power, $(1-\rho)P$, is utilized for information processing, where $0 \leq \rho \leq 1$. In the second half slot, the

Table 2 Comparison of various EH techniques

Literature	Protocol used	System model	Channel	Objective	CSI
Son and Kong [14]	EH-PS relaying using AF and DF	Three node wiretap channel with M number of EH relays	Rayleigh	To maximize the achievable secrecy rate	Perfectly known at destination
Xing et al. [23]	EH-PS relaying	Three node wiretap channel with a relay	AWGN	To minimize SOP or to maximize average transmission rate	Known at transmitter
Liu et al. [21]	HJ using EH-TS relaying	One source, one destination and one multi-antenna jammer	Rayleigh	To maximize throughput subjected SOP constraint	Available at transmitter and receiver both
Zhang et al. [22]	OFDMA based SWIPT	One base station, K receivers, N subcarriers	Rayleigh	Optimal resource allocation to maximize secrecy rate	Available at base station
Nguyen et al. [30]	EH-TS relaying using DF	Three node wiretap channel with K relays and one multi-antenna power beacon	AWGN	To enhance SOP and to maximize EH at source and relay using PRS and ORS	Full and partially known at source
Hamdi et al. [20]	EH-TS, EH-PS and IRR	Three node wiretap channel with multi-antenna relay	Rayleigh	Comparison of different EH techniques	Full knowledge at relay
Bi and Chen [25]	AnJ EH-TS relaying consisting DEH and OEH	Three node wiretap channel	Both Rayleigh and Rician	To minimize SOP and obtain a nonzero secrecy capacity	Known perfectly at source while imperfectly at jammer
Feng et al. [15]	Joint CB–ES using AF relaying	Three node wiretap channel with M eavesdroppers and K relays	AWGN	To maximize secrecy rate	Estimated CSI at eavesdropper
Sinha and Jindal [28]	EH-TS relaying using AF and DF	Three node wiretap channel with one relay and one power beacon	AWGN	To enhance secrecy rate over conventional techniques	N.A.
Jindal and Sinha [31]	EH-PS using AF and DF	Three node wiretap channel with one relay and one power beacon	AWGN	To maximize secrecy rate	N.A.
Zhu et al. [32]	EH-PS relaying	MIMO system with transmitter, receiver and energy harvesters	Rayleigh	To minimize transmit power	Both perfect and imperfect knowledge

Fig. 3 Structure of time frame for EH-TS relay protocol

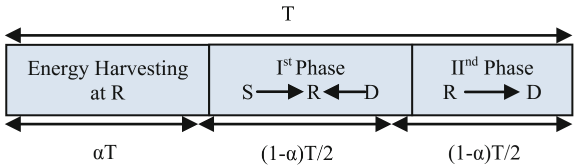
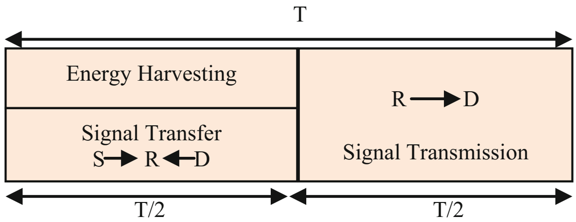


Fig. 4 Structure of time frame for EH-PS relay protocol



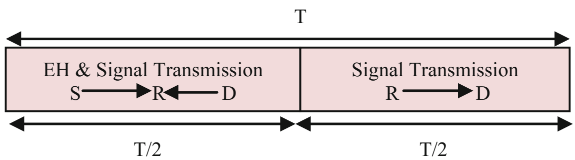
message signal is relayed to the legitimate receiver using the harvested energy in phase I.

Like TSR, PSR also requires optimal value of power splitting factor ρ. In [31], comparison of the secrecy performance of EH–AF and EH–DF protocols using power splitting scheme for single-hop wireless relay network shows that the EH–AF protocol performs better than EH–DF. In a MIMO system which consisting of IRs and energy receivers (ERs), and ERs may act as eavesdropper. For this, the authors in [32] proposed a transmit beamforming scheme to minimize power for a multi-user MIMO SWIPT communication system using PSR scheme and an AN injection scheme to secure the information.

3.3 *Ideal Relay Receiver*

Considering the ideal case, the IRR system can simultaneously process the signal as well harvest energy from the same received signal [16]. The time duration T is splitted into two equal portions. During the first half, the relay performs EH and processes the received signal, while in the second half the relay uses the harvested energy to transmit the received signal to the destination (Fig. 5).

Fig. 5 Structure of time frame for IRR protocol



4 Conclusion

In this paper, a study on physical layer security with cooperation schemes using EH techniques is presented. Physical layer security when implemented with cooperative communication and EH techniques offers a secure, reliable and energy efficient communication. The concept of physical layer security with cooperative communication using EH schemes can be beneficial for the upcoming generation of wireless networks. The literature available in this field employs only fixed cooperative protocols, i.e. AF and DF. For future work, the EH schemes must also exploit the adaptive cooperative relay protocols. Moreover, the fixed protocols mainly considered the single-hop systems. Multi-hop scenarios are a more practical approach and need to be explored more.

References

1. Liu, S., Hong, Y., Viterbo, E.: Artificial noise revisited. IEEE Trans. Inf. Theor **61**(7), 3901–3911 (2015)
2. Van Der Meulen, E.C.: Three-terminal communication channels. Advanc. Appl. Probab. **3**(1), 120–154 (1971)
3. Liu, K.R.: Cooperative Communications and Networking. Cambridge University Press (2009)
4. Varsney, L.R.: Transforming information and energy simultaneously. In: Proceedings of IEEE International Symposium on Information Theory (ISIT), July 2008, pp. 1612–1616 (2008)
5. Biason, A., Laurenti, N., Zorzi, M.: Achievable secrecy rates of an energy harvesting device. IEEE J. Sel. Areas Commun. **34**(5), 1502–1517 (2016)
6. Chen, G., Gong, Y., Xiao, P., Chambers, J.A.: Physical layer network security in the full-duplex relay system. IEEE Trans. Inf. Forensics Secur. **10**(3), 574–583 (2015)
7. Cumanan, K., Xing, H., Xu, P., Zheng, G., Dai, X., Nallanathan, A., Karagiannidis, G.K.: Physical layer security jamming: theoretical limits and practical designs in wireless networks. IEEE Access **5**, 3603–3611 (2017)
8. Gollakota, S., Katabi, D.: Physical layer wireless security made fast and channel independent. In: 2011 Proceedings of IEEE on INFOCOM, pp. 1125–1133. IEEE (2011)
9. Le, T.A., Vien, Q.T., Nguyen, H.X., Ng, D.W.K., Schober, R.: Robust chance-constrained optimization for power-efficient and secure SWIPT systems. IEEE Trans. Green Commun. Netw.
10. Zou, Y., Zhu, J., Wang, X., Leung, V.C.: Improving physical-layer security in wireless communications using diversity techniques. IEEE Netw. **29**(1), 42–48 (2015)
11. Hong, S., Brand, J., Choi, J., Jain, M., Mehlman, J., Katti, S., Levis, P.: Applications of self-interference cancellation in 5G and beyond. IEEE Commun. Mag. **52**(2), 114–121 (2014)
12. Bharadia, D., McMilin, E., Katti, S.: Full duplex radios. ACM SIGCOMM Comput. Commun. Rev. **43**(4), 375–386 (2013)
13. Sinha, R., Jindal, P.: Performance analysis of cooperation schemes under total transmit power constraint in single hop wireless relaying system. In: IEEE International Conference on CCIS (2016)
14. Son, P.N., Kong, H.Y.: Cooperative communication with energy-harvesting relays under physical layer security. IET Commun. **9**(17), 2131–2139 (2015)

15. Feng, Y., Yang, Z., Zhu, W.P., Li, Q., Lv, B.: Robust cooperative secure beamforming for simultaneous wireless information and power transfer in amplify-and-forward relay networks. IEEE Trans. Veh. Technol. **66**(3), 2354–2366 (2017)
16. Wang, Y., Sun, R., Wang, X.: Transceiver design to maximize the weighted sum secrecy rate in full-duplex SWIPT systems. IEEE Signal Process. Lett. **23**(6), 883–887 (2016)
17. Medepally, B., Mehta, N.B.: Voluntary energy harvesting relays and selection in cooperative wireless networks. IEEE Trans. Wirel. Commun. **9**(11), 3543–3553 (2010)
18. Grover, P., Sahai, A.: Shanon meets Tesla: wireless information and rate-energy tradeoff. IEEE Trans. Commun. **61**(11) (2013)
19. Zhou, Z., Peng, M., Zhao, Z., Li, Y.: Joint power splitting and antenna selection in energy harvesting relay channels. IEEE Signal Process. Lett. **22**(7), 823–827 (2015)
20. Salem, A., Hamdi, K.A., Rabie, K.M.: Physical layer security with RF energy harvesting in AF multi-antenna relaying networks. IEEE Trans. Commun. **64**(7), 3025–3038 (2016)
21. Liu, W., Zhou, X., Durrani, S., Popovski, P.: Secure communication with a wireless-powered friendly jammer. IEEE Trans. Wirel. Commun. **15**(1), 401–415 (2016)
22. Zhang, M., Liu, Y., Zhang, R.: Artificial noise aided secrecy information and power transfer in OFDMA systems. IEEE Trans. Wirel. Commun. **15**(4), 3085–3096 (2016)
23. Xing, H., Liu, L., Zhang, R.: Secrecy wireless information and power transfer in fading wiretap channel. IEEE Trans. Veh. Technol. **65**(1), 180–190 (2016)
24. Chen, X., Zhong, C., Yuen, C., Chen, H.H.: Multi-antenna relay aided wireless physical layer security. IEEE Commun. Mag. **53**(12), 40–46 (2015)
25. Bi, Y., Chen, H.: Accumulate and jam: towards secure communication via a wireless-powered full-duplex jammer. IEEE J. Selected Topics Signal Proces **10**(8), 1538–1550 (2016)
26. Xing, H., Wong, K.K., Chu, Z., Nallanathan, A.: To harvest and jam: A paradigm of self-sustaining friendly jammers for secure AF relaying. IEEE Trans. Signal Process. **63**(24), 6616–6631 (2015)
27. Nasir, A.A., Zhou, X., Durrani, S., Kennedy, R.A.: Relaying protocols for wireless energy harvesting and information processing. IEEE Trans. Wirel. Commun. **12**(7), 3622–3636 (2013)
28. Sinha, R., Jindal, P.: A study of physical layer security with energy harvesting in single hop relaying environment. In: 2017 4th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 530–533. IEEE (2017)
29. Nasir, A.A., Zhou, X., Durrani, S., Kennedy, R.A.: Wireless-powered relays in cooperative communications: time-switching relaying protocols and throughput analysis. IEEE Trans. Commun. **63**(5), 1607–1622 (2015)
30. Nguyen, N.P., Duong, T.Q., Ngo, H.Q., Hadzi-Velkov, Z., Shu, L.: Secure 5G wireless communications: a joint relay selection and wireless power transfer approach. IEEE Access **4**, 3349–3359 (2016)
31. Jindal, P., Sinha, R.: Energy efficiency and secure communication with power splitting energy harvesting technique for single relay network. In: 2017 International Conference on Computer and Applications (ICCA), pp. 215–219. IEEE (2017)
32. Zhu, Z., Chu, Z., Wang, N., Huang, S., Wang, Z., Lee, I.: Beamforming and Power Splitting Designs for AN-Aided Secure Multi-User MIMO SWIPT Systems (2017). arXiv:1703.04711

One-Tip Secure: Next-Gen of Text-Based Password

Zainab Zaheer, Aysha Khan, M. Sarosh Umar and Muneeb Hasan Khan

Abstract As this century marked the beginning of new era in the world of Internet, the need for user authentication system becomes a basic requirement to ensure secure communication with confidentiality and integrity. With the development of authentication systems, the techniques for intruding are also developed by hackers in parallel. So, in order to prevent or at least minimize the attacks, many encryption techniques and authentication schemes also came into existence. This paper also describes one such novel scheme using text-based authentication. Firstly, the present scenario in the context of authentication system and different methods of hacking and breaking into the system are demarcated. Then proposed scheme is mentioned along with the discussion of various features and security issues exposed by this scheme. This scheme will be a time-saving and efficient alternative on measures of user-friendliness and convenience to current authentication system. It can be used in bank ATMs, and other public and private organizations. The aim is to provide fast and efficient method at user convenience so that it can be deployed in various activities relating to common users.

Keywords Authentication · Information · Hackers · Confidentiality
Security · Integrity

Z. Zaheer (✉) · A. Khan · M. Sarosh Umar · M. H. Khan
Department of Computer Engineering, Zakir Husain College of Engineering and Technology, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India
e-mail: zainab.zaheer@zhcet.ac.in

A. Khan
e-mail: ayshakhan@zhcet.ac.in

M. Sarosh Umar
e-mail: saroshumar@zhcet.ac.in

M. H. Khan
e-mail: muneebhkhan@zhcet.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_24

1 Introduction

Whenever there is a communication between two individuals, the first and foremost requirement is authentication. That is, sender has to identify its receiver correctly when confidential information is being exchanged during communication. The common cited situations for secret communication in past century were love affairs, wartime communications, and business transactions. But now in the time of Internet of Things, when everything is getting digitized, it is inevitable to deny secure communication. That's what is provided by cyber and network security in the present world. The desirable properties of secure communication are confidentiality (only sender and receiver understands the contents of message), message integrity (unaltered contents), end-point authentication (identity of receiver and sender) and operation security (content reside within an organization, if any) [1]. The point of discussion in this paper is end-point authentication.

In past, Sentries used to create some kind of signals for authentication of their kinsmen. For face-to-face communication, this can be done easily by visual recognition. In cyber security, it can be easily done by authenticating the users with an id and a password because of the fear of intruders. The face-to-face communication may be eavesdropped; secure communication in cyber world may be hacked. Now the technology is advancing and the methods of authentication are becoming more and more sophisticated, and the hackers have also increased their expertise level. In today's world, the most pronounced authentication techniques include mainly three approaches—knowledge-based approach, token-based approach and biometric approach. The techniques for hacking that challenge these approaches include spoofing, shoulder surfing, social engineering, key logging, brute force approach.

Knowledge-based authentication schemes are the most feasible and henceforth are popularly deployed in many systems that may be either text-based or graphical password. Though it is most convenient to remember the password by user for one account, it becomes cumbersome when users tend to have multiple accounts. Also, these techniques are more prone to shoulder surfing, spyware attacks, hidden camera, social engineering, dictionary attack and brute force attacks, etc. Hence to overcome this shortcoming, there is a need for a change in the technique without changing the traditional method of logging in through passwords. Hence, we are proposing an alternative to the current scheme of text-based passwords.

The rest of the paper is organized as follows. Section 2 sets the background for user authentication. It discusses different methods of authentication and bearable attacks on them. In Sect. 3, the proposed scheme is described. This is followed by a detailed analysis and discussion on various aspects of the scheme in next section. Finally, paper is concluded telling about the future work.

2 Background

The science of user authentication in network security and potential attacks on the system develops simultaneously. So, here we discuss different methods along with tolerable attacks on them.

2.1 Methods of Authentication

These methods of authentication are divided into three main classes [2]. These are:

Token-based authentication assigns a unique token to users, and the process of authentication requires the submission of that token at every login. The main examples are ATMs, DND Flyways membership card, credit cards, airport parking, shoppers stop membership card.

Biometric-based authentication makes use of the physical characteristics of a human being for authentication. It establishes a unique identity of every user according to their features, such as retina scan, thumb scan, fingerprint. These are widely deployed in examination halls, Aadhaar ID verification, property dealing agreements, etc.

Knowledge-based authentication includes both text- and graphics [3]-based passwords. Here the user first registers with a unique id and password of their choice and later uses it for further logins. It is the most common technique in today's world and is used in almost every possible domain. These types of passwords can be seen in social networking websites, mailing websites or college databases, and many more.

2.2 Common Attacks on Text-Based Passwords

The most common attacks [4] on text-based passwords are –

Dictionary Attack In this attack, the hacker may try to systematically enter every combination of password.

Brute Force Attack Brute force attack is a trial-and-error method to decode encrypted data such as passwords using Data Encryption Standard through exhaustive effort [5].

Shoulder Surfing In this attack, the intruder may spy from the users back when he/she is logging into the password [6].

Hidden Camera In this, there is an attempt to spy the password by putting the hidden cameras in ATMs, Cyber Cafes, etc.

Key Logging (Spyware Attack) This can be done either manually by looking from the back or through some kind of key logging software.

Guessing The hacker may simply guess the password. It will be easy if the hacker knows the user personally and also knows the length of password.

Social Engineering Social Engineering consists of guessing the password through the user's social circle. It is either by scanning his profile on social networking websites or personally being social to the user.

3 Proposed Scheme

3.1 System Overview

Our system comprises a module for logging into the system. This lets the user to register with necessary credentials such as name, date of birth, genuine email ID and password of at least 12 characters or more consisting of any combination of alphabets, numbers and special characters according to user convenience. This has no restriction on database used in background for storing the data of registered users which states that when user logins next time, he doesn't need to enter his full password. Instead he will be provided with a captcha, in which he has to search for the occurring of any character abbreviating the month as follows.

Then, he has to enter that character of his password which is at month's position in the location where the abbreviation for month had occurred along with dummy values at other places. The above screen shown in Fig. 1 appears as soon as user submits the registration form. And the captcha may looks like as in Fig. 2.

This captcha contains only one of the reserved characters by our scheme at a time ensuring no repetition. So that user never gets confused with ambiguity.

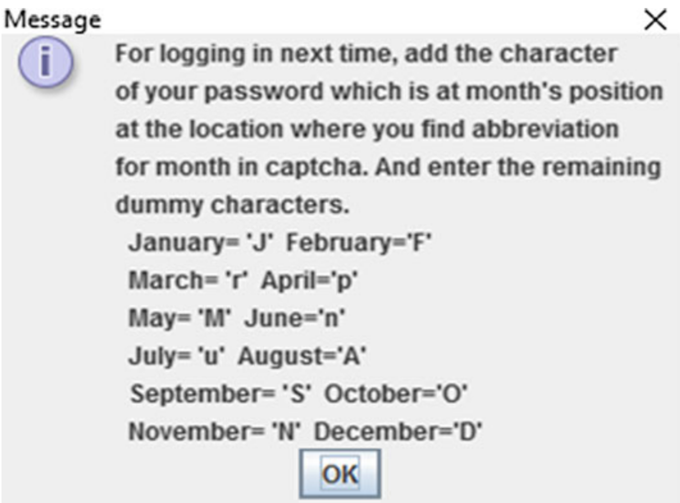


Fig. 1 Abbreviation of months according to our scheme

Fig. 2 Generated captcha

3.2 Testing the System

Suppose a user sets his/her password as “aishazainab1” during registration along with other necessary details as required in the form shown in Fig. 3. Now, for logging into the system (Fig. 4), he has to enter the registered username. But in password field, he doesn’t need to enter the actual password. Instead he has to enter a dummy password which contains as many as dummy characters he wants of his choice except only 1 variable in a particular position. For the generated captcha in Fig. 2, he encountered “n” at fourth position. In our scheme’s coding, “n” abbreviates as “June”. Since June is the sixth month of year, so, user has to enter the sixth character of his password at fourth position. In this case, it is “z”. If the user is able to enter password according to this scheme only, then he will be directed to “Login Successful”.

This is a sample registration screen in our scheme. Other fields may be added according to the need of organization installing our scheme. On submission of values for registration, user is directed to the screen shown in Fig. 1.

Figure 4 is login screen, which shows the generated captcha. And user is asked to enter the registered username as he knew the scheme from registration time. So, he doesn’t need to enter the exact password instead he makes use of captcha to

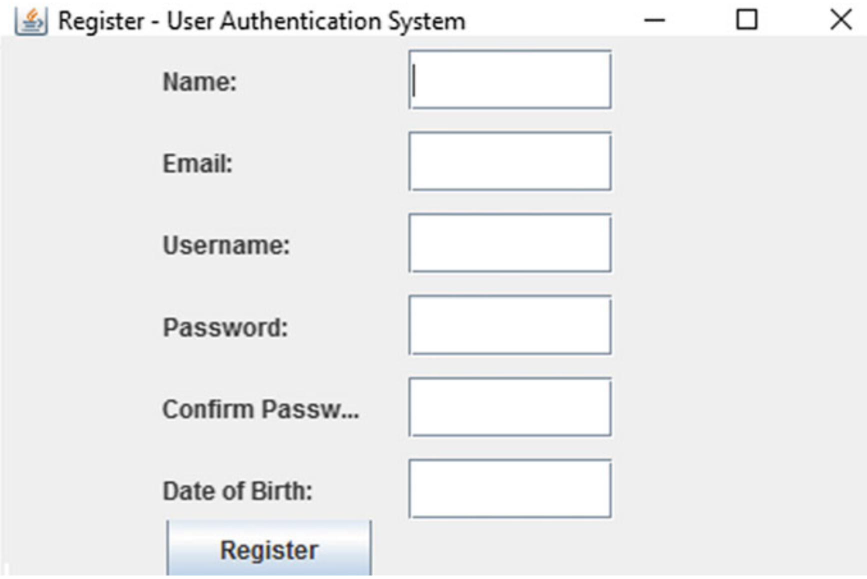


Fig. 3 Registration screen

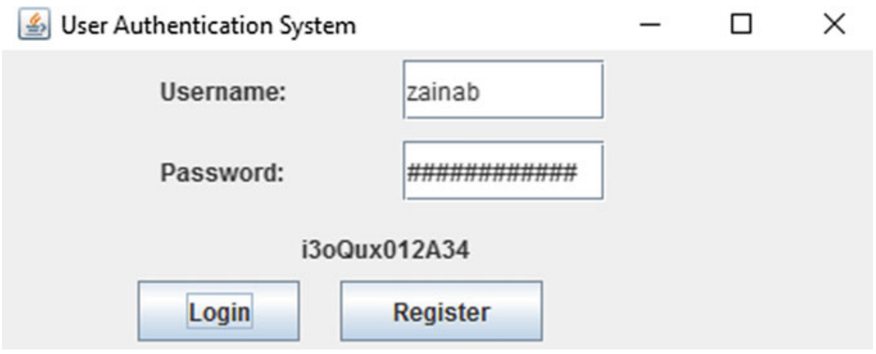


Fig. 4 Login screen

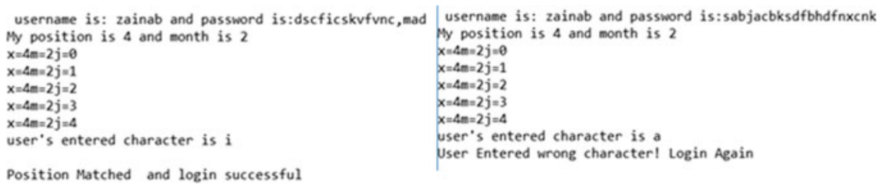


Fig. 5 Background screen when user entered correct and incorrect character, respectively

find the location and character of his password and if it matches, the user is logged in successfully. In case, if he is a trespasser, or if he forgets the scheme, he is asked to login again. Figure 5 shows background processing of two cases of our scheme, where left part shows the case of logging in successfully and right part shows invalid login.

4 Analysis and Discussion

This scheme was tested on a set of 20 common people which mostly work in non-technical areas and have minimal knowledge and acquaintance with computers. The users were first asked to authenticate themselves by filling the registration form as shown in Fig. 3. And then their time for three consecutive logins (Fig. 4) was noted. These experimental values are then plotted into the column chart (Fig. 6).

Users were allowed to set the passwords of their own choice without any restriction. They can make use of any combination of characters, symbols or numbers. Then, on the basis of the values, we had made the following observations.

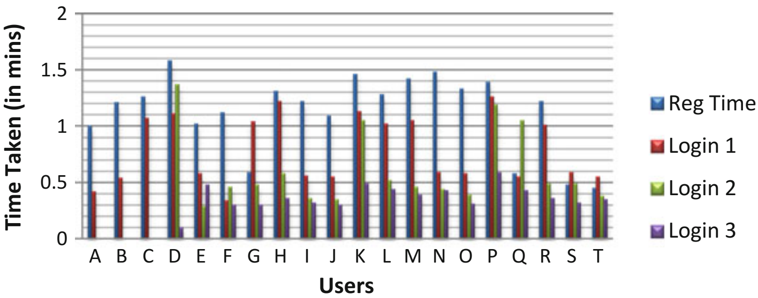


Fig. 6 Experimental chart

4.1 User-Friendliness

Since users have no restriction for forming the typically strong password, it solely depends on user convenience to remember their passwords. Also, the registration and login time depend on the typing speed of user.

It can be easily depicted from the graph that time taken in third login decreased to a considerable extent, which shows that once the user get used to, then login time depends mainly on their own typing speed and memory.

4.2 Efficiency

Though the time taken depends on an individual, a keen view on graph shows constancy in time in third login. It shows that most of the users' time for login is nearly about 30 s. This can also be seen from the graph that login time in consecutive attempts falls at an exponential rate. So, this makes an efficient scheme as it gets more popular, users will be able to login within quite less time.

4.3 Password Space

The user does not need to enter his full password, rather he can enter dummy values except at a single position which is manipulated by the captcha in our scheme. This makes its password space reasonably good. It is given by $S = A^N$ where N = No. of characters in password and A is the set of characters allowed to be used as password. So, here password can be of 12 characters and alphabet range is of 96 characters, so our password space is about 96^{12}.

4.4 Attacks Resistant

It is also noted that passwords are safe from shoulder surfers, key loggers, and other spyware or hidden camera attacks. This was so because the users are entering many dummy characters and the password characters are summoned up in their minds to match the position with captcha. Since the same user would be entering any random different characters in every logins, it would be tedious for hackers to crack the password through dictionary attack or brute force attacks.

Hence, this scheme will prove to be an efficient alternative on user convenience for secure and confidential communication.

5 Conclusion

Our project can be used for any application requiring authentication of users. These application areas may range from simple software in different fields requiring login to more confidential system as in banks, military system, etc. This application is more user-friendly as compared to existing methods. This eliminates the need to remember the passwords for multiple accounts. The user authentication scheme has many advantages over the existing methods of knowledge-based passwords in our cyber security. Some of them are listed below:

- First of all, this scheme is user-friendly as the user won't take much time to login. They can login in the minimum stipulated time as that much required in their traditional method.
- Secondly, it will be easier to remember and use as there is no need to deploy additional resources for this task. Further our system is cheaper than the existing schemes too.
- Thirdly, this method will be much more efficient and less erroneous because it eliminates the key logging and hidden camera attacks. Text-based passwords are more prone to attacks than graphical or biometric, and in this system most of the possible attacks will be avoided.

It is much more convenient because many of the passwords are prone to dictionary attack or may be hacked by social engineering. Thus, it is protected method for both user and organization.

6 Future Work

Our project can be further expanded. We can incorporate hashing technique in determining the month and position which also extends the length of password. This can increase the password space to a considerable extent. Thus, our system will be more secure and robust. It also ensures its safety against dictionary and brute force attacks as they are directly related to the length of password.

This project can also be used in many other services rather than only in bank ATMs. For example, it can be used in colleges, private organizations, such as MNCs and much more. We believe our project has a lot of potential in benefitting humanity.

References

1. Kurose, J.F., Ross, K.W.: Computer Networking—A top down Approach. 6th Edn. (2013)
2. Kushwaha, B.K.: An approach for user authentication one time password (numeric and graphical) scheme. J. Global Res. Comput. Sci. **3**(11) (2012)
3. Saeed, S., Umar, M.S.: A hybrid graphical user authentication scheme. In: Proceedings of Communication, Control and Intelligent Systems (CCIS). IEEE Nov 7–8, 2015
4. Van Oorschot, P.C., Wan, T.: Two step: an authentication method combining text and graphical passwords. School of Computer Science, Carleton University, Ottawa, Canada
5. http://searchsecurity.techtarget.com/definition/brute-force-cracking (2017). Last accessed 9 Oct 2017
6. Zhao, H., Li, X.: S3PAS: A Scalable Shoulder- Surfing Resistant Textual-Graphical Password Authentication Scheme. Scalable Software Systems Laboratory, Department of Computer Science, Oklahoma State University, USA (2007)
7. Umar, M.S., Khan, I.A., Maqsood, T., Ali, J.: A novel two stage graphical password scheme. In: IEEE-Sponsored National Conference on Emerging Trends in Electronics and Communication Engineering, Sept 14–15, 2006, Goa, India (2006)
8. Umar, M.S., Rafiq, M.Q., Ansari, J.A.: Graphical user authentication: a time interval based approach. In: Proceedings of 2012 IEEE International Conference on Signal Processing, Computing and Control (2012 IEEE ISPCC) Jaypee University of Information Technology, March 15–17, 2012, Solan, HP, pp. 204–210 (2012)
9. Rajavat, R., Gala, B., Redekar, A.: Textual and graphical password authentication scheme resistant to shoulder surfing. Int. J. Comput. Appl. **114**(15) (2015)
10. Styaq, S., Agrawal, L.: A new technique for user authentication using numeric one time password scheme. Int. J. Comput. Sci. Eng. **4**(5) (2016)

A Technique to Envisage Investigative Codes for Enduring Maladies Using Data Mining Techniques

Vinitha Dominic, Shaleen Bhatnagar and Garima Goyal

Abstract Data mining (DM) is a new methodology to data investigation and knowledge discovery The goal of data mining is to gain novel and deep insights and of large datasets which can then be used to support decision making. The information gained can be used for applications ranging from business management, production control, and market exploration to emerging design and science exploration and medical data analysis. Data mining techniques can be applied to medicare domain to catalyze and support goals like avoiding clinical diagnostic tests, finding adverse drug reactions, reducing hospital acquired infections, and rooting out fraud. The goal is to derive life rescuing information which assures better health for the society. Data mining techniques can be used to accomplish this, by using effective analytic tools to discover hidden relationships and trends in medicare data. This paper focuses on studying the enduring maladies like kidney maladies, osteoporosis, arthritis from the medicare dataset. The relation among the enduring maladies and the conforming investigative codes are analyzed by using several data mining (DM) techniques. Then, active conclusion on the various diagnosis tests for each of the chronic maladies is defined, by acknowledging its clinical relevance.

Keywords Medicare data · ICD9 codes · Data mining

1 Introduction

There is a growing trend of applying data mining techniques on medical data. The medicare domain generates exchanges and stores patient-specific data. Focused efforts to make medicare inexpensive and more convenient without conceding on

V. Dominic (✉) · S. Bhatnagar · G. Goyal
Presidency University, Bengaluru 560064, India
e-mail: vinithadominic@presidencyuniversity.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_25

the quality of medicare service were made. The other issues were to condense the number of clinical trials and moderate the medicare expense of an individual, and the necessity to address these issues has become an essential requisite for the society.

Over the past decades, there is an emergent trend of enduring maladies such as osteoporosis, depression, stroke, arthritis. This is due to the change in lifestyle and nature of diet. These maladies are diagnosed through various tests and by monitoring different control factors such has blood pressure, calcium density. This is one of the significant reasons for high expense of medicare for an individual having chronic maladies. Prototypes should be derived to condense the investigative tests for chronic maladies which will significantly decrease the medicare cost.

One of the feasible solutions is to use DM on the medicare data. In recent years, data mining techniques were introduced, and they have the objective of finding the recurrently occurring patterns and derive meaningful acquaintance from it and design prediction models. Data mining has three categories of algorithm such as classification, association, and clustering. Each of these has characteristic set of rules when applied on medicare data [1]. These techniques will help derive substantial acquaintance from the medicare data useful for the benefit of the society.

In this paper, the goal is to study about the six enduring maladies namely Alzheimer's, obsessive pulmonary disorder, osteoporosis, depression, arthritis, and stroke based on its investigative codes. Then, various DM techniques like attribute reduction techniques using the Genetic Search algorithm and classification algorithms were used to obtain the reduced set of investigative codes. The paper concludes by distinguishing the relevant and significant codes.

2 Related Work

Research on medicare data has been carried out in varied ways. Before applying the data mining techniques, efforts were made to normalize the structure of medicare data; i.e., an electronic health record prototype was defined which makes the data easy to use and analyze [2]. Then, research was done on the enduring maladies like cancer, osteoporosis, and cardiac disorders independently using DM techniques. Classification techniques like SVM were used for early detection of osteoporosis [3]. Classification techniques like regression, decision tree, and neural networks were used for renal maladies prognosis [4]. Techniques like perceptron networks were used for perceiving the likelihood of cardiac maladies [5]. Similarly approach for condensing the attribute set for a particular maladies, this was applied on cardiac maladies was analyzed [6, 7]. Association rules were used for medicare data mining [8]. These rules were applied to derive acquaintance different types of maladies [9]. Major part of the research focussed on understanding the data mining techniques that can be used on an enduring malady, to reduce the attributes, for probability of occurrence [10–12]. After exhaustive study of these papers, analysis was done on the cardiac maladies dataset using InfoGain and Genetic Search algorithms for attribute selection and these attributes were then validated using six classifiers. The analysis resulted Adaboost classifier exhibiting significantly good performance [13]; thus, it can be used for better analysis and effective predictions. Another noteworthy observation was that UCI repository was commonly used for the medicare data analysis. There are many other repositories available. One of the popular repositories is CMS, and this paper uses it for the experiments.

Exhaustive research was done on the CMS dataset. This dataset was used to predict the medicare expenses of an individual based on the various medicare conditions [14]. Then, efforts were made for diagnosis code summarization by applying clustering techniques [15], and the Naïve Bayes and greedy stepwise algorithms were combined for different parameters of the data [16]. Another remarkable work on predicting the severity of cancer using the covariates was carried out using the same dataset. Then, research on allocating standard medical codes to patient records was done using the CMS data. Each patient record is correlated to specific maladies using several codes/labels. Comparison and analysis of algorithms for diagnosis coding on a large patient dataset was done [17–20]. Lot of queries on lack of exploration of information available about the various enduring maladies and the various procedure codes and diagnosis that lack exploration. The association between chronic maladies and the investigative codes was not studied for deriving useful acquaintance. The focus of this paper is to apply data mining techniques on the CMS data to derive the association between investigative codes and the enduring maladies, validate the relevance and significance of investigative codes, and derive acquaintance which facilitated in effective implication of these diagnosis trails for each malady by validating the anatomic relevance.

3 Data Description

The experiments for this paper are performed suing CMS dataset. The types of files in the dataset—the Beneficiary File, the Outpatient Claims (OP), the Inpatient Claims (IP), the Carrier Claims, and the Prescription Drug files (PDE) for the year 2008–2010. An individual who submits the claim is called a beneficiary. Every beneficiary is identified using a distinct id called Desynpuf_id. The presence of chronic maladies for a beneficiary is assigned a value 1, else the value will be 2. The official system for assigning codes to procedures and diagnosis is ICD-9-CM [16]. There codes are 2, 3, and 4 digits. The decimal digit for these codes is used for more specification. The dataset manual gives a better understanding about the medicare data. The Beneficiary files, OP files, IP files, and ICD9 codes are used for the experiment.

4 Projected Approach

The representation in Fig. 1 gives the comprehensive design of the projected approach. The projected approach consists of training (to learn and derive knowledge) as well as testing (to apply the derived knowledge) phase. The training phase includes the following stages.

4.1 Data Categorization

The dataset is reorganized, by correlating the beneficiary dataset, outpatient dataset, and inpatient dataset by using the distinct key, i.e., the beneficiary id. For each of the enduring maladies, the corresponding ICD9 codes are extracted. This helps in understanding and analyzing the count of diagnostic codes and their correlation used for each malady.

4.2 Data Mining

Based on the enduring maladies, the dataset is divided into samples. If a beneficiary has an enduring malady the class/target (C1), variable is set to 1, and it is set to 2 for

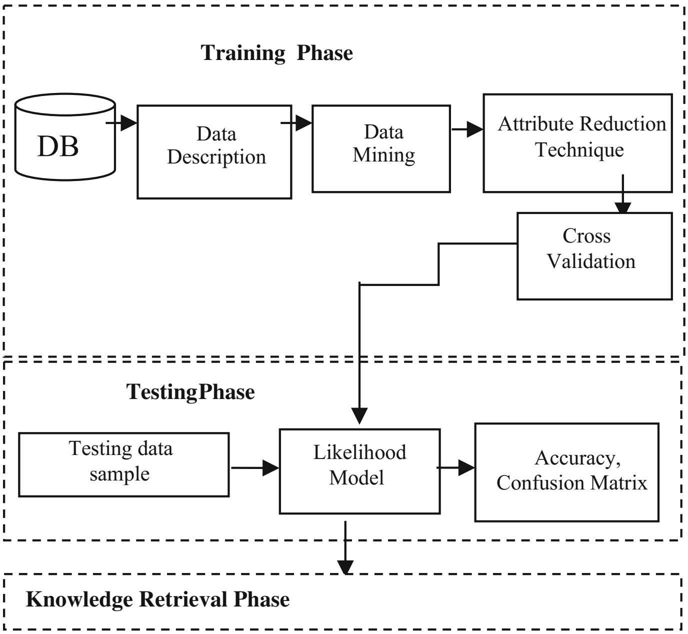


Fig. 1 Proposed design

healthy beneficiary (C2). The distinctive ICD9 codes from the inpatient and outpatient files were identified, and the count of occurrence of these codes was calculated; N codes were filtered for the experiment.

4.3 Resampling and Attribute Reduction Techniques

The probability of bias in the dissemination of healthy and unhealthy beneficiaries is high for medicare data; to avoid this dissemination, the resample filter was used. This filter ensures uniform class dissemination by restructuring the data. After this attribute reduction technique, InfoGain and Genetic Search are used.

4.4 Clasificación

Clasificación is done using Adaboost classifier. The validation is done for n folds. The performance metric is accuracy, i.e., the number of correctly classified instances. The output of the training phase is a likelihood model, which is tested in the data test phase.

4.5 Data Test Phase

Data from a different sample is validated. The procedure of data extraction and data definition is applied. This data sample is validated using the likelihood model based on accuracy, i.e., correctly classified instances. This procedure is repeated for each of the enduring maladies mentioned in the previous section.

4.6 Acquaintance Retrieval Phase

This phase uses the likelihood prototype and the condensed set of investigative codes for each malady as inputs. The model is studied by using the performance metric accuracy and confusion matrix for the five enduring maladies. This phase will analyze and study the condensed set of investigative codes for the five enduring maladies, identify the common set of investigative codes, and finally explore its anatomic relevance.

5 Investigational Setup

The investigational setup has several stages. The data is redefined, extracted based on the top N codes. Attribute reduction technique used is InfoGain and Genetic Search, and testing is done using tenfold cross-validation.

The data statistics for each of the enduring maladies for training and testing phase is shown in Table 1. From the table, it is implicit that osteoporosis has the maximum count of instances because it is the most commonly observed among the youth due to the lack of calcium intake because of poor diet habit. Next is depression, which is also major emerging enduring malady due to the lifestyle which leads to excessive stress.

Data mining techniques were applied to this data by using a software tool called Weka. It provides set of data mining algorithms which help in carrying out data mining tasks.

6 Result Analysis

The extracted data is subjected to attribute reduction technique. After applying these techniques, the reduced attribute set obtained were validated using the Adaboost classifier using cross-validation. Figure 2 gives the accuracy of the classifier with the condensed set of attributes for the training and testing dataset. It is perceived that the likelihood model of the proposed approach gives accuracy above 90%. It is observed that the highest accuracy for classification is recorded for Strk. It is perceived from Table 3 that during the testing phase, most of the enduring maladies record accuracy between 80 and 90%. The accuracy for the enduring maladies is between 75 and 79%, because for diagnosis of depression the nature of tests vary for each beneficiary, it depends on an individual's state of mind. After obtaining the condensed set of investigative codes for each enduring maladies, the next goal was to analyze it. The ICD9 codes along with the specification for the condensed set of attributes obtained after applying GST are given in Table 2.

From the table, it can be perceived that the number of reduced attributes is 3, because during repeated iterations it was observed maximum accuracy was

Table 1 Training and testing data statistics

Maladies	Training	Testing
Alzhdmta	5854	450
Copd	4234	453
Osteoprs	8987	462
Depressn	6779	767
Ra_oa	3967	908
Strketa	3100	713

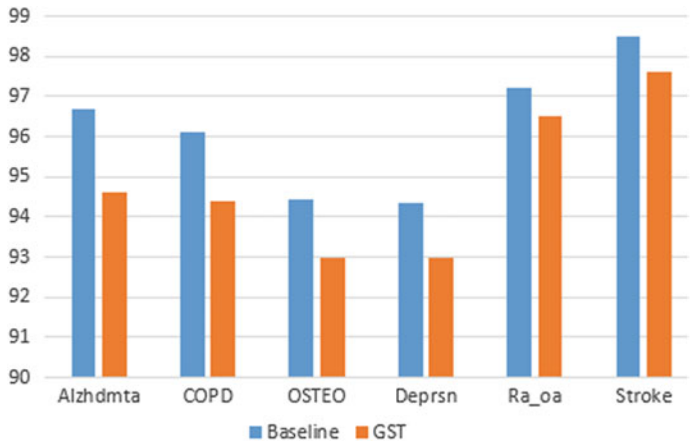


Fig. 2 Classification accuracy for condensed set of attributes on training and testing

Table 2 Condensed list of ICD9 codes for chronic maladies after applying attribute selection

Maladies	Condensed set of attributes using IG
Alzheimer's	331, 294, 401
Compulsive pulmonary disorder	518, 496, 786,
Osteoporosis	733.03, 733.09, 272
Depression	311, 298, 296.2
Ra_oa	714, 720, 716
Strketa	401, 780, 427

achieved with 3 attributes. It was observed that the first attribute obtained was the investigative tests conducted for the detection of the particular maladies.

For osteoporosis, the first attribute in the attribute election techniques is 733.03 and 733.09, i.e., which indicates the diagnosis test for osteoporosis the calcium density analysis. Similarly for compulsive obsessive order, the first code is 518 which is the test for pulmonary edema; it helps in the detection of COPD (pulmonary disorder).

The code for Alzheimer's is 331 and 294 which is for monitoring the behavioral changes of the beneficiary for early detection of dementia which leads to Alzheimer's in the final stage. The code for rheumatoid arthritis is 714,720 which are used for the detection of the muscle spasm and monitoring the muscle and joint degeneration. The codes for depression are 311, 296.2 which give the different stages and reasons which led to the malady. Finally, the code for stroke is monitoring of blood pressure which is the major reason for this maladies.

These reduced set of codes were tested on a new sample of data extracted. The performance of these codes is shown in Table 3. The range of accuracy observed is

Table 3 Classification accuracy for testing phase

Maladies	Test data accuracy
Alzheimer's	81.39
Compulsive pulmonary disorder	89.61
Osteoporosis	83.67
Depression	78.02
Rheumatoid arthritis	86.39
Stroke	92.45

between 80 and 93%. Thus, these reduced set of codes can be used for early and timely detection of the maladies. These codes were validated for anatomic significance with a medical practitioner.

7 Conclusion and Future Work

The data was restructured, and data mining techniques were applied to derive meaningful knowledge from it. In the first experiment, a study of the number of tests done by a beneficiary was done; it was understood that there is need to reduce the number of tests. To achieve this goal, the second experiment was done in which attribute selection technique GST was used and the reduced set of attributes were obtained. These attributes were then validated on a new data sample and were found to be significant. A study of the reduced attributes that is the investigative tests was made. It was understood that these were relevant and important for the chronic maladies diagnosis. Thus, these are the important codes for particular chronic maladies, and these codes help in taking precautionary measures for the beneficiary so that these diagnosis tests are not required in future. It is of greater benefit for the claims agent to introduce more successful claims in the market. In future, more number of codes can be extracted and these experiments can be performed to get more acquaintance about the diagnosis tests for chronic maladies. Also occurrence of multiple chronic maladies for a beneficiary can be studied. And study based on the demographic information for each enduring malady can be carried out.

References

1. Jensen, P.B., Jensen, L.J., Brunak, S.: Mining electronic health records: towards better research applications and clinical care. Nat. Rev. Gen. **13**, 395–405 (2012)
2. Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J.F., Hua, L.: Data Mining in Medicare and Biomedicine: A Survey of the Literature. Springer Science + Business Media, LLC (2011)
3. GurbuzE, B.M., Ondokuz, Diagnosis of osteoporosis using adaptive SVM and attribute selection. In: IEEE 19th Conference on Signal Processing and Communications Applications (SIU) (2011)

4. Lakshmi, K.R., Nagesh, Y., Krishna, M.V.: Performance comparison of three data mining techniques for predicting kidney dialysis survivalility. IJAET **7**(1), 242–254 (2014)
5. Dimple, classification of data for heart maladies likelihood system using DMP. Int. J. Multidisc. Acad. Res. (SSIJMAR) **2**(5) ISSN 2278–5973 (2013)
6. Ratnakar, S., Rajeshwari, K. Jacob, R.: Likelihood of heart maladies using genetic algorithm for selection of optimal reduced set of attributes. Int. J. Advanc. Comput. Eng. Netw. ISSN (p) **1**(2), 2320–2106 (2013)
7. Bhatla, N., Jyoti, K.: A novel approach for heart maladies diagnosis using data mining and fuzzy logic. Int. J. Comput. Appl. (0975–8887), **54**(17), 16–21 (2012)
8. Hu, R.: Medical data mining based on association rules. Comput. Informat. Sci. **3**(4) (2010)
9. Chan, C.-L., Chen, C.-W., Liu, B.-J.: Discovery of association rules in metabolic syndrome related maladies. In: International Joint Conference on Neural Networks (IJCNN) (2008)
10. Krishnaiah, V., Narsimha, G., Chandra, N.S.: Diagnosis of lung cancer likelihood system using data mining classification techniques. Int. J. Com. Sci. Inf. Technol. **4**(1) (2013)
11. Priya, S., Rajalaxmi, R.R.: An improved data mining model to predict the occurrence of type-2 osteoporosis using neural network. In: International Conference on Recent Trends in Computational Methods, Communication and Controls (2012)
12. Patel, S.B., Yadav, P.K., Shukla, D.P.: Predict the diagnosis of heart maladies patients using classification mining techniques. IOSR J. Agri. Veterinary Sci. (IOSR-JAVS) e-ISSN 2319-2380, p-ISSN 2319-2372. **4**(2) (2013)
13. Dominic, V., Gupta, D., Khare, S.: An effective performance analysis of data mining techniques for cardiovascular maladies. Appl. Med. Informat. (2015)
14. Lahiri, B., Agarwal, N.: Predicting Medicare Expenditure Increase for an Individual from Medicare Data. ACM (2014)
15. Popescu, M., Keller, J.M.: Summarization of Patient Groups Using the Fuzzy C-Means and Ontology Similarity Measures. In: IEEE International Conference on Fuzzy Systems (2006)
16. Hickey, S.J.: Naive Bayes classification of public health data with greedy attribute selection. Commun. IIMA (2013)
17. Smith, G.L., Shih, Y.C.T., Giordano, S.H., Smith, B.D., Buchholz, T.A.: A method to predict breast cancer stage using medicare claims. Epidemiol. Perspect. Innovations (2010)
18. Lita, L.V., Yu, S., Niculescu, S., Bi, J.: Large Scale Investigative Code Classification for Medical Patient Records (2006)
19. Rao, R.B.: Improved cardiac care via automated mining of medical patient records. In: IEEE International Conference on Data Mining: Data Mining Case Studies Workshop (2005)
20. Perotte, A., Hripcsak, G.: Temporal properties of diagnosis code time series in aggregate. IEEE Biomed. Health Informat. **17**(2) (2013)

Review of Machine Learning Techniques for Optimizing Energy of Home Appliances

Jasmeet Kaur and Anju Bala

Abstract The energy consumption of resources in a household has been a periodical issue in the recent study. Most of the energy used in homes is invisible to the users which lead to increase in energy consumption. Energy consumption of the appliances has been estimated by using various machine learning techniques. As per the literature survey, several machine learning techniques have been utilized for optimizing energy consumption. Therefore, in this paper, we present a review related to various machine learning techniques for energy efficiency and tabulated their various parameters and approaches used.

Keywords Cloud computing · Energy efficiency · Electrical home appliances

1 Introduction

Energy consumption of household appliances represents a large part of the total energy consumed. Many approaches have been used for energy efficiency of the household appliances. Using machine learning technique for energy efficiency is one of the methods adopted by many researchers. Machine learning techniques are useful for predicting the energy based on different parameters.

Cloud computing has changed the paradigm which has led to more innovation without worrying about the hardware availability or its maintenance. The concept of cloud computing has spread widely and has gained an importance with upcoming technologies. With its many unique advantages, it can also be utilized in energy scheduling techniques for household appliances.

With the continual growth of the household appliances, there has been a significant increase in power in the system; there are two types of appliances used in homes,

J. Kaur (✉) · A. Bala
Thapar University, Patiala 147004, PB, India
e-mail: jkaur3_me16@thapar.edu

A. Bala
e-mail: anjubala@thapar.edu

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_26

i.e., active and non-active appliances. Active appliances are those that consume continuous power, while non-active appliances are those who have on/off states.

Energy is an important element for economic growth and improving the quality of life. Energy efficiency is the process of minimizing the amount of energy required to provide goods and services. For example, building a home requires using less or more heating and cooling devices according to weather to maintain energy efficiency of a home. Till now, there are very few surveys which describe various machine learning techniques for optimizing energy efficiency. As machine learning techniques can be used for proactive predicting energy, so there is a need to explore various machine learning techniques for optimizing energy consumption. Hence, in this paper, various machine learning techniques have been surveyed out which can be utilized for reducing energy consumption for household appliances.

In Sect. 2, we discussed the previous work done and various challenges. Section 3 describes the metrics on machine learning technique with home appliances. We conclude the paper with final conclusion and future work.

1.1 Motivation

Energy Efficiency of home appliances has received a large attention and share of interest. Electrical home appliances are responsible for the increase in the bill of the homes. There is a need to use the energy in homes in an optimized way so that the bills in our homes should reduce. Many techniques and approaches have been proposed till date for optimizing the energy of household appliances. Salem et al. [7] have reviewed the residential energy on machine learning and data mining techniques. Since now there are no such papers has been there which have surveyed about the different home appliances on machine learning techniques.

2 Related Work

Researchers have long been interested in optimizing the energy efficiency of home appliances. Several approaches have been proposed till now. Some of the approaches are intrusive load monitoring (ILM), non-intrusive load monitoring, supervised approach, unsupervised approach, home energy management system, and demand-side management. These approaches have been used along with various machine learning techniques.

Table 1 describes the previous work done on machine learning techniques. It describes the work of authors with their approach and parameters used for their work. Aiad et al. [2] proposed a cluster splitting approach on cluster cohesion.

Yao et al. [10] proposed the mixed integer linear programming algorithm and the method that took load scheduling problem and energy dispatch problem together. Tsai et al. [8] used a genetic algorithm for optimizing the energy efficiency of the appliances. Flora et al. [5] initiated adaptive k-means algorithm by considering the load shapes for measuring the power consumption of appliances based on hourly basis. Wang et al. [1] propound a demand-side energy management for grid-connected households with a locally generated photovoltaic energy.

Kleiminger et al. [4] investigated the occupancy detection by undertaking the ground truth occupancy data. Wang et al. [9] proposed the mean shift clustering and multidimensional linear discriminate method.

Han et al. [3] developed a home energy management system by comparing the energy of the same household appliances. Murata et al. [6] used the non-intrusive monitoring system and the regression methods for calculating the accuracy of the appliances. However, previous studies have not taken all the household electrical appliances together.

From Table 1, it can be concluded that most of the authors have predicted energy using quality of service (QOS) parameters such as accuracy, power, threshold probability, voltage, and so on. They have also used various machine learning techniques such as cluster splitting, genetic algorithm, load clustering for optimizing the energy for home appliances, but none of them have taken all the home appliances into consideration. So, there is a need to implement techniques in which all the home appliances have been used for optimizing the energy.

2.1 Challenges

To attain energy efficiency using machine learning for optimizing home appliances is not an easy endeavor. Various challenges have been summarized below from Table 1:

- There is a entail for predicting energy for home appliances.
- All the home appliances have not been taken into consideration by any of the technique, so there is a need to implement all the appliances by using efficient technique.
- Most of the techniques are ignoring varying loads appliances, so there is also a requirement to develop a technique which should consider varying load appliances.
- There is a desideratum to develop the scheduling techniques for all the home appliances.

3 Metrics Required for Optimizing the Energy Efficiency

Many techniques have been used for various home electrical appliances by using different parameters and approach. In Table 2, several techniques along with home appliances are shown. Some of the home appliances such as printer, straightener, smoke alarms, CCTV's have not been implemented by any of the machine learning

Table 1 Existing machine learning techniques for optimizing energy of home appliances

Author	Technique used	Parameters	Application	Approach	Challenges
Aiad et al. [2]	Cluster splitting	Accuracy, PValues	Non-intrusive load monitoring	Non-intrusive	To estimate the power consumption of appliances that switch their state frequently. To measure the power of continuously varying loads appliances.
Yoa et al. [10]	Mixed integer linear programming	Power, cost	Optimal energy management	HEMS	Scheduling of all household appliances in all the environmental conditions with different load characteristics. Integrate the energy saving method with carbon reduction framework.
Tsai et al. [8]	Genetic algorithm	Energy rate, power, crossover rate, mutation rate	Scheduling	Unsupervised approach	Calculate the energy consumption for all household appliances. Diversify the energy saving resources.
Flora et al. [5]	Adaptive K-mean algorithm Hierarchical clustering	Variability	Segmentation	Structured	Integrating the energy saving

(continued)

Table 1 (continued)

Author	Technique used	Parameters	Application	Approach	Challenges
Wang et al. [1]	Load clustering, load scheduling optimization	Hourly energy, power, threshold probability	Autonomous appliance scheduling, prosumer concept	DSEM	Develop an energy efficiency method with a carbon reduction framework.
Kleiminger et al. [4]	Classification algorithm	Mean, SD, S.A.D Matthews correlation coefficient, accuracy	Occupancy detection	Unsupervised approach	To detect the home appliances during sleep hours. Investigation of sensor fusion methods.
Wang et al. [9]	Mean Shift Clustering and MLDM	Active power, current, voltage	Non-intrusive appliances	Non-intrusive/Unsupervised	To find an innovative and more of identification of appliances.
Han et al. [3]	Green Home energy management system	Hourly, daily, Weekly, Monthly energy usage	Energy comparison of appliances	HEMS	To diversify energy efficiency of home appliances. Increasing the magnitude of saving the energy instead of replacing it.
Murata et al. [6]	Regression Methods	MAE, RMSE	Estimating the states of household appliances with inverter	Non-intrusive	To estimate the power consumption of all electrical appliances.

Table 2 Metrics considered by using existing machine learning techniques for energy efficiency

Techniques used	Classification algorithm	Regression methods	Cluster splitting	Mean Shift clustering	MLDM	Adaptive K-mean algorithm	Hierarchical clustering	Load clustering	Load scheduling optimization	Mixed integer linear programming	Genetic algorithm
Refrigerator	✓	✓	✓	✓	✓	×	×	✓	✓	×	×
TV	✓	✓	✓	✓	✓	✓	✓	✓	✓	×	×
Washing machine	✓	×	×	✓	✓	×	×	✓	✓	×	✓
Air conditioner	×	✓	✓	✓	✓	✓	✓	✓	✓	✓	×
Dishwasher	×	×	✓	×	×	×	×	×	✓	✓	✓
Chimney	×	×	✓	×	×	×	×	×	×	×	×
Lights	×	✓	✓	×	×	✓	✓	✓	✓	×	×
Fan	×	✓	✓	✓	✓	×	×	✓	✓	×	×
Water filters	×	×	✓	×	×	×	×	×	×	×	×
Grinder	×	×	✓	×	×	×	×	×	×	×	×
Mixer	×	×	✓	✓	✓	×	×	×	×	×	×
Toaster	×	×	✓	×	×	×	×	×	×	×	×
Kettle	✓	✓	✓	×	×	×	×	×	×	×	×
Iron	×	×	✓	×	×	×	×	×	×	×	×
Routers	✓	×	×	×	×	×	×	×	×	×	×
Printer	×	×	×	×	×	×	×	×	×	×	×
Geyser	×	×	✓	×	×	×	×	×	×	×	×

(continued)

Table 2 (continued)

Techniques used	Classification algorithm	Regression methods	Cluster splitting	Mean Shift clustering	MLDM	Adaptive K-mean algorithm	Hierarchical clustering	Load clustering	Load scheduling optimization	Mixed integer linear programming	Genetic algorithm
Heaters	×	×	✓	✓	✓	✓	✓	✓	✓	×	×
CCTV's	×	×	×	×	×	×	×	×	×	×	×
Hair dryers	×	×	×	✓	✓	×	×	×	×	✓	×
Bulbs	×	×	✓	×	×	×	×	×	×	×	×
Cooler	×	×	×	×	×	×	×	×	×	×	✓
Microwave oven	✓	×	✓	✓	✓	×	×	✓	✓	×	×
Straightener	×	×	×	×	×	×	×	×	×	×	×
Smoke alarms	×	×	×	×	×	×	×	×	×	×	×
Tumble dryer	✓	×	✓	×	×	×	×	✓	✓	×	×
Induction cooker	×	×	×	✓	✓	×	×	×	×	×	×
Vacuum cleaner	×	×	×	×	×	×	×	×	×	×	✓

techniques as per author's review. There are some appliances such as refrigerator, TV, air conditioner have been implemented by many techniques such as load clustering, multidimensional linear discriminate method(MLDM), cluster splitting. So, there is a dire need for predicting energy by considering all of the possible home appliances.

4 Conclusion

Energy efficiency is the foundation of increasing prosperity and economic growth since it saves natural and non-renewable resources. Some recent survey related to energy efficiency of household appliances indicated that there have not been any techniques implemented which have taken into account all the household appliances by using machine learning technique. In this paper, we have surveyed the various machine learning techniques on different parameters such as cost, accuracy, threshold probability, power, and so on for optimizing the home appliances for energy efficiency and tabulated their different approaches along with their challenges. Parameters used to depend on which machine learning technique is used. Existing techniques do not consider all the home appliances under one roof. Finally, it has been concluded that none of the authors have implemented machine learning techniques for predicting the energy of all the home appliances. Therefore, there is a need to implement a technique that should consider all the home appliances based on any state.

In the future, we can propose such a technique which will consider all the home appliances and help in optimizing the energy along with the reduction in the bills and reducing the carbon emission.

5 Abbreviations Used

See Table 3.

Table 3 Abbreviations

Abbreviation	Explanation
HEMS	Home energy management system
S.A.D	Sum of absolute error
SD	Standard deviation
DSEM	Demand-side energy management
MLDM	Multidimensional linear discriminate method
RMSE	Root mean squared error
MAE	Mean absolute error

References

1. Adika, C.O., Wang, L.: Autonomous appliance scheduling for household energy management. In: IEEE Transactions on Smart Grid, vol. 5, no. 2, pp. 673–682 (2014). https://doi.org/10.1109/TSG.2013.2271427
2. Aiad, M., Lee, P.H.: Non-intrusive monitoring of overlapping home appliances using smart meter measurements. In: Power and Energy Conference, pp. 1–5. IEEE, Champaign, IL (2017). https://doi.org/10.1109/PECI.2017.7935717
3. Han, J., Choi, C.S., Park, W.K., Lee, I.: Green home energy management system through comparison of energy usage between the same kinds of home appliances. In: 15th International Symposium on Consumer Electronics, pp. 1–4. IEEE, Singapore (2011). https://doi.org/10.1109/ISCE.2011.5973168
4. Kleiminger, W., Beckel, C., Staake, T., Santini, S.: Occupancy detection from electricity consumption data. In: 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, pp. 1–8. ACM, New York (2013). https://doi.org/10.1145/2528282.2528295
5. Kwac, J., Flora, J., Rajagopal, R.: Household energy consumption segmentation using hourly data. In: IEEE Transactions on Smart Grid, vol. 5, no. 1, pp. 420–430. (2014). https://doi.org/10.1109/TSG.2013.2278477
6. Murata, H., Onoda, T. : Estimation of power consumption for household electric appliances. In: 9th International Conference on Neural Information Processing, vol. 5, pp. 2299–2303. IEEE, Singapore (2002). https://doi.org/10.1109/ICONIP.2002.1201903
7. Salem, H., Sayed-Mouchaweh, M., Hassine, A.B.: A review on machine learning and data mining techniques for residential energy smart management. In: 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1073–1076. IEEE, Anaheim, CA (2016). https://doi.org/10.1109/ICMLA.2016.0195
8. Tsai, P., Pan, B.: Intelligent electricity bill saving system for smart home. In: International Conference on Fuzzy Theory and Its Applications (iFUZZY), pp. 105–108. IEEE, Yilan (2015). https://doi.org/10.1109/iFUZZY.2015.7391902
9. Wang, Z., Zheng, G. : Residential appliances identification and monitoring by a nonintrusive method. In: IEEE transactions on Smart Grid, vol. 3, iss. 1, pp. 80–92. (2012). https://doi.org/10.1109/TSG.2011.2163950
10. Yao, L., Shen, J.Y., Lim, W.H.: Real-time energy management optimization for smart household. In: IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), pp. 20–26. IEEE, Chengdu (2016). https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData.2016.31

Performance Analysis of "En-SPDP: Enhanced Secure Pool Delivery Protocol" for Food Delivery Network

Havan Somaiya, Radhakrishna Kamath, Vijay Godhani, Yash Ahuja and Nupur Giri

Abstract Online food ordering systems have made it possible for a restaurant to increase their customer base, popularity, and profit. However, these systems lack in providing security to the user's data and there is always a possibility of data breach. Also, while placing the order for food items from a restaurant, the customer has to pay delivery charges, minimum order policy or minimum amount policy, high service charges which sometimes lead to unhappiness of the customer and result in loss of customer's interest. The En-SPDP protocol proposed in our work not only ensures a secure delivery of messages between food network and delivery network but also presents optimized solution in terms of throughput, efficiency, and minimum blocking period for any customer.

Keywords Queuing theory · Performance evaluation · Food delivery system

1 Introduction

Online food delivery systems are growing on a faster pace in India. The major contenders include Zomato, Swiggy, and FAASOS, and majority of them face challenges in terms of providing security to user data [1], have minimum order

H. Somaiya · R. Kamath (✉) · V. Godhani · Y. Ahuja · N. Giri
Department of Computer Engineering, V.E.S. Institute of Technology, Chembur, Mumbai, India
e-mail: radhakrishna.kamath@ves.ac.in

H. Somaiya
e-mail: havan.somaiya@ves.ac.in

V. Godhani
e-mail: vijay.godhani@ves.ac.in

Y. Ahuja
e-mail: yash.ahuja@ves.ac.in

N. Giri
e-mail: nupur.giri@ves.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_27

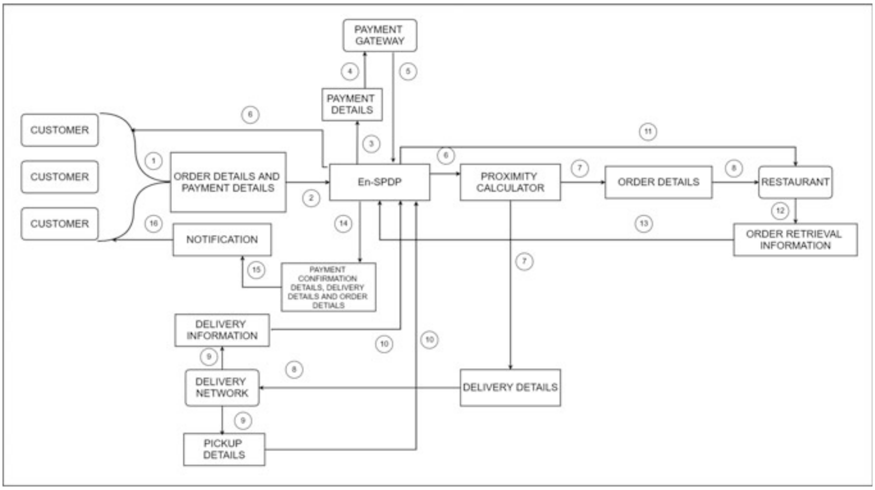


Fig. 1 Overview of En-PDS system

policy or minimum delivery charges, and have ad hoc route decision policy so end up wasting fuel, time, and energy even for small deliveries.

The En-SPDP system overcomes the above-stated issues, by pooling the delivery orders. The work in this paper presents the performance evaluation of En-SPDP for delivery system so that they could plan minimal number of delivery person required for maximizing the customer base without any queuing delay, maximum number of orders which can be serviced with the set number of delivery resources [2]. Thus, the paper presents the benefits of using En-SPDP over generic delivery protocols in terms of efficiency, peak load, throughput, etc., for a delivery network provider.

The En-SPDP also ensures the data security by incorporating an enhanced version of secured electronic transaction (SET) protocol which uses triple signature. The detailed description of the En-SPDP system is presented in the paper "En-SPDP: Enhanced Secure Pool Delivery Protocol" For Food Delivery System [3]. The following section thus explains performance evaluation with respect to above matrices for delivery network (Fig. 1).

2 Overview of Enhanced Pool Delivery System (En-PDS)

The above diagram shows the overview of the system, and it is composed of two parts

- En-SPDP,
- Proximity calculator.

The detailed explanation of each step will be given as follows:

1. When customer makes an order and does payment for that order.
2. The order details as well as payment details get encapsulated in a packet and are sent to En-SPDP system.
3. The system separates the payment details and order details.
4. The system sends the payment details to payment gateway.
5. Payment gateway sends the message of payment failure or payment success to En-SPDP.
6. After successful payment, En-SPDP sends the geo-location of respective customer and hotel (represented as L_{cust} and L_{hotel}) to proximity calculator, which then calculates the X-factor for selecting the delivery person, and if the payment fails, then a notification is sent to the customer stating that payment has failed.
7. Proximity calculator takes the order details from the encapsulated packet. Proximity calculator has current location of all the delivery person (represented as L_{curr}), and the system considers a default value for delay due to traffic (D), preparation delay (T(pd)), and probability for getting a new order along the way. (Poisson probability distribution P(t; μ)). The proximity calculator calculates time required to reach from L_{hotel} to L_{cust} using default speed of delivery person as 30 km/hr and calculating distance between the geo-locations. The time to reach L_{hotel} from current location is calculated for each of the delivery person and using the above default speed. With these default and calculated values, X-factor is calculated and the delivery person with minimum X-factor is selected for that order. X-factor can be calculated using the below formula:
$$X = \mathrm{Max}(T(L_{curr}, L_{hotel}) + D, T(pd)) + T(L_{hotel}, L_{cust}) + D + \int_{t = L_{hotel}}^{L_{curr}} P(t; \mu)dt \text{ [3]}$$
8. Order details are sent to the respective restaurant and delivery details are sent to delivery network.
9. Delivery network will now decide the time for pick up from restaurant and time to deliver order to the customer.
10. Pickup details and delivery information are now sent to En-SPDP system.
11. Pickup details are sent to restaurant by the En-SPDP system.
12. Restaurant makes an order retrieval information packet, when order gets picked up.
13. Order retrieval information is sent to En-SPDP system.
14. En-SPDP system makes a packet of triple signature (which includes payment confirmation details, delivery details and order details).
15. This triple signature packet is converted into a notification.
16. Notification is now sent to customer.

3 System Model and Performance Analysis

The Erlang formula can be used to determine the maximum number of orders that can be immediately serviced by the proposed system without placing the orders in the queue. Since the proposed system has no service-level parameter defined and no queue is maintained when the orders arrive, the Erlang B formula can be used to find out the blocking probability that describes the probability of new order getting blocked/rejected from entering into delivery phase because all the delivery person are busy serving old orders [5].

The Erlang B formula is given as [4]

$$P_b = B(E, m) = \frac{\frac{\mathbf{E^m}}{\mathbf{m!}}}{\sum_{\mathbf{i}=0}^{\mathbf{m}} \frac{\mathbf{E^i}}{\mathbf{i!}}} \tag{1}$$

where

λ Order arrival rate per hour,
h Average delivery time in minutes,
m Number of delivery persons,
E Offered load stated in Erlang and it is given as,

$$E = \lambda h \tag{2}$$

P_b The probability of blocking.

Assumptions and Calculation

$\lambda = 100$
$m = 50$

Existing system:

$h = 30$

$E = \lambda h = 100 \times \left(\frac{30}{60}\right) = 50$ using (2)

$$P_b = B(50, 50) = \frac{\frac{50^{50}}{50!}}{\sum_{i=0}^{50} \frac{50^i}{i!}} = 0.1048 \text{ using } (1)$$

Proposed system:

For a restaurant, preparation of any food item may take maximum of 10 min.

Restaurant preparation time = 10 min

One of the delivery persons' gets the order, as he has the best X-factor value. Now, by the time he picks up the order he was again chosen by the system for another order delivery which is nearby the first order. He gets the second order because he was a good candidate for that order and since he was already carrying an order that area, this order will now result in a pool delivery. So for reaching the first restaurant, which is in the range of 2 km for the delivery person, he takes around 6–7 min at a speed of 30 kmph.

He will wait for 3–4 min, and then he leaves for another restaurant where the food is already prepared. So now the time taken to reach and pick up the deliverables is 20 min. Now, the delivery person leaves to deliver the food items and reaches the first customer in 7 min and then the second consumer in 7 min. Total time taken to deliver both the orders is 14 min, and if we take preparation time, then it is 34 min. Due to traffic jams and signals, he may have wait for 4 min and still he will be completing two orders in 38 min.

In an average single order in 38/2 = 19 min

h = 19

$$E = \lambda h = 100 \times \left(\tfrac{19}{60}\right) = 31.25 \text{ using } (2)$$

$$P_b = B(31.25, 50) = \frac{\frac{31.25^{50}}{50!}}{\sum_{i=0}^{50} \frac{31.25^i}{i!}} = 0.0005 \text{ using } (1)$$

The results reveal that the probability of the new order being rejected for the proposed system is less than the existing system with a margin of 10.42%.

Using the system model and Erlang B formula, the following parameters have been calculated which measures the quantitative and qualitative performance of En-SPDP system.

3.1 Customer Base

Customer base defines the total number of customers whose orders can be serviced immediately by the En-SPDP system without being blocked. It is calculated based on the Erlang B formula (Fig. 2).

From the above graph, it can be concluded that the blocking probability for current system reaches it threshold at 200 orders, while the proposed pooling system reaches its threshold at 320 orders, thus achieving a high level of parallelism as compared to current system.

$$\begin{aligned} \text{Customer base} = {} & \text{Number of Orders handled by Pooling System at threshold} \\ & - \text{Number of Orders handled by Current System at threshold} \end{aligned} \tag{3}$$

$$\text{Customer base} = 320 - 200 = 120 \text{ using } (3)$$

λ	h(current)	h(pooling)	m	P(current)	P(pooling)
100	30	19	50	0.1048	0.0005
120	30	19	50	0.2161	0.0087
140	30	19	50	0.3138	0.0434
160	30	19	50	0.3931	0.1048
180	30	19	50	0.457	0.175
200	30	19	50	0.5093	0.2423
220	30	19	50	0.5526	0.3026
240	30	19	50	0.589	0.3556
260	30	19	50	0.62	0.4018
280	30	19	50	0.6467	0.4423
300	30	19	50	0.6699	0.4775
320	30	19	50	0.6903	0.5093

Fig. 2 Data for calculating blocking probability

Thus, pooling system can handle 120 orders more in comparison to current system leading to increased customer base.

3.2 *Number of Delivery Person*

This factor indicates the number of delivery persons required to deliver λ orders for the system with blocking probability P_b.

Min Delivery Personnel Required:

$\lambda = 100$

$P_b = 0.5$

Existing System:

$E = 50$

$h = 30$

$$0.5 = \frac{\frac{50^m}{m!}}{\sum_{i=0}^{m} \frac{50^i}{i!}} \text{ using (1)}$$

$m = 26$

Proposed System:

$E = 31.25$

$h = 19$

$$0.5 = \frac{\frac{31.25^m}{m!}}{\sum_{i=0}^{m} \frac{31.25^i}{i!}} \text{ using (1)}$$

$m = 17$

$$\begin{aligned}&\text{Reduction in delivery persons (R)}\\&\quad = \text{Minimum number of Delivery person required for Current System}\\&\qquad - \text{Minimum number of Delivery person required for pooling System}\end{aligned} \tag{4}$$

R = 26 − 17 = 9 using (4)

Thus, the En-SPDP system requires nine less delivery persons as compared to existing system to deliver $\lambda = 100$ orders with blocking probability of 0.5 resulting in reduction of human resource.

3.3 Peak Load

Peak load is defined as the maximum number of orders that can be serviced immediately by the system before reaching the blocking probability Pb of the system.

From Fig. 3, it is evident that the maximum no. of orders handled by existing system and proposed system is 200 and 320 orders, respectively, at the blocking probability of 0.5.

Peak load for existing system = 200 orders
Peak load for proposed system = 320 orders

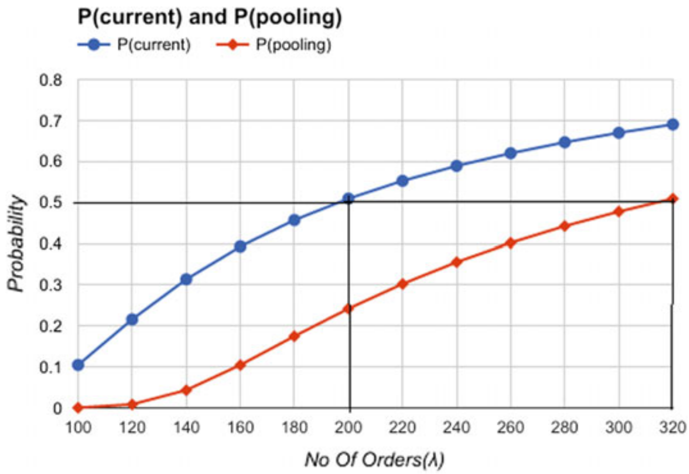


Fig. 3 Graph showing blocking probability for current system and pooling system

3.4 Efficiency

Efficiency is defined as the number of orders that can be delivered by the En-SPDP system as compared to the number of orders that can be delivered by the existing system using m delivery persons. It is calculated as follows:

$$\text{Efficiency}\,(\eta) = \frac{\text{New System Output}}{\text{Existing System Output}} \times 100 \tag{5}$$

$\eta = \frac{105}{85} \times 100 = 123.529\%$ using (5)

Therefore, it can be stated that the proposed system is 23.529% more efficient than existing system.

3.5 Throughput

Throughout is defined as the actual number of orders that can be serviced by the system using m delivery persons. It is given as follows:

$$\text{Throughput} = \frac{\text{Load Carried}}{\text{Load offered}} \times 100 \tag{6}$$

Therefore,

Throughput for existing system $= \frac{85}{120} \times 100 = 70.833\,\%$ using (6)

Throughput for proposed system $= \frac{105}{120} \times 100 = 87.5\,\%$ using (6)

It is observed that the throughput of the proposed system has improved by 16.67% from the existing system.

4 Conclusion

In a given scenario, the proposed system is efficient than the existing system in the following ways. The proposed system can handle 120 customers more the current system. The proposed system is 123% more efficient in handling orders than the current system. The throughput of the proposed system is 17% more than the current system. And all this can be achieved with nine less delivery persons'. Also, the system implements a layer of security which is designed using customized version of SET protocol thus ensuring robustness and data privacy. There is an optimal utilization of the delivery network as apt number of delivery persons' are assigned for delivering the orders. This guarantees increased revenue and minimizes wastage of resources.

References

1. Prachi, P.-P.: Data breaches [the data]. IEEE Spect. **45**(9) (2008)
2. Calvete, H.I., Galé, C., Oliveros, M.J., Sánchez-Valverde, B.: A goal programming approach to vehicle routing problems with soft time windows. European J. Operat. Res. **177**(3), 1720–1733 (2007)
3. Somaiya, H. et. al.: En-SPDP: enhanced secure pool delivery protocol for food delivery system. Informat. Commun. Technol. Intelligen. Syst. (ICTIS) **2**, 554–563 (2017)
4. Erlang, A.K.: The theory of probabilities and telephone conversations. In: Nyt Tidsskrift for Matematik, 1909, vol. 20, no. B, pp. 33–39 (1909)
5. Moshrefjavadi, M.H., Dolatabadi, H.R., Nourbakhsh, M., Poursaeedi, A., Asadollahi, A.: An analysis of factors affecting on online shopping behavior of consumers. Int. J. Market. Stud. **4** (5) (2012)

Survey on Fault-Tolerance-Aware Scheduling in Cloud Computing

Chesta Kathpal and Ritu Garg

Abstract Nowadays, to a large extent, clients look at cloud not just as service provider but also as partner. So, they want cloud to deliver timely and accurate services. Cloud nodes must be reliable in order to provide quality of services as per the customer requirements. Further, physical size of high-performance computing environment is also increasing day by day. Larger the system, more failures are likely to occur that eventually results in the poor reliability of the system which is highly undesirable for the time-critical applications. To deal with the reliability, service provider must know the failure characteristics of the cloud computing nodes in order to better handle the failure using fault-tolerance-aware techniques at the time of scheduling the application tasks. Thus, in this paper, we presented the survey of fault-tolerance-aware techniques which are classified as proactive and reactive fault tolerance. This survey provides the foundation for the researchers to work in the area of fault-tolerance-aware scheduling in order to have better scheduling decisions with the aim to enhance the performance and reliability of application execution.

Keywords Reliability · Fault tolerance · Virtualization

1 Introduction

Cloud is an Internet-based computing paradigm that provides basic services as Infrastructure as a Service (IaaS), Software as a Service (SaaS), Platform as a Service (PaaS) [1]. Different types of cloud providers, i.e., public, private, or hybrids, are responsible for providing above services to user. Nowadays, usage of

C. Kathpal (✉) · R. Garg
Department of Computer Engineering, National Institute of Technology, Kurukshetra, Kurukshetra, Haryana, India
e-mail: Chestakathpal93@gmail.com

R. Garg
e-mail: ritu.59@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_28

cloud computing is increasing because of its enormous features such as sharing of resources, on-demand resource provisioning and virtualization. As users need to pay for required resources on basis of pay as you go model, cloud service provider must provide reliable services as per the QoS [2] requirements of the users. Although cloud computing makes the computing very reliable, dynamic, fast, and easy, it is still facing numerous challenges due to its large-scale and complex architecture. Reliability is the key challenge that needs careful attention.

Reliability and performance of applications in cloud environment depend upon many factors like scheduling of application tasks on resources and occurrence of failure of resources. Aim of scheduling algorithm is to assign the tasks to the available resources with the objective to maximize the execution performance and reliability of an application. To achieve these goals, numerous research works have been carried out in the past on scheduling problems. As we know, scheduling problem is NP-hard in nature, different approaches like heuristics, meta-heuristics, or approximations are the possible solutions. Because of failure of resources, reliability of application execution is decreasing that affects the financial losses to both service providers and users. For performance improvement, it is essential to maximize the reliability of a system. For the same purpose, understanding of foundering and rectifying attributes of cloud resources is very crucial. Garraghan et al. [3] put efforts in knowing the failure characteristics of wide-scale cloud applications. They investigated on the Google Cloud Trace Log for measuring the failure characteristics of cloud servers and tasks, comprising more than 12,500 servers that took 29 days to operate. They observed that the failures' rate of application tasks follow different theoretical distribution like Lognormal or Weibull. Further, analysis on distributed systems [4–6] indicates that the computing resources have Poisson failure distribution.

2 Fault-Tolerance-Aware Techniques in Cloud

To improve the reliability of application execution, failures [7] can be handled either in proactive manner or reactive manner. In case of proactive fault tolerance, it is handled before the occurrence of fault.

2.1 Proactive Fault Tolerance

In cloud computing environment, to minimize the effect of failure in proactive manner, tasks are scheduled to reliable virtual machines according to prior failure information. Various reliability-aware scheduling techniques are proposed in literature for scheduling the tasks in order to maximize the reliability of application execution as shown in the Fig. 1.

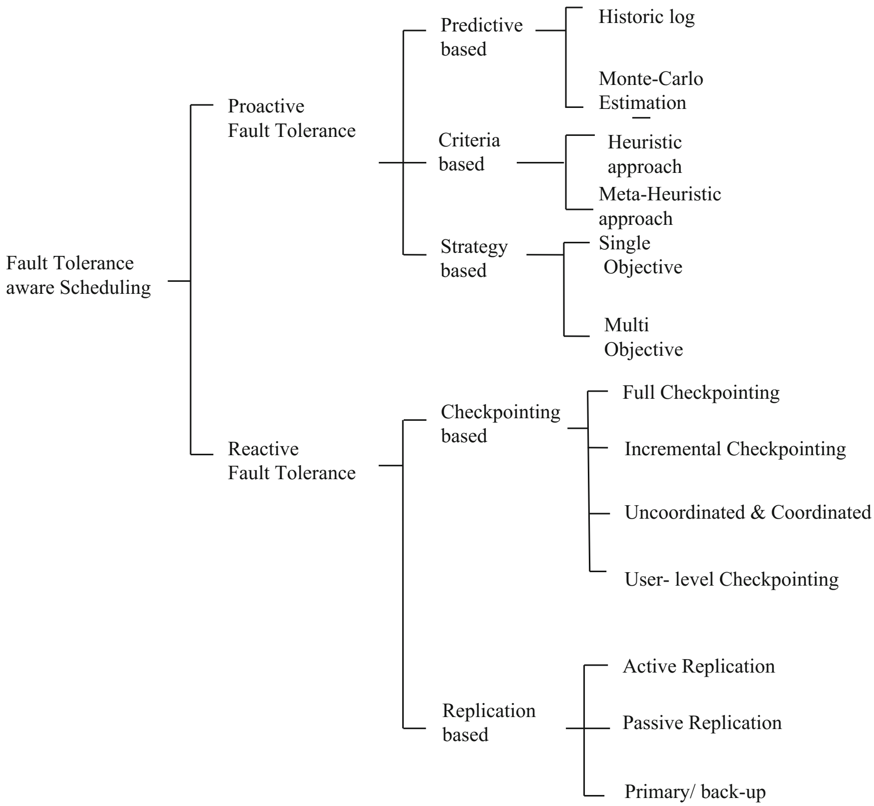


Fig. 1 Taxonomy of fault tolerance aware scheduling techniques

Predictive-Based Reliability-Aware Techniques.

Historic Log Basic. In Ref. [5], authors proposed the scheduling of tasks considering reliability of processors and communication links under the *Poisson failure distribution* [8] of networks. In Ref. [9], proactive approach is proposed for resource reliability considering three parameters: CPU (MIPS), memory (RAM), and bandwidth (BW). The system model finds out the reliability of each virtual machine and then assigns the cloudlets to best reliable VM using prior information like available memory, available MIPS ratio, and available bandwidth ratio. *Weibull failure distribution* is considered in Ref. [10] for maximizing the reliability by scheduling the tasks to reliable VM.

Monte Carlo Estimation Technique. Monte Carlo Failure Estimation algorithm is developed to investigate the future patterns for scheduling the tasks by estimating the failure of virtual machines using Weibull failure distribution in cloud. In Ref. [11], to schedule the tasks to various virtual machines, Failure-Aware Resource Scheduling (FARS) algorithm is proposed in which the reliability is taken into consideration to map the tasks in workflow application. This mechanism reduces the schedule length of FARS because more reliable virtual machines are always selected for execution.

Mapping Strategy-based Reliability-Aware Techniques. As the problem of scheduling in cloud belongs to the class of NP-Hard problems, heuristic and meta-heuristics are the preferred choices. The heuristic-based approaches [10] are problem specific and may not find near-optimal solution. Meta-heuristic approaches are used to handle problem of local optima.

Heuristic Approach. In Ref [12], a heuristic algorithm is developed to achieve one objective, i.e., maximizing reliability under the constraint of another objective, i.e., minimizing end-to-end delay (EED) for distributed computing systems. The reliability is maximized under time bound constraint for mapping the task to processor. Mapping scheme is different from previous centralized mapping scheme where only one central server collects all the information related to networks and decides about mapping, but in this approach all the distributed severs participate in taking the decision of mapping to achieve scalability while keeping the status of failure rates. In Ref. [4], a simulated annealing algorithm is developed for scheduling the tasks to processors to maximize the system reliability and evaluated its performance in comparison with branch-and-bound technique with satisfactory results.

Meta-Heuristic Approach. In Ref. [5], authors developed a scheduling algorithm using meta-heuristic approach along with the maximization of the reliability. NSGA-II approach is used with ENLU technique to avoid applying the non-domination sorting from scratch, and ENLU helps in sorting the solution by taking the advantages of existing knowledge of current population for scheduling the applications on system. Also maintain the energy consumption and reliability of system. The main advantage of the proposed algorithm is that if we scale up then it improves the performance of the system.

Criteria-Based Reliability-Aware Technique.

Single Objective. In Ref. [10], hill-climbing heuristic approach is applied within particle swarm optimization (PSO) in order to maximize the system reliability. This algorithm outperforms the genetic algorithm (GA) because particles can communicate with each other which is the main feature of PSO.

Multiobjective. Various multiobjective algorithms [13, 14] have been developed based on multiple QoS parameters that include reliability, energy consumption, makespan, deadline, scalability. In Ref. [15], the bi-objective genetic algorithm (BOGA) is developed for heterogeneous systems to minimize the energy consumption and maximize the reliability of system by executing the tasks concurrently as a combinatorial optimization problem. In Ref. [16], a double strategy was developed to minimize the Euclidean distance between the generated solutions to a set of user-specified constraints for four-objective problem, i.e., makespan, economic cost, energy consumption, and reliability optimization. In Ref. [17], authors developed a scheduling algorithm using meta-heuristic approach to optimize the three conflicting objectives, i.e., minimization of energy consumption and makespan of tasks along with the maximization of the reliability.

2.2 Reactive Fault Tolerance

Reactive fault tolerance policies handle the effect of failures on application execution when the failure effectively occurs. Reactive fault tolerance policies reduce the effect of failures on application execution when the failure effectively occurs. Various reactive techniques are proposed in the literature for handling the failure of application tasks.

Checkpointing-Based Fault Tolerance Techniques. Checkpointing is a mechanism that records the system state periodically to establish recovery points. At the time of failure, computation restarts from the last saving state. For example, an application needs to access 100 files to complete the execution, and after reading of 99 files, there is some failure and all the computation gets lost. Just to avoid the restarting of application from the scratch, restart it from last saving state. Checkpointing is the most popular technique for recovery of unreliable systems, but some questions also arise: (i) How to optimize the number of checkpoints for each task? (ii) Since the task failure rate is not fixed so how to dynamically tune the optimal solution with the checkpoint/restart mechanism at run-time? (iii) How to know the size of checkpointing interval, etc.? In checkpointing mechanism, storage of snapshot of system state can be done locally or globally. On storing, checkpointing in shared disk achieves higher reliability and degrades the implicit process migration; in contrast, storing process status in local disk increases the cost of process migration. Different kinds of checkpointing techniques have been introduced by researchers like:

Full Checkpointing. It is a mechanism in which complete state of the process is saved to some media after a fixed interval of time. It should be considered that checkpoints should be applied after optimal interval so that it does not cost much overhead that result in maximization of application execution time. In Ref. [18], author developed an algorithm for optimal checkpointing by efficiently selecting checkpointing storage and recovery server.

Incremental Checkpointing. This mechanism helps in reducing the checkpoint overhead by saving those pages in which there have been any changes instead of saving the whole process. In the scheme [19], firstly, save the state of system as full checkpoint, and after that, some mechanism is applied to update the pages which have been modified since the last checkpoint. In Ref. [20], they develop an approach of smart checkpointing based on read-only and read–write part in VM image. During the time of first checkpoint, it saves the state of read-only part only once while rest of checkpoints will save modifications for read–write part. Checkpoints are logged into Hadoop

Distributed File System and replication is made to each node of the service provider. This mechanism is efficient when several tasks are recovering concurrently rather than one task recovery because checkpoints can concurrently recover from different nodes.

Uncoordinated Checkpointing and Coordinated Checkpointing. Each process takes its checkpoint independently; there is no synchronization among the processes to form a consistent global checkpoint. Hence, there is a main drawback called the domino effect. Coordinated checkpointing [21] is an approach for fault tolerance in distributed applications because all the processes come to an agreement to synchronize their checkpoints to make system consistent. This consistent set can be used to limit the rollback.

User-Level Checkpointing. The application programmer identifies program points at which state, status of task must be captured from within the application. In user-level checkpointing, explicit linking is required with the user-level library which is responsible for recovery from failure. In Ref. [22], the algorithm user specifies the different reliability requirements of each task. Each task synchronously checkpointing all its VMs using coordinated checkpointing scheme. One advantage of the algorithms is that it provides equal reliability to multiple users at same time using *peer-to-peer checkpointing*.

Replication-Based Fault Tolerance Techniques. As compared to failure of multiple resources simultaneously, a single resource is much more vulnerable to failure. Replication works as a guard against single point of failure and widely used as a fault-tolerant technique. The primary concern here is to decide the number of replicas. In replication, as the number of replicas increases, the fault coverage increases due to which management of backup systems is very costly. Various types of replication approaches have been studied in cloud computing environment, i.e., job replication, component replication, and data replication.

Active Replication. Each job is replicated on several processors and all the redundant processors are invoked simultaneously and the job will succeed if there is at least one processor which executed the job completely. This technique handles the f arbitrary faults by assigning the tasks to $3f + 1$ replica. In Ref. [23], an algorithm is developed for a Byzantine fault tolerance framework for a reliable system. In BFTCloud, a BFT group is chosen which consists of one primary and $3k$ replications of a task and response to current request will be committed or not committed are judge by cloud module. In case of 'k' faulty nodes in BFT group, execution of tasks will be done again on newly selected primary node and replicas. After identification of k faults, faulty resources will be substituted with other resources. BFTCloud achieves higher throughput where probability of failure of BFT group members is very low. In Ref. [24], MaxRe algorithm is developed for handling the fault during execution. The number of replicas of application tasks depends upon execution time.

Passive Replication. Once there is a crash in primary processor, the task will be scheduled on backup processor for completing their execution. In Ref. [25], algorithm belongs to the category of rescheduling strategy. Compared to MaxRe [24] algorithm, the number of replicas depends upon number of failures but in former depends upon execution time. Problem with the algorithm is that it is not suitable for multiple failures and costly also, because once the failure is detected multiple replicas of tasks are executed. In Ref. [26], algorithm is developed using double strategy. The backup copies are overlapped with the backup copy of their precedence task on the same processors to further reduce schedule length and eventually maximize the performance.

Checkpointing along with replication. In Ref. [27], algorithm dynamically selects the most suitable reactive fault tolerance technique for task execution. The proposed algorithm adaptively determines the length of checkpoint interval and number of replicas of tasks based on previous history. In Ref. [28], an algorithm Replication-based scheduling for Maximizing System Reliability (RMSR) is suggested while considering the task communication. If the number of tasks increases, reliability decreases, but the algorithm RMSR increases the reliability by dynamically replicating the tasks according to threshold value (λ) determined by the user.

Primary-backup replication. In Ref. [29], an algorithm FASTER is developed for real-time applications in virtualized cloud. The author has extended the primary-backup (PB)-based scheduling by incorporating the cloud features like elasticity and virtualization. In Ref. [30], algorithm considers response time and replication cost for dependent tasks and independent tasks in the scheduling process with primary-backup approach. For dependent tasks, backup copies are scheduled with consideration of precedence tasks, and for independent tasks, backup copies can be scheduled with any backup copy of another task on same processor with minimum replication time and cost. In Ref. [6], the given algorithm first selects the number of host servers for placing the replicas of application and then selects the VM by considering the proximity of primary VM and backup VM, and if there is requirement of recovery, then selects the best virtual machine for completion of task as backup VM in the same subnet of hosts. The major aim is to moderate the consumption of network resources by using fat-tree network structure when there is a failure in primary virtual machine.

3 Conclusion

Among all the major challenges in cloud computing, fault tolerance is significant issue. Fault tolerance deals with errors and faults occurred in the system by ignoring, tolerating, and fixing. This paper identified various types of fault-tolerance-aware scheduling techniques. As we know, scheduling problem is NP-hard in nature, so different approaches like heuristics, meta-heuristics, or approximations are the possible solutions that are discussed in the survey. Nowadays, there are several mechanisms for fault tolerance but still there are number of challenges which need to be considered.

References

1. Sadiku, M.N., Musa, S.M., Momoh, O.D.: Cloud computing: opportunities and challenges. IEEE Potent. **33**(1), 34–36 (2014)
2. Patel, P., Ranabahu, A.H., Sheth, A.P.: Service Level Agreement in Cloud Computing (2009)
3. Garraghan, P., Townend, P., Xu, J.: An empirical failure-analysis of a large-scale cloud computing environment. In: 2014 IEEE 15th International Symposium on High-Assurance Systems Engineering (HASE), pp. 113–120. IEEE (2014)
4. Attiya, G., Hamam, Y.: Task allocation for maximizing reliability of distributed systems: a simulated annealing approach. J. Parallel Distrib. Comput. **66**(10), 1259–1266 (2006)
5. Rehani, N., Garg, R.: Meta-heuristic based reliable and green workflow scheduling in cloud computing. Int. J. Syst. Assur. Eng. Manag. 1–10
6. Zhou, A., Wang, S., Cheng, B., Zheng, Z., Yang, F., Chang, R., Buyya, R.: Cloud service reliability enhancement via virtual machine placement optimization. IEEE Trans. Serv. Comput. (2016)
7. Heddaya, A., Helal, A.: Reliability, Availability, Dependability and Performability: A User-Centered View. Boston University Computer Science Department (1997)
8. Qin, X., Jiang, H.: A dynamic and reliability-driven scheduling algorithm for parallel real-time jobs executing on heterogeneous clusters. J. Parallel Distrib. Comput. **65**(8), 885–900 (2005)
9. Charity, T.J., Hua, G.C.: Resource reliability using fault tolerance in cloud computing. In: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 65–71. IEEE (2016)
10. Zhou, A., Wang, S., Cheng, B., Zheng, Z., Yang, F., Chang, R., Buyya, R.: Cloud service reliability enhancement via virtual machine placement optimization. IEEE Trans. Serv. Comput.
11. Rehani, N., Garg, R.: Reliability-aware workflow scheduling using monte carlo failure estimation in cloud. In: Proceedings of International Conference on Communication and Networks, pp. 139–153. Springer, Singapore (2017)
12. Cao, F., Zhu, M.M.: Distributed workflow mapping algorithm for maximized reliability under end-to-end delay constraint. J. Supercomput. **66**(3), 1462–1488 (2013)
13. Dongarra, J.J., Jeannot, E., Saule, E., Shi, Z.: Bi-objective scheduling algorithms for optimizing makespan and reliability on heterogeneous systems. In: Proceedings of the nineteenth annual ACM symposium on Parallel algorithms and architectures, pp. 280–288. ACM (2007)
14. Wang, X., Yeo, C.S., Buyya, R., Su, J.: Optimizing the makespan and reliability for workflow applications with reputation and a look-ahead genetic algorithm. Fut. Generat. Comput. Syst. **27**(8), 1124–1134 (2011)
15. Zhang, L., Li, K., Li, C., Li, K.: Bi-objective workflow scheduling of the energy consumption and reliability in heterogeneous computing systems. Inf. Sci. **379**, 241–256 (2017)
16. Fard, H.M., Prodan, R., Barrionuevo, J.J.D., Fahringer, T.: A multi-objective approach for workflow scheduling in heterogeneous environments. In: Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), pp. 300–309. IEEE Computer Society (2012)
17. Zhang, L., Li, K., Xu, Y., Mei, J., Zhang, F., & Li, K.: Maximizing reliability with energy conservation for parallel task scheduling in a heterogeneous cluster. Inf. Sci. **319**, 113–131 (2015)
18. Zhou, A., Sun, Q., Li, J.: Enhancing reliability via checkpointing in cloud computing systems. China Commun. **14**(7), 1–10 (2017)
19. Paun, M., Naksinehaboon, N., Nassar, R., Leangsuksun, C., Scott, S.L., Taerat, N.: Incremental checkpoint schemes for Weibull failure distribution. Int. J. Foundat. Comput. Sci. **21**(03), 329–344 (2010)
20. Goiri, Í., Julia, F., Guitart, J., Torres, J.: Checkpoint-based fault-tolerant infrastructure for virtualized service providers. In: 2010 IEEE Network Operations and Management Symposium (NOMS), pp. 455–462. IEEE (2010)

21. Cao, G., Singhal, M.: On coordinated checkpointing in distributed systems. IEEE Trans. Parallel Distrib. Syst. **9**(12), 1213–1225 (1998)
22. Zhao, J., Xiang, Y., Lan, T., Huang, H.H., Subramanian, S.: Elastic reliability optimization through peer-to-peer checkpointing in cloud computing. IEEE Trans. Parallel Distrib. Syst. **28** (2), 491–502 (2017)
23. Zhang, Y., Zheng, Z., Lyu, M.R.: BFTCloud: a byzantine fault tolerance framework for voluntary-resource cloud computing. In 2011 IEEE International Conference on Cloud Computing (CLOUD), pp. 444–451. IEEE
24. Zhao, L., Ren, Y., Xiang, Y., Sakurai, K.: Fault-tolerant scheduling with dynamic number of replicas in heterogeneous systems. In: 2010 12th IEEE International Conference on High Performance Computing and Communications (HPCC), pp. 434–441. IEEE (2010)
25. Mei, J., Li, K., Zhou, X., Li, K.: Fault-tolerant dynamic rescheduling for heterogeneous computing systems. J. Grid Comput. **13**(4), 507–525 (2015)
26. Chen, C.Y.: Task scheduling for maximizing performance and reliability considering fault recovery in heterogeneous distributed systems. IEEE Trans. Parallel Distrib. Syst. **27**(2), 521–532 (2016)
27. Amoon, M.: Adaptive framework for reliable cloud computing environment. IEEE Access **4**, 9469–9478 (2016)
28. Wang, S., Li, K., Mei, J., Xiao, G., Li, K.: A Reliability-aware task scheduling algorithm based on replication on heterogeneous computing systems. J. Grid Comput. **15**(1), 23–39 (2017)
29. Zhu, X., Wang, J., Guo, H., Zhu, D., Yang, L.T., Liu, L.: Fault-tolerant scheduling for real-time scientific workflows with elastic resource provisioning in virtualized clouds. IEEE Trans. Parallel Distrib. Syst. **27**(12), 3501–3517 (2016)
30. Zheng, Q., Veeravalli, B., Tham, C.K.: On the design of fault-tolerant scheduling strategies using primary-backup approach for computational grids with low replication costs. IEEE Trans. Comput. **58**(3), 380–393 (2009)

Video Redundancy Investigation and Cooperative Cache Model in Wireless Network

E. M. Malathy and N. Radha

Abstract Video hosting service is of great interest to the Internet user. The video-sharing Web sites allow multiple users to upload the video file. The server in wireless network stores the video file. The issue of redundancy rises when the same video file is stored in different codes in the server. This process initiates additional network load condition on the server. Therefore, bottleneck of the server emerges due to redundancy issue. This paper addresses the redundancy through a combination Motion History Images and discrete wavelet transform-based feature extraction investigation model, and we design an efficient cooperative cache support in the wireless network. The first phase carries the redundancy check for the server, and the second phase adds quick access to the video file through cooperative cache support. This will build the wireless network into more robustness to access streaming traffic mobile nodes. The simulation performance proves video redundancy analysis is important in network-based real-time applications through computation time performance measure and delay measure to access video data in the wireless network.

Keywords Video redundancy · Motion history images · Discrete wavelet transform · Cooperative cache · Wireless network

1 Introduction

Video sharing enables mobile user in Internet to upload and view video data files. The video file varies from short clip to full movie, and video hosting is popularly demanding with Internet facility in wireless networks [1]. Video host features have

E. M. Malathy (✉) · N. Radha
Depatment of Information Technology, SSN College of Engineering, Chennai, Tamilnadu, India
e-mail: malathyem@ssn.edu.in

N. Radha
e-mail: radhan@ssn.edu.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_29

significant features on bandwidth usage. The cost of hosting hassle free experiences for video streaming. Wireless network offers Web sites to support hosting of video data files. However, the user uploads video files from multiple mobile nodes through social networks. This creates impact on generating redundancy in wireless network server. Redundant video storage in server consumes heavy bandwidth. Therefore, it affects the broadcasting quality of the network. Author studied extensive work on masking video of YouTube with half-life and redundancy check. However, the study shows that several 1000 videos exist in given URL at a time. The real-time video detection is possible through security algorithm presented in [2]. In [3], the author used the fingerprint technology with little human intervention in indoor wireless network. Literature [4, 5] shows redundancy check in ad hoc network. Study on social network [6] for performance in wireless network. Caching techniques [6, 7] have been studied in wireless network. However, the multiple file removal in wireless server is still an open challenge.

1.1 Problem Statement and Objective

The wireless network resource and time delivery of data seek attention for research. The redundant video in wireless server takes large storage space and memory. There heavy time delay in accessing such video file in the network. Therefore, this paper presents the redundancy removal algorithm and efficient cooperative cache supports to further enhance the data access to mobile user by designing cooperative cache. The next section deals with the proposed system design of cooperative cache model of wireless network.

2 Proposed System Design

This section deals with two parts that are video redundancy and cooperative cache support which is shown in Fig. 1. In video redundancy, database for study, motion detection, DWT features, and modeling is discussed. Cooperative cache deals with where the cache to be made by Cache node identification and Cache replacement policy.

2.1 Motion History-Based Video Redundancy

A video dataset corpus collected from samples of five male (age group of 20–25 years) and five female speakers (age group of 20–25 years) is considered under this proposed study. For video recording purpose, SONY Handycam is utilized. About 40 video samples with available speaker for uttered digit out of 20 with an angle of

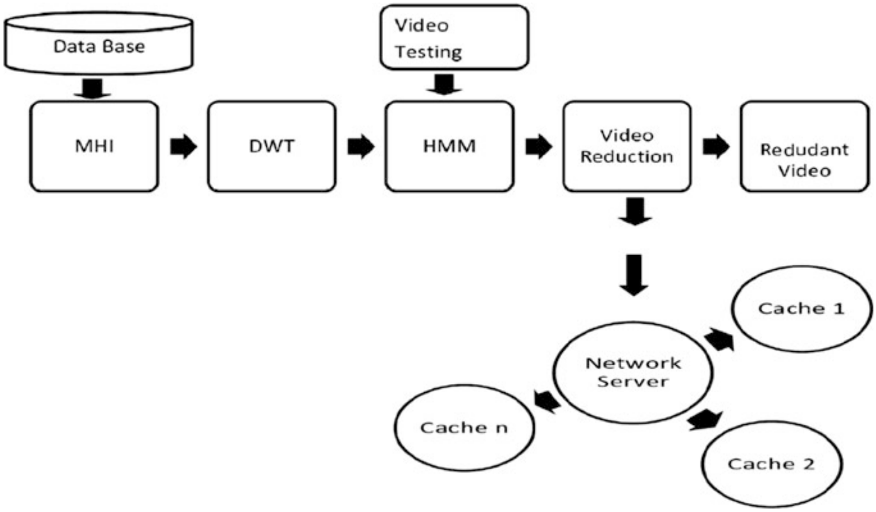


Fig. 1 Proposed system design

45° and the remaining videos with 90° are carried out. Such samples are used for training (30 videos) and testing (10 videos) purposes. Therefore for ten speakers, 300 videos are used for training and 100 videos are used for testing. Figure 2 shows the setup of the utterance of video recordings with male age group of 20 years.

The simulations are carried out under the same lighting and normal environmental conditions for all the video samples. The specification includes 50 f/s with 640-pixel widths and a height of 480-pixel widths for each frame. The horizontal distance between speaker's face position and the camera is about 35 cm and the height includes 65 cm from the ground level to the camera. Effective motion analysis and wavelet transform-based feature extraction are vigorously for video samples, and the performance discussion is carried in the next section.

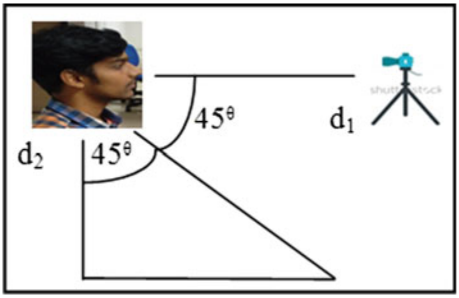


Fig. 2 Setup position of camera

2.1.1 Effective Motion Estimation for Visual Features

Motion features are extracted from videos by using MHI algorithm processed further by image transform approaches such as DWT. The MHI computes DOF of videos, and the DOF of an image $I(x, y)$ for the consecutive frames at time interval t is calculated as follows

$$DOF_{\mathrm{t}}(x,y)/MOF_{\mathrm{t}}(x,y) = |I_{\mathrm{t}}(x,y) - I_{\mathrm{t}-k}(x,y)| \\ \text{where}\, k = 1, N/3,\ N/2\ldots \tag{1}$$

Single-frame DOF of an image contains only limited dynamic information, because the DOF is only between the adjacent frames and binary thresholding is performed in queue. Another type of DOF is MOF, and the motion features are extracted using from multiple (group) frames. But the movement between current and previous frame information has been lost in MOF. This algorithm (DOF/MOF) captures only the information changes in the video region with respect to time and does not indicate the direction of the movements. MHI is the extension of DOF/MOF and captures the spatial location and temporal information of movements in the videos. Thresholding is a step performed on the DOFs/MOFs to produce binary images $B_t(x,y)$ and grayscale images $G_t(x,y)$ which is given by

$$B_t(x,y) = \begin{cases} 1 \ \mathit{if}\ DOF_t(x,y) \geq t \\ 0 \quad \mathit{otherwise} \end{cases} \\ G_t(x,y) = \begin{cases} 0 \quad \mathit{if}\ DOF_t(x,y) \geq t \\ 255 \quad \mathit{otherwise} \end{cases} \tag{2}$$

MHIs are obtained from the union of the multiplication grayscale image $G_t(x, y)$, and weighting factor w are calculated at pixel location (x, y) of the frame which is given by

$$MHI_t(x,y) = \sum_{n=1}^{N} MHI_t(x,y) \cdot w + G_t(x,y) \quad N = 40 \tag{3}$$

Gray values of the MHI are the temporal feature descriptor of the motion. The pixels in the region of movement have higher intensity compared with the pixel where there is no movement [8].

Motion analysis is given in Table 1. The MHI of digit datasets is further processed by image transform approach DWT which is given in the next section.

2.1.2 Image Transform Technique

From the given input MHIs, DWT coefficients are extracted from every MHI. DWT is an iterative process of sub-band decomposition (fourth level) of the visual signal

Table 1 Motion analysis

		
Input Videos	Output of **MHIs - 1**	Output of **MHIs - 2**

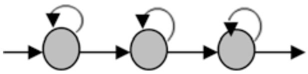


Fig. 3 L-to-R HMM

and converts visual speech signals into approximated coefficients and the detailed coefficients. The approximation coefficients are used in this work because this band contains the significant amount of information about the MHIs as compared to the other DWT coefficients [9]. Therefore, approximation coefficients are alone considered as a feature in this proposed work. The total 256 of DWT coefficients $f(x)$ per MHI's is obtained as:

$$f(x) = MHI(F(x))$$

where $F(x)$ is represented as

$$F(x) = \frac{1}{\sqrt{M}}\left[\sum_{k} w_{\varnothing}(r_o, l)\varnothing_{r,l}(x) + \sum_{r=r_0}^{\infty}\sum_{k} w_{\gamma}(r,l)\gamma_{r,l}(x)\right]$$

where $w_{\varnothing}(r_o, l), w_{\gamma}(r,l)$ are called the approximation or scaling coefficients, detailed or wavelet coefficients, respectively.

A single-stream (L-to-R) Gaussian HMM [10] is used to model the DWT features which are represented in Fig. 3. Three HMM models were created for training. The below-mentioned equation shows the probability vector o for any of I video models, where w_i represents the mixture weights.

$$p(o/\lambda) = \sum_{i=1}^{I} w_i\ b_i(o) \tag{4}$$

Therefore for every MHI's Gaussian Mixture Model with mean, covariance and effective weight parameters are established by $\lambda = \{w_i, \mu_i, \Sigma_i\}$.

Each sample model represents five states, with two states as starting and the end ones. The remaining three states are allotted for feature representation.

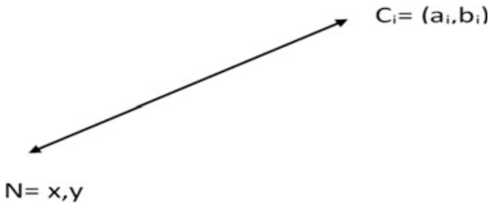
2.1.3 Cooperative Cache Model

Video data access consumes lot of resource in network server. There exists end-to-end delay from server to user. To address this issue, the paper designs a cooperative cache in network server. Once the redundant data is removed, the caching of video data is enabled. The node where the cache is designed is first checked by node cache identification policy. Then, the cache is maintained in the identified wireless nodes. Then, the proposed model has cache replacement policy to have recent file of the video in the cache.

2.1.4 Cache Identification Policy

Cache identification policies help in identifying nodes that will store processed data from feature extraction model. Equation 5 represents the distance computation of the server node to the cache node. The policy where the processed information needs to be stored is effectively identified by distance computation in wireless network.

$$d(N, C_i) = \sqrt{(X - a_i)^2} + \sqrt{(Y - b_i)^2} \tag{5}$$



The problem is designed by dividing the nodes of any wireless network into a number of clusters that minimizes the edges between different clusters. We use Euclidean distance computation to measure the distance value of nodes to server as presented in Eq. 5.

$$Minf(x, y) = \sum_{i=1}^{m} W_i X[|X - a_i| + |Y - b_i|] \tag{6}$$

The minimization function computes the frame rate and cost of the node to cache the data. This way node is identified to cache the video data.

2.1.5 Cache Replacement Policy

The cache data in wireless node has to certify the freshness of the data that is stored. To enable this, we accommodate replacement policy. The TTL value of the data is checked in the cache. Once the TTL value is over, the file is replaced with server and fresh data is again stored.

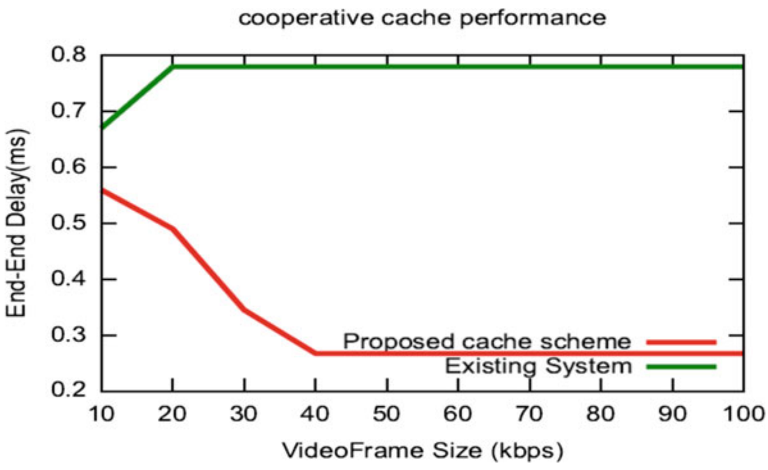
Table 2 Video redundancy performance (%)

	Number of videos	Correctly recognized videos	% Performance (videos)
Angle 90	50	47	94.35
Angle 45	50	8	16.13

2.1.6 Simulation Results

The performance of the video redundancy is analyzed by recognition accuracy. Accuracy is measured by number of video recognized and total number of videos. The performance analysis of video redundancy system with recognition accuracy is presented in Table 2.

In next, the validity of the work proposed is carried by simulation work. The cache support is measured by the delay performances, and the video availability to user without redundancy is given with variation to frame size, and the results are compared [11].



3 Conclusion

An effective method which uses a markov function which helps in identifying whether a video is watched already by a user and this function proves to be more efficient and effective way of identifying which parts of a multimedia video file is already available redundantly in previous videos and it helps in providing a better user experience while using video-sharing sites and social medias. Cooperative cache enables better way for data access and gives good performance in wireless network, thereby removing the bottleneck of the server.

References

1. Aldahoud, A., Singhai, R., Kumar, P., Singhai, J.: Video fingerprinting using AED of CGO based signatures. In: IEEE International Conference in Computer Science and Automation Engineering (CSAE), vol. 1, pp. 292–295 (2012)
2. Mao, J., Xiao, G., Sheng, W., Hu, Y., Qu, Z.: A method for video authenticity based on the fingerprint of scene frame. Neuro Comput. **20**, 22–32 (2016)
3. Yang, Z., Wu, C., Liu, Y.: Locating in fingerprint space: wireless indoor localization with little human intervention. In: Proceedings of the 18th Annual International Conference on Mobile Computing and Networking, pp. 269–280. ACM (2012)
4. Siersdorfer, S., San Pedro, J., Sanderson, M.: Automatic video tagging using content redundancy. In: Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 395–402. ACM (2009)
5. Rezende, C., Pazzi, R., Boukerche, A.: A reactive solution with a redundancy-based error correction mechanism for video dissemination over vehicular ad hoc networks. In: Proceedings of the 15th ACM international conference on Modeling, analysis and simulation of wireless and mobile systems, pp. 343–352. ACM (2012)
6. Katsaros, D., Dimokas, N., Tassiulas, L.: Social network analysis concepts in the design of wireless ad hoc network protocols. IEEE Netw. **24** (2010)
7. Zhao, J., Zhang, P., Cao, G., Das, C.R.: Cooperative caching in wireless p2p networks: Design, implementation, and evaluation. IEEE Trans. Parallel Distrib. Syst. **2**, 29–41 (2009)
8. Radha, N., Shahina, A., Prabha, P., BT, P.S., Khan, N.: An analysis of the effect of combining standard and alternate sensor signals on recognition of syllabic units for multimodal speech recognition. Pattern Recogn. Lett. (2017)
9. Radha, N., Shahina, A., Nayeemulla Khan, A.: A person identification system combining recognition of face and lip-read passwords. In: International conference on Computing and Network Communications (CoCoNet), IEEE (2015)
10. Radha, N., Shahina, A., Nayeemulla Khan, A.: An improved visual speech recognition of isolated words using combined pixel and geometric features. Indian J. Sci. Technol. **9** (2016)
11. Paschos, G., Bastug, E., Land, I., Caire, G., Debbah, M.: Wireless caching: technical misconceptions and business barriers. IEEE Commun. Mag. **54**, 16–22 (2016)

Efficient Energy Sensor Arrays with Uneven Transport Interval in WSN

Chinmya Kumar Nayak and Satyabrata Das

Abstract The trouble of the power gap is an important issue affecting the existence of wireless sensor networks (WSNs). Inside a spherical network with multihop sensors (model as center in common crowns), the best possible communication sequences of all crowns can efficiently recover network existence. Herein article, our concentration is with uneven transmissions, where sensors built in diverse region can be different in their utmost communication speed. So, we advise a power competence method for the highest communication spread, which can seek the best possible communication fields of all crowns to expand the life of the network. In addition, the results of the simulation indicate that the effective power algorithm for the highest transmission range performs superior than the previously existing procedure.

Keywords Power gap · Node allocation · Power stability · Network life span

1 Introduction

In general, node sensors are used with energy source battery and when the energy source battery is gone, the sensor nodes no longer effort. In other words, the key boundary of sensor nodes is the power of your battery exhausted and weak. Because sensor nodes are often used in harsh environments such as battlefields, deep volcanoes, charging, or replacing batteries are often difficult or impossible. The wireless sensor network method must be effective to extend the life of the network.

C. K. Nayak (✉) · S. Das
Department of CSE & IT, Veer Surendra Sai University of Technology (VSSUT), Burla, Sambalpur 768018, Odisha, India
e-mail: cknayak85@gmail.com

S. Das
e-mail: teacher.satya@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_30

Most wireless search network sensors often assume this information gathered by sensors or fast forwarding nodes at a target area in the WSN. It has been reported that the nearest detergent sensors are trying to expel the sensors energy budget second faster [1] that creates an empty space this means that it is sinking. Immediately, after the power opening is displayed, no data is transmitted on stage. Therefore, a significant quantity of power is lost as a result the purpose of network fail at the right time. Experiments [2] have shown that high energy was not used because the life span of the network exceeded the large-scale networks that can reach up to 90%. As a result, network energy efficiency and longevity are a major problem.

As a result, the network's energy efficiency and longevity are a serious problem. However, in some applications the network is heterogeneous. Sensors nodes, in different locations, for example, may differ in the maximum communication range that reaches unmatched primary communications. In this article, we analyze power disruption problems for wireless sensor networks with uneven communication sequences and propose a best possible energy method for sensor networks with an uneven communication array creates resource potential crisis. Results of simulation show that the energy efficiency algorithm at uneven intervals between transmissions in antenna networks increases the energy efficiency and network life of heterogeneous networks.

2 Correlated Work

Recent studies of WSN have shown with the intention of the power of sensor nodes is randomly scattered across the grapevine. Perillo et al. [3] converse two inequalities in power utilization in WSN through flat knot distributions. Let's assume that all nodes are sent directly to zinc noodles, so the nodes over the pools put away a great deal power with the nodes closest with wash basin. Each and every nodes send information to multistage synchronization. In this way, the closest to the washing process is their energy fast, because they need to get a lot more information from the other nodes.

It is observed in the paper, Li and Mohapatra [4] issue of irregular power utilization in a big WSN by means of multiple messages. In several telecommunication networks, each and every one sensor nodes provide information at steady rate. The information is transmitted for synchronization with the same multiplexing. The authors describe the energy disturbance of the tire model (such as the crown model) and determine the traffic load and energy consumption in the nodes (ECR). Of course, the nodes close to the sync button enclose extra passage. In this paper, it is analyzed that the inner nodes of tire sensors devour power quicker in comparison to the external circle; therefore, it is having very less duration operating time. Irregular power utilization is calling an power gap that can have serious consequences. In this paper, various techniques used for solving an energy gap problem, including distribution aids, compression, and merging of traffic.

Xiaobing and Guihai [5] suggest an uneven distribution strategy for nodes to solve the problem of low energy balance. The authors claim that amount of crown nodes to CR-1 crown will upgrade to the crown C1 inside the arithmetical series of the ordinary relation $q > 1$, the crown node CR is NR-1/($q - 1$), it can achieve an energy balance in equilibrium. Here, the underside is within the radius denoted as ri, number of nodes in the crown of Ci is represented as Ni. While this node approach can not work because the nodes in the majority are arbitrary, so the compact level of local nodes is also uncontrollable.

In another paper represented by Stojmenovic et al. converse about length of a mesh and width of every ring of the central crown replica. In this paper, it is viewed that the total amount of energy from the sensor to the sink and the sink must be reduced. Every one crowns definitely has the equal breadth. Though, it is assumed with the intention of all crown-shaped data must be transferred to the crown of Ci-1, and the crown's transfer period is Ci (re-ri-1).

Jarry et al. [6] represented that balance the power load between the ears of the network, mixing direction-finding method that allows every sensor must forward a information its instant neighbors or straight to the bottom node, and the judgment is base on the possible operation based on its residual energy. However, if the radius of the area network exceeds the maximum transmitter field, the planned method not suitable. Shiue et al. [7] also propose a method to solve a power failure of mobile sensors to handle power jets. The price of these aids is far above the ground.

3 Analysis of the Problem and Model of the Network

In this section, we will first introduce a structure replica, together with the system replica and power replica. A better crown template intended for antenna clusters by means of unequal transfer rates will be introduced, representing the idea of a ladder in the conventional crown model.

Model of the Network

We presuppose that the wireless WSN has the subsequent features:

(1) Each and every one node remains Uncovered after startup and stays stationary.
(2) Power of the sensor nodes initially set to a particular cost without charging or replace the battery. This means that the power of the sensor nodes cannot be simplified. When power is consumed, the nodes no longer work.
(3) Different crown sensor nodes have different maximum transmission range. The utmost communication area I of the crown is marked with a tx that is significantly smaller than R (so far from the sensor as its closest probe).
(4) Energy management is accessible. The sensors use dissimilar broadcasting influence to spread information in dissimilar areas.

Table 1 Notations use.

Data	Explanations
Ci	Set the circles with radii ri − 1 and ri
ni	Amount of Corona in Ci
t xi	I-th Crown's maximum coverage
ri	Circle radius I
l	Each sensor node generates and transmits 1 data bit per unit of time

(5) In a convinced speed each sensor transmit out information. Intended for effortlessness assumes that every sensor node generate and transmit 1 data bit per unit.

(6) According to the Network Life Definition [4], we set this document for the network life starting of the network awaiting the first node in the network expires. The net sensor node crown is considered dead when it cannot send whichever data or send its possess data.

Entries in this paper are viewed in Table 1.

4 Enhanced Energy-Efficient Procedures for Non-uniform Utmost Communication Array Plan

Since indicated earlier, for maximizing viability of network, it is important to find the best possible list of broadcasts for all crowns. Without the generality, we are going to talk about it in the six-crown sensor network. Presumptuous the utmost communication array represented as (tx) of a sequence c1, c2, and c3 is 2d. Similarly, c4 and c6 the utmost communication array assumed (tx) is 4d, then each discharge can be transmitted to new crowns by means of dissimilar communication array. A crown allotment table, as shown in Fig. 1. For example, Crown c1 can transmit information to the sink at the node in the communication area d, then its next bound is Crown c which is the sink node. Corona c5 be able to choose communication array d, 2d, 3d, 4d for data transfer to c4, c3, c2, and c1 corona discharge states, in that order. Every crown should choose the next jump in the crown. A representation is given in Fig. 1, under the trees substance are the next jump of every crown. Every one of in the shade items communicate to the communication range list. For example, the communication approach for each coral in Fig. 1 is {c0, c0, c1, c2, c2, c3}. Then, the matching list of communication range is {1, 2, 2, 2, 3, 3}.

Thus, for extending life span of network, we must find best possible list sent by Korona. We offer a power economy method for WSN by means of unparalleled utmost communication range. The purpose of this method introduced as ENMT to

Fig. 1 Relationships of corona

		1d	2d	3d	4d
K1=2	C1	C0			
K2=2	C2	C1	C0		
K3=2	C3	C2	C1		
K4=4	C4	C3	C2	C1	C0
K5=4	C5	C4	C3	C2	C1
K6=4	C6	C5	C4	C3	C2

discover an estimated list of communication list from the internal crown to the external nuclei in stages. As the internal crowns have a shorter life than the outer crowns, ENMT provides better coverage of the internal crowns. The range of external crowns is calculated to provide accommodation the transfer area of the internal crown. Before starting structure, it is possible a list of communication ranges using the ENMT method based on place in order, for example, the total radius of the controlled region, the distribution frequency and the speed at which the corona data is generated. When using sensor nodes, each node in each corona transmits its data according to a fixed list of communication numbers. The ENMT method is explained as given below:

1. The crown binding bench is initialized as Si ($1 \le i \le m$) to unfilled. Add c0 crown to set s0, set i0. Now, if you have a set of communication array lists which network life span is close to the best possible communication range.
2. For every crown, attempt adding the next jump $i = i + 1$ according to the crown relations table to the set Si1 and create a temporary list linked to the Si-1 lists. Of course, if crown has crowns in its next jump, and if Set-1 has a list of transmission ranges p, they set as q X p impermanent list. With these time lists, calculate useful life of the network t and to ti.
3. Presumptuous so as to Tmax which is the utmost life span of ti network. Put in provisional list which network life span stuck between tmax and tmax X (1-timerange) to valid. At this time, the parameter specify the proportion of tmax it find temporary list si. Find the number of selected temporary lists. If it is greater than MAXCOUNT, just add MAXCOUNT provisional trees whose network life is longer than others in Si. Here, MAXCOUNT represent the upper bound of lists in Si.

5 Experiment and Outcome

To calculate the presence of a power competent method for unequal highest communication activity, we simulate a power competent method for a dissimilar method for the highest communication area. Olariu and Stojmenovic [1], different two methods are chosen as base for simulation. We choose first as Max, where every sensing nodes send the highest communication area tx. The sensor nodes which distance to the sink node is less than tx will send information directly to sink node. Another is similar, in which all senor nodes transmit using sensors the same transmission range that is the minimum transmission range among transmission ranges of the coronas.

It is assumed the power of all nodes is initialized to 50 J in simulation. The highest communication range level initialized as 4 and density of the node initialized to 5 knots/m2. Given power competent method, MAXCOUNT of the non-uniform highest communication area initialized to 200. Different factor variables of the power model are set such as β1, β2, β3 and the α factor variable is initialized 4.

Simulate in four different cores (m) 6, 8, 10, and 12 in the corresponding scenarios. Netting crowns are divided into two groups, each of which is Crown m/2. The highest permissible array set to 2 and 4 indoor units to the outdoor clusters. In Fig. 2, we experimented the relationship between life span of network and enduring power ratio using dissimilar methods. We conclude that life span of the network with different three methods decrease according to the growing of the network. Because as the data traffic increase as the radius increase, particularly in the case of internal crowns. In Fig. 2b, it is noticed that increasing the radius of the network life span of the network is declining, as the total initial power is increasing, as a result the residual energy ratio is rising gradually. It is obvious that the power economy method for the non-uniform maximum transmission algorithm works better in comparison to other methods, both in terms of life span of the network and the enduring power ratio.

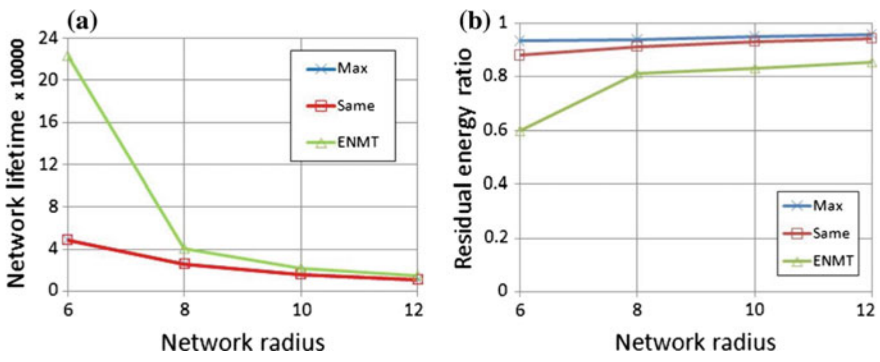


Fig. 2 Outcome of simulation. **a** Life span of network. **b** Enduring power ratio

In contrast to the simulations shown in Fig. 3, the highest communication array of 4 and 2 was changed from the internal cluster to external cluster. In Fig. 4, the relationship between life span of network and enduring power ratio is using dissimilar methods. It is clear that the energy saving algorithm for an unmatched maximum input algorithm is even best in comparison with other method for life span of network, and enduring power ratio.

Simulate in four scenarios corresponding to 6, 9, 12, and 15 SEK (m). Table crowns are divided keen on three clusters, every cluster containing following m/3 crowns. Highest allowable communication array is 2, 3, and 4 indoor units to an external cluster. Figure 3 viewed relationship between n life span of the network and enduring power ratio to a different a procedure. Note that the power efficiency procedure for the unmatched broadcasting algorithm exceeds two other algorithms for both network life and res energy.

In Fig. 4 simulations, the highest ground transferred to the groups 4, 3, and 2 of the internal cluster to the external cluster. In simulation, Fig. 5 viewed relationship between network life and enduring power ratio using different algorithms. It is clear

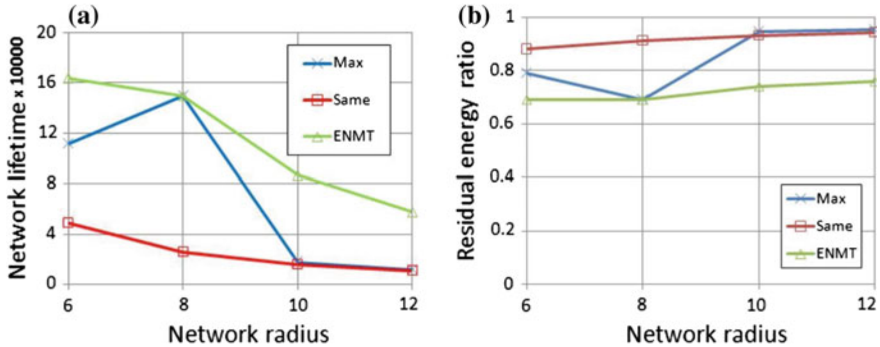


Fig. 3 Result of simulations. **a** Life span of network. **b** Enduring power ratio

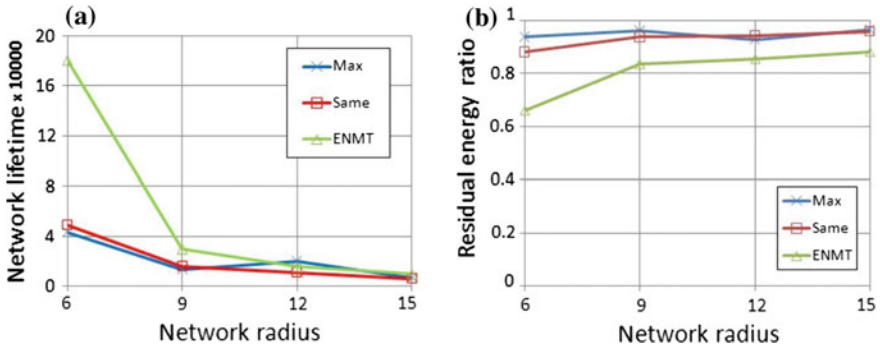


Fig. 4 Results of simulation. **a** Life span of Network. **b** Enduring power ratio

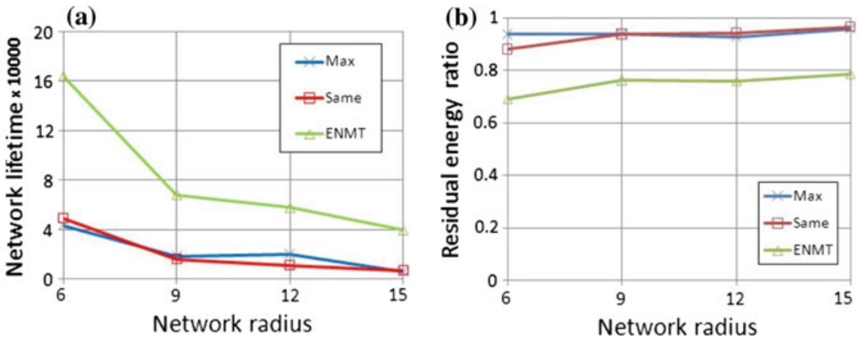


Fig. 5 Result of simulation. **a** Life span of network. **b** Enduring power ratio

with the intention of energy efficiency method of unequal transmission algorithm is also more efficient in comparison to previous method, together life span of the network and enduring power ratio.

6 Conclusions

Power failure problems significantly reduce the length of the sensor body and cause residual energy losses to the wireless sensor networks. In this paper, we suggest energy efficiency method named an efficient energy algorithm that solves the problem of unmatched maximum power level as a problem by electric drilling for sensor frames with unequal communication sequences. The goal of an efficient energy procedure for unequal communication must be found more appropriate transfer records which broaden wireless sensor network life span. The outcome of simulation showed an efficient energy method for unbound maximum power speeds exceeds other algorithms and effectively increases network durability.

References

1. Olariu, S., Stojmenovic, I.: Design guidelines for maximizing lifetime and avoiding energy holes in sensor networks with uniform distribution and uniform reporting. In: Proceedings of IEEE Infocom, Barcelona, Spain
2. Lian, J., Naik, K., Agnew, G.: Data capacity improvement of wireless sensor networks using non-uniform sensor distribution. Int. J. Distrib. Sens. Netw. **2**, 121–145 (2006)
3. Perillo, M., Cheng, Z., Heinzelman, W.: On the problem of unbalanced load distribution in wireless sensor networks. In: Proceedings of IEEE Globecom, Dallas, TX, USA, pp. 74–79 (2004)
4. Li, J., Mohapatra, P.: Analytical modeling and mitigation techniques for the energy hole problems in sensor networks. Pervas. Mob. Comput. **3**, 233–254 (2007)

5. Xiaobing, W., Guihai, C.: Avoiding energy holes in wireless sensor networks with nonuniform node distribution. IEEE Trans. Parall. Distrib. Syst. **19**, 710–720 (2007)
6. Jarry, A., Leone, P., Powell, O., Rolim, J.: An optimal data propagation algorithm for maximizing the lifespan of sensor networks. In: Proceedings of Dcoss, San Francisco, CA, USA, pp. 18–20 (2006)
7. Shiue, H.Y., Yu, G.J., Sheu, J.P.: Energy hole healing protocol for surveillance sensor networks. In: Proceedings of Workshop on WASN 2005, 29 July–2 August 2005, Taoyuan, Taiwan (2005)
8. Wadaa, A., Olariu, S., Wilson, L., Jones, K., Eltoweissy, M.: Training a sensor networks. In: Proceedings of Monet, Orlando, FL, USA (2005)

Stress and Modal Analysis of Six-Axis Articulated Robot Using ANSYS

Supriya Sahu and B. B. Choudhury

Abstract Industrial robots are used in pick and place, and various other operations in industries. So in this paper, the aim is to find out the stresses and modal analysis of different points of a six-axis industrial robot to determine its maximum shear stress, natural frequencies, and mode shapes. The optimum stress and modal analysis are done by finite element analysis (FEA) using the ANSYS workbench. For this analysis, the mesh size is taken as 0.01 mm. Different values of loads are applied on the griper to find out the maximum value of stress. For modal analysis, different cracks on the robot are considered. The modal shapes and natural frequencies for robot with crack and without crack are compared to find the weak part on the robot structure so that any design modifications can be done in order to make the robot more efficient for industrial work.

Keywords ANSYS workbench · FEA · Industrial robot · Mode shapes Stress analysis

1 Introduction

ARISTO is a six-axis articulated robotic arm having six axes such as base, shoulder, elbow, wrist, pitch, and roll. Modal analysis is performed to find the fundamental frequencies (modes) and their associated behavior (mode shapes). This can be done by analyzing the deformation shape of structure from FEA model. High deformation area can be used for sensors placement as it can capture accurately the frequency response. Stress analysis gives the idea about the weak part of the robot

S. Sahu · B. B. Choudhury (✉)
Department of Mechanical Engineering, Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India
e-mail: bbcigit@gmail.com

S. Sahu
e-mail: supriyaigit24@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_31

structure so that any design modifications if possible can be applied to get a better robotic structure.

Bhusnar and Sarawade [1] analyzed the dynamic behavior of structures like a rectangular plate, bolted lap joint and studied the natural frequencies and mode shapes using finite element analysis (FEA) and analytical methods. Chitte et al. [2] minimized the structural deformation of a six-axis robot manipulator in all the three x, y, and z directions by using calculus method on applied forces by means of which the stiffness can be improved. Ghiorghe [3] considered the criteria for minimizing the material used for building the robot structure. Finite element method has been used to determine the optimum value of design parameters. Jevan and Rao [4] analyzed the structural parameters of industrial robot, and from the analysis it has been found that the square-shaped robot arm vibrates less as compare to circular-shaped sustains. Kumar et al. [5] used FEA to analyze the natural frequencies, vibrations, and mode shapes in order to get the fracture locations in the bone by simulation. Kumar and Sambaiah [6] used finite element tool in the industrial robot for keeping the tasks in right location, with right force and torque in right time. FEA tool is also used by Nor et al. [7] to find out high stress value to determine maximum deflection in low loader structure with modeling and simulation. Pachaiyappan et al. [8] used ANSYS, the commercially available analysis tool is to analyze various critical loads acting on the articulated robot. Ristea [9] designed the control system of a motorized robot structure, along with differentiating the material such as composite and aluminum. Sridhar et al. [10] studied the behavior of industrial robot structure and used finite element method to develop the model of robot for correctness in different loading conditions.

2 The CATIA Model and Analysis

2.1 Modeling of Robot in CATIA V5

The modeling of the robot is prepared using the software CATIA V5. First of all, models of different parts of robot structure are prepared and then assembled to get the total structure. The model is then imported from CATIA into ANSYS workbench software. As the robot is of six-degree freedom structure and has to be fixed on the floor, when in use, the base surface is used as a fixed constraint.

2.2 The Finite Element Stress Analysis by ANSYS

The stress analysis is performed to test the robotic arm to withstand specific load conditions. In FEM, the model of the structure is evaluated to determine if it can handle the load conditions calculated before in the dynamic model or not. It is important to determine the stiffness of the structure. Physical analysis of the model

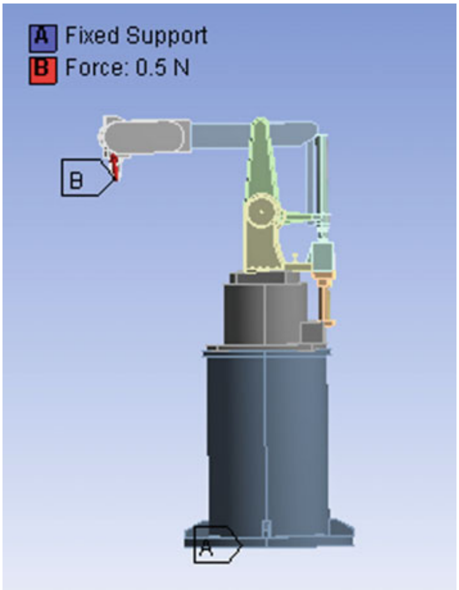


Fig. 1 Static structural setup of robot

is carried out in finite element environment. For this, the base nodes need to be constrained. For the analysis, the mesh size is taken as 0.01 mm and different loads are applied to the gripper end. The material selected is structural steel. Constraints for the assembly parts are included manually. The static structural setup of the robot is shown in Fig. 1.

For analysis of the structure, the loads applied are 0.5, 25, 50, 75, 100, and 125 N. Analysis is done to obtain shear stresses for different loads applied, and the maximum value of shear stress obtained is noted.

2.3 The Finite Element Modal Analysis by ANSYS

Vibration characteristics of the model such as natural frequency and mode shape are determined by means of modal analysis in FEA. The requirement of more capable material leads the modal analysis in systems like industrial robots. The occurrence of crack or damage in any engineering structure, rotating machines, causes premature failure and creates different operational problems. One of the criteria for fault detection is the change in dynamic behavior of structures. Due to the presence of crack, alteration of parameters such as the natural frequencies, mode shapes, and amplitude of vibration takes place.

In this present experiment, an effort has been made to explore the behavior of cracked robot. In order to achieve the three mode shapes of robot with and without crack has been analyzed.

3 Results and Discussion

3.1 Results of Stress Analysis

Optimized structural design for the structures of the industrial robots has to meet criteria regarding dimensional design and shape, material consumption and adapt this to the functional requirements. For an optimized design of the robot structure, all the aspects of industrial applications where the structure will be integrated are considered. The results achieved for maximum shear stress for each load applied are presented in Table 1.

The analysis on deformation and stress of the structure gives the idea about life, damage, and failure of robot. The shear stresses for six different gripper loads are shown in Fig. 2. The bottom part of the structure having the lowest value of shear stress is shown as dark blue color, and the top part of the structure shows the maximum value of shear stress is shown as red color. The maximum value of shear stress obtained is for 125 N which is near the gripper. The dialog box which is on the left side of the figure gives the values. The dark blue indicates the lowest value, the light blue indicates the lower value, the yellow color indicates the higher value, and red color indicates the highest value of shear stresses.

Taking these output values of shear stresses, a graph has been plotted by taking load (N) in x-axis and shear stress (Pa) in y-axis and is shown in Fig. 3.

From the graph, it is seen that as load increases the shear stress increases uniformly. And for the force of 125 N, it is the highest value of $1.0606e^{7}$ Pa is obtained.

3.2 Results of Modal Analysis

The output frequencies of generated mode shapes at three different modes are obtained. The generated first mode shape for the robot without crack is presented in Fig. 4. The dialog box at the left side shows the values of deformations. The minimum value of deformation is shown in dark blue color which is near the base, and the maximum value of deformation is shown in red color which is near the

Table 1 Maximum shear stresses for different loads applied

SI. no.	Gripping loads (N)	Maximum shear stress (Pa)
1	0.5	42,425
2	25	$2.1212e^{6}$
3	50	$4.2425e^{6}$
4	75	$6.3637e^{6}$
5	100	$8.485e^{6}$
6	125	$1.0606e^{7}$

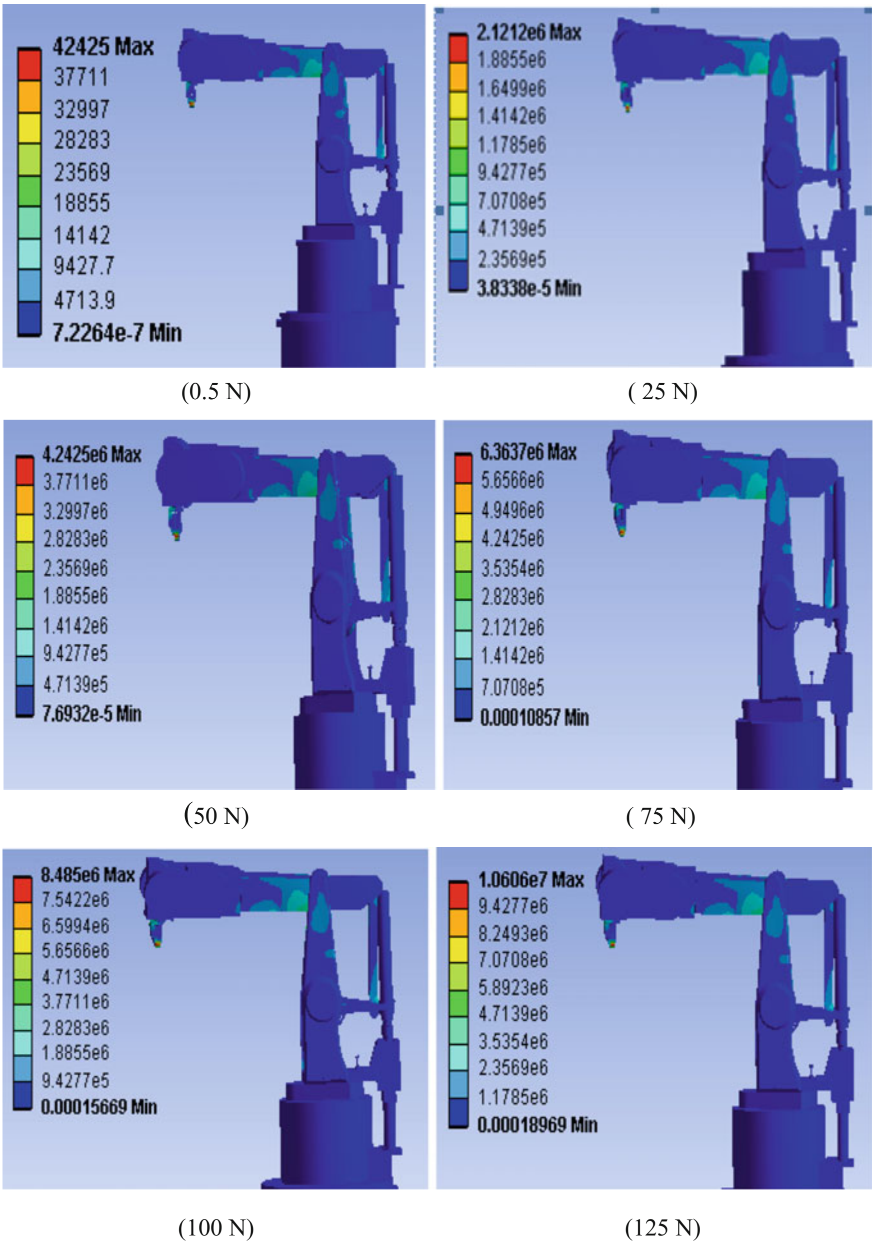


Fig. 2 Shear stresses of robot model structure at various loads

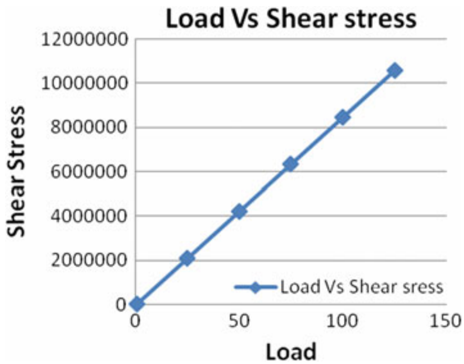


Fig. 3 Graphical presentation of load (N) versus shear stress (Pa)

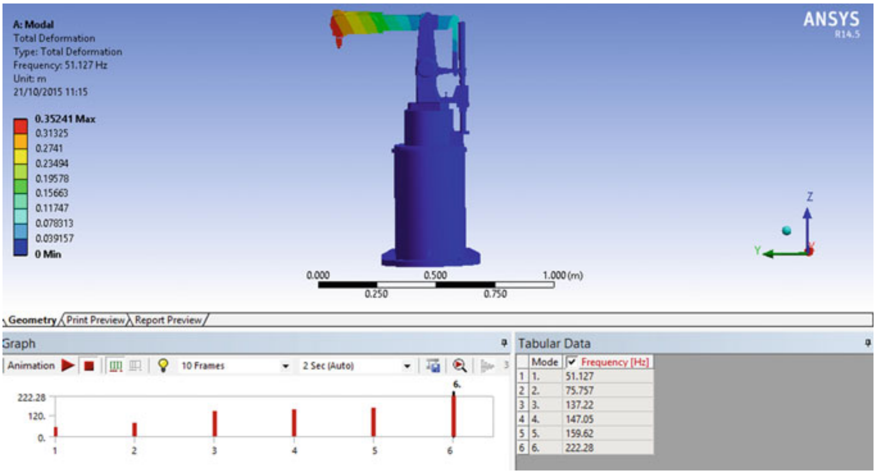


Fig. 4 Generated first mode shape for robot without crack

gripper of the color scale. The first three natural frequencies are obtained which is shown in the form of tabular data.

Out of three mode shapes generated, the first mode shape for robot with crack 1 is shown in Fig. 5 and for robot with crack 2 is shown in Fig. 6. From these figures, it can be observed that there are significant variations in mode shapes. The first three natural frequencies were noted down.

From these, it has been noticed that the crack had a great influence on mode shapes. In dynamic structures, the presence of crack causes vibration and affects the mechanical behavior of the structure. So the identification of cracks in the structures is a relevant issue.

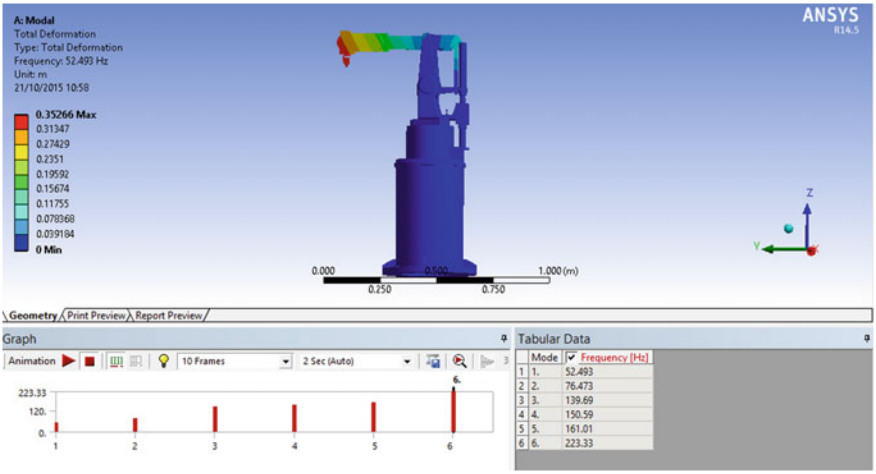


Fig. 5 Generated first mode shape for robot with crack 1

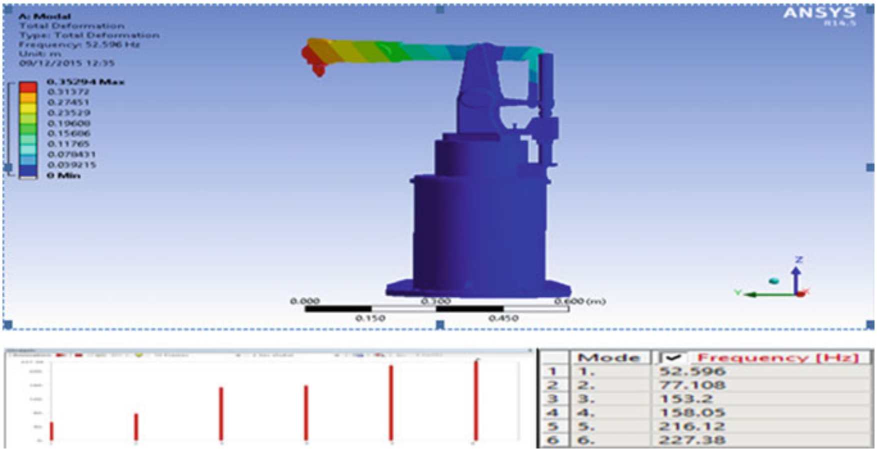


Fig. 6 Generated first mode shape for robot with crack 2

Table 2 Frequencies for different mode shapes for robot without and with crack

Mode	Frequency without crack	Frequency with crack 1	Frequency with crack 2
1	51.123	52.528	52.596
2	137.22	139.7	153.2
3	159.62	161.06	216.12

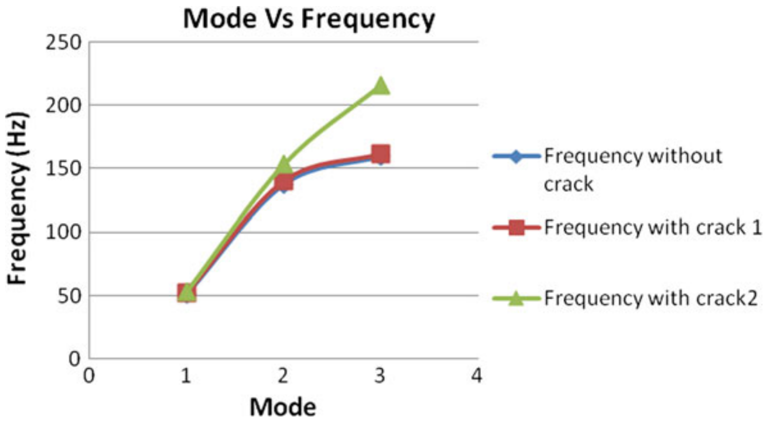


Fig. 7 Graphical presentation of mode versus frequency (Hz)

Output frequencies of ANSYS workbench for the robot with crack and without crack have been given in Table 2, and it is found that due to the presence of crack the frequencies for the three different modes increases. Comparing the frequencies of robot with crack 1 and crack 2, it can be observed from the table that for crack 2, the frequencies for three mode shapes are more than crack 1.

Taking these output values of frequencies, graphs have been plotted by taking mode in x-axis and frequency in y-axis and are shown in Fig. 7. The blue line indicates the frequencies without crack, the red line which is above the blue line indicates the frequencies with crack 1, and the green line indicates the frequencies with crack 2 for three different modes of 1, 2, and 3 as presented in the graph.

From the figure, it can be noticed that with the presence of crack, frequency of vibration increases for first mode, second mode, and also for the third mode of vibration. Due to the presence of crack, it is observed that there are significant variations in mode shapes and frequencies also depend upon the location of cracks.

4 Conclusion

The stress analysis that has been performed shows that the safety factor is reduced. Therefore, in order to maintain the safety the lifting capacity would need to be reduced to 2.5 N force. It is important to analyze the robot's structural integrity after its complete fabrication and assembly to ensure the safety factor is not adversely affected.

Modal analysis using FEA was capable of giving information about all the nodal diameter or modes for the structure. Then, the frequencies so obtained can be compared with the frequencies of sources of excitation. From the analysis, it is observed that whether those frequencies lie within the range of that structure or not, which might lead to deformation or failure of the component in some later point in time.

References

1. Bhusnar, M.S., Sarawade, S.: Modal analysis of rectangular plate with lap joints to find natural frequencies and mode shapes. IOSR J. Mech. Civil Eng. (IOSRJMCE), pp. 06–14 (2016). e-ISSN: 2278-1684, p-ISSN: 2320-334X
2. Chitte, P.G., Bansode, S.S., Rathod, S.S., Motgi, N.S.: Structural and irrational analysis of six axis ARISTO robot using ANSYS. IJIRT **3**(1) (2016). ISSN: 2349-6002
3. Ghiorghe, A.: Optimization Design for the Structure of an RRR type industrial robot. U.P.B. Sci. Bull., Series D **72**(4) (2010)
4. Jeevan Rao, A.N.: Modeling and analysis of robot ARM using ANSYS. Int. J. Sci. Eng. Technol. Res. **04**(33), 6692–6697 (2015). ISSN 2319-8885
5. Kumara, A., Jaiswala H., Garga, T., Patil, P.: Free vibration modes analysis of femur bone fracture using varying boundary conditions based on FEA. Proc. Mat. Sci. 1593–1599 (2014)
6. Manoj, K.K., Sambaiah, C.H.: Modelling and study of motion control system for motorized robot arm using mat lab and analysis of the arm by using ansys. Int. J. Mech. Ind. Eng. (IJMIE) **2**(1) (2012). ISSN No. 2231–6477
7. Nor, M.A.M., Rashid, H., Mahyuddin, W.M.F.W., Azlan, M., Mahmud, J.: Stress analysis of a low loader chassis. Proc. Eng. 995–1001 (2012)
8. Pachaiyappan, S., Balraj, M.M., Sridhar, T.: Design and analysis of an articulated robot arm for various industrial applications. J. Mech. Civil Eng. 42–53 (2014). e-ISSN: 2278-1684, p-ISSN: 2320-3334X
9. Ristea, A.: FEA Analysis for Frequency Behavior of Industrial Robot's Mechanical Elements, pp. 26–28. Sinaia, Romania (2011)
10. Sridhar, C., Doukasa, J., Pandremenosa, P., Stavropoulosa, P., Foteinopoulosa, G.: On an empirical investigation of the structural behavior of robots. Proc. CIRP **3**, 501–506 (2012)

New Text-Based User Authentication Scheme Using CAPTCHA

Atiya Usmani, Amrah Maryam, M. Sarosh Umar and Muneeb Hasan Khan

Abstract In recent times, there is a pressing needs to develop more robust and efficient user authentication techniques in order to protect the confidentiality and integrity of crucial information. Thus, many new techniques are being developed like graphical passwords, two-step verification, and others. In this domain, the conventional text-based passwords are often criticized and considered less secure but surprisingly they are still the most common authentication technique. This popularity is due to their ease of implementation and use. In this paper, we have proposed new text-based user authentication technique using CAPTCHA. This technique overcomes most of the challenges faced by both conventional and contemporary authentication methods and offers a higher level of security than them. The password is user-friendly, quick and easy to compute, and more secure.

Keywords One-time password (OTP) · Authentication · Security Attacks · CAPTCHA

1 Introduction

Nowadays, user authentication and cyber security are a grave concern among computer scientists and engineers. There is an increase in reported cyber crimes such as data theft, cyber extortion, hacking, and various other scams and frauds.

A. Usmani (✉) · A. Maryam · M. Sarosh Umar · M. H. Khan
Department of Computer Engineering, Zakir Hussain College of Engineering
and Technology, Aligarh Muslim University, Aligarh 202002, Uttar Pradesh, India
e-mail: atiya.usmani@zhcet.ac.in

A. Maryam
e-mail: almas36912@zhcet.ac.in

M. Sarosh Umar
e-mail: sarosh.umar@zhcet.ac.in

M. H. Khan
e-mail: muneebhkhan@zhcet.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_32

Most of the contemporary software and Web sites used text-based password. These passwords have their demerits and limitations such as: They are susceptible to security attacks like phishing attacks, dictionary attacks, and brute force attack. The user often has to write down his password which the intruder can easily get hold off. Thus, we can conclude that text-based password is not the ideal user authentication systems, has many limitations of their own, and is not truly secure.

This means that there is an opportunity to develop new techniques [1] and methods for user authentication. These can be entirely different techniques like token-based authentication, biometrics-based authentication, we can also create a slightly modified text-based technique, or it can be a hybrid of any two technologies mentioned above. But there are certain requirements that this newly developed user authentication technique should satisfy [2]. These are listed below:

1. It should overcome the security drawbacks of contemporary password that is it should not be prone to shoulder surfing, dictionary attacks, brute force attack, etc.
2. It should be quick, easy to compute, and user-friendly.
3. Moreover, if it is a one-time password and fairly random, then it can be classified as a good user authentication technique.

In this paper, we have proposed a new text-based user authentication technique that satisfies all the above requirements. Then, we discuss its implementation and results.

2 Background

2.1 Authentication Methods

The main methods of user authentication are:

1. **Token-based authentication** requires token, for example, an ATM card, id card, RFID card. The user is authenticated based on the token. It is also referred as "what you have" type of authentication [3].
2. **Biometrics-based authentication** is also referred as "what you are" type of technique. Every human being has some unique features, biometric uses these features for authentication like the iris scan, fingerprint, facial recognition. These signs are highly secure but extremely expensive [3].
3. **Knowledge-based authentication** technique is also referred to as "what you know" type of technique. It includes both text-based and picture-based authentication [4, 5].

2.2 Possible Attacks on Text-Based Passwords

1. **Brute Force Attack**: Brute force attack is a trial and error method used by the attacker to decode encrypted data such as password through exhaustive effort [6].
2. **Dictionary Attacks**: In dictionary attacks, the attacker attempts to gain illicit access to a computer system by using a very large set of words, generally words from a dictionary to generate potential passwords.
3. **Spyware Attacks**: Spyware is a software that gathers information about the user without his knowledge and may send it to another entity without the user's consent, for example, key listening spyware or keyloggers.
4. **Shoulder Surfing**: Shoulder surfing refers to using direct observation techniques, such as looking over someone's shoulder to get information. This can also be done by using a video camera and analyzing the video footage later.
5. **Eavesdropping**: In eavesdropping, the attacker intercepts the communication between the user and the server and obtains crucial information of the user.

3 Proposed Scheme

We have proposed a new text-based authentication scheme which generates a new password each time using CAPTCHA. This technique has been devised in order to achieve a higher level of security and overcome the challenges faced by conventional text-based passwords.

3.1 Method

User authentication system provides the following functions.

1. **Registration**: During registration, the user has to enter his email and an eight character password (i.e., his real password). This real password serves as a private key for the user. He will never have to enter it again. The user also has to enter a location digit. This digit will be used later at the time of login. It gives the location of the user's answer based on the CAPTCHA challenge in the user's one-time login password. The rest of the characters in the login password are dummy characters.
2. **Login**: When the user logins every time, he sees a CAPTCHA of 15 characters. It is a random character string but it will have any one character from the user's real password or key. Let's call it C'. Using this key and the location where his password character appears in the CAPTCHA, he has to calculate new character *NC* by (1).

$$NC = Key\,[loc] \tag{1}$$

Here, *NC* is the character in the key at location *Loc*. *Loc* is given by (2).

$$Loc = loc1 + loc2 \tag{2}$$

In (2), *Loc1* is the location of C' in user's real password or key, and *Loc2* is given by (3). If the real password string contains same characters at two different indexes, the first index is taken as *Loc1*.

$$Loc2 = integer\ value\ of\ \left(\left(Loc_{Cp}\right)/10\right) + \left(\left(Loc_{cp}\right) modulus(10)\right) \tag{3}$$

Loc_{Cp} is the location of C' in CAPTCHA, where C' is the first character which is both in the random string CAPTCHA and also in the real password. If the location is a two-digit number, then the two-digit will be added.

This new character *NC* has to be entered at the correct location given by the location digit specified during registration. The other characters will be some dummy characters in the password. The password should altogether be 15 characters long. The user just need to identify the common character C' and note down the location. Then, all he needs to do is count the characters from C' based on the location in the real password.

Example 1:

Real Password: Alics%91

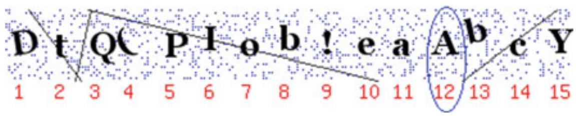
Location Digit: 4

CAPTCHA (Fig. 1):

C' = A, First common character in CAPTCHA and real password.
Loc2 = 3, location of C' in CAPTCHA.
Since it is a two-digit no. (12) both the digits are added (1 + 2).
Loc1 = 1, location of C' in real password.
New Character NC = c.

In simple terms, the user identifies the common character "A" that comes in his real password (Alics%91) and the CAPTCHA and notes down its location, i.e., 12. Since the location is a two-digit number, he adds the two digits, 1 + 2 = 3. So he counts three characters more from "A" in Alice%91, which will be "c". Thus, the new character "sc" has to be entered at the correct location in the login password, indicated by the location digit 4. The other characters are dummy.

Fig. 1 CAPTCHA for example 2



Password:

A2pc9Mxq81nls*

Successfully logged in.

Thus, the password is completely random each time the user logs in it may not even have the same characters.

4 Implementation

The system is developed by using MySQL and Apache server. PHP has been used to develop the front-end software [7], and MySQL has been used to store and retrieve the data from the database. The user interface is made using HTML/CSS.

4.1 Registration Page

During registration, the user has to provide necessary information which will be saved in the database. These include his email, the result location, and a password. The result location is between 1 and 15. It indicates where the user wants to enter the result in the 15 character password at the time of login. The password should be eight characters long and must contain a special character and a numeral. The user is advised to keep a password that is easy to remember (Fig. 2).

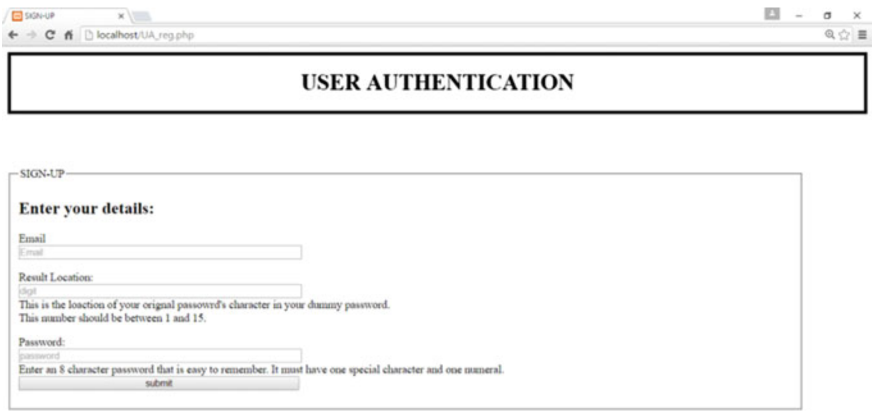


Fig. 2 Registration page

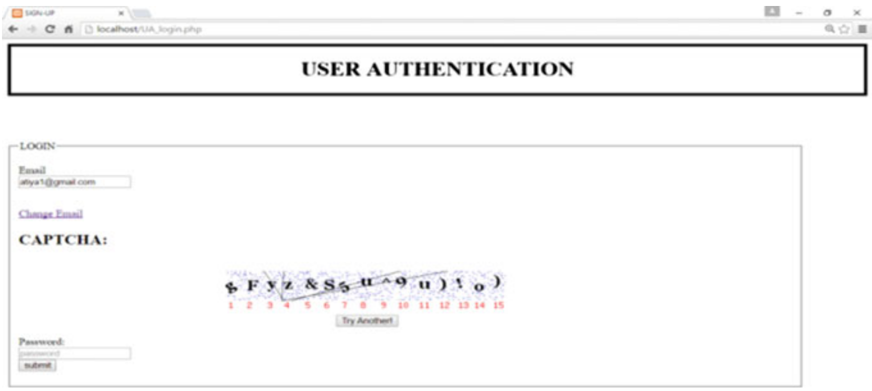
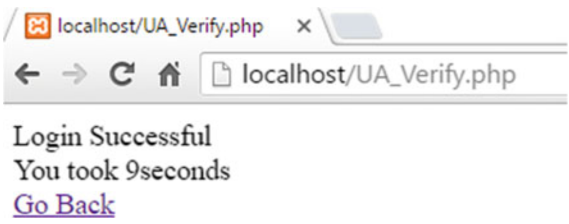


Fig. 3 Login page

Fig. 4 After login



4.2 *Login*

The user is given a challenge based on the information he has provided during registration. The user can login only if he successfully overcomes that challenge (Fig. 3).

The time taken by the user is automatically recorded in the database for each login attempt (Fig. 4).

5 User Study and Results

We asked different people to create an account in our authentication system. Altogether ten people volunteered. Then, we asked each user to log into his account three times and the time taken by the user was automatically calculated and stored in the database. We have assessed the quality of our software based on three main factors: time taken to log in, user-friendliness, and security.

username	time
amrah4@gmail.com	14
amrah4@gmail.com	10
amrah4@gmail.com	14
amrah4@gmail.com	8
amrah4@gmail.com	14
amrah4@gmail.com	11
amrah4@gmail.com	9
amrah4@gmail.com	14
amrah4@gmail.com	9
amrah4@gmail.com	12
amrah4@gmail.com	12
amrah4@gmail.com	7
amrah4@gmail.com	12
amrah4@gmail.com	15
atiya1@gmail.com	9
atiya1@gmail.com	13
atiya1@gmail.com	13
atiya1@gmail.com	7
atiya1@gmail.com	17
atiya1@gmail.com	9
amrah4@gmail.com	9
amrah4@gmail.com	13
amrah4@gmail.com	14
amrah4@gmail.com	51
nadiasid@gmail.com	36

username	time
nadiasid@gmail.com	23
nadiasid@gmail.com	17
ramsha.fatima@zhcet.ac.in	19
ramsha.fatima@zhcet.ac.in	24
ramsha.fatima@zhcet.ac.in	19
zeba.khanam@zhcet.ac.in	12
zeba.khanam@zhcet.ac.in	52
zeba.khanam@zhcet.ac.in	32
zeba.khanam@zhcet.ac.in	25
zeba.khanam@zhcet.ac.in	23
shivangi@gmail.com	46
shivangi@gmail.com	28
shivangi@gmail.com	29
bushrakha9198@gmail.com	28
bushrakha9198@gmail.com	26
bushrakha9198@gmail.com	19
shira@gmail.com	39
shira@gmail.com	22
shira@gmail.com	22
shira@gmail.com	16
bob@gmail.com	31
bob@gmail.com	45
bob@gmail.com	21

Fig. 5 Time taken by each candidate login table

Fig. 6 Average time query

```
MariaDB [ua]> SELECT AVG(time) FROM time_table AS Average_Time;
+-----------+
| AVG(time) |
+-----------+
|   18.6203 |
+-----------+
1 row in set (0.07 sec)
```

5.1 *Time Taken to Login*

We have maintained a database which stores the time taken to log in by each of our ten volunteers. An instance of the database is shown in Fig. 5.

We calculated the average time taken by each volunteer for the entire database and the result was 18.6203 s (Fig. 6).

When we first explained the scheme to a new user, he took significant amount of time to log in. But by the third attempt, the time to log in decreased dramatically. Some experienced users could login in just 6 s. Also, the time depended on the result location chosen by the user. Locations 1 and 15 give the fastest result. The average time take by each user in 3–5 attempts is shown (Fig. 7).

Fig. 7 Average time for each user

username	AVG(time)
amrah4@gmail.com	14.9756
atiya1@gmail.com	10.5000
bob@gmail.com	26.2000
bushrakha9198@gmail.com	24.3333
nadiasid@gmail.com	25.3333
ramsha.fatima@zhcet.ac.in	20.6667
shira@gmail.com	24.7500
shivangi@gmail.com	34.3333
zaonab.zaheer@zhcet.ac.in	21.2500
zeba.khanam@zhcet.ac.in	28.8000

5.2 User-Friendliness

The software is easy to use and does not require mathematical calculations. The given scheme is easy enough to be used by children's as well. The user has to remember only his/her real password. Additional memorization is not needed.

5.3 Security

The real password is never entered or shared. It acts as a key.

Sample space [8] for the real password (key) is given by Eq. (1)

$$S = A^N \tag{4}$$

where

N Length of the password,
A Total keyboard characters (usually 94).

If we calculate the key space using (4) we get $(94)^8$.

5.3.1 Password Space

For every instance of the random CAPTCHA string, the attacker must get the correct character at the correct location. This can be represented mathematically as follows:

$^{94}C_1 * {}^{15}C_1 = 1.4 \times 10^3$

The password space is comparatively less but the scheme is still very secure due to the following reasons:

1. As the proposed challenge is based on the CAPTCHA, a software or a robot cannot be employed to crack the user's password.
2. For each CAPTCHA challenge, the user has just one trial, for next attempt the CAPTCHA would be different. So even by manual brute force, the chances of cracking are nearly 0.001%. And this chance does not increase with subsequent attempts.
3. If by chance the intruder breaks into the user's account once which is very less likely to happen, he won't be able to do so again because he does not know the users key.
4. Each time the user's password is completely different and random.
5. The password is also immune to keyloggers, spyware, dictionary, brute force, shoulder surfing, guessing, eavesdropping, and other attacks which is discussed in the next section.

6 Comparison with Existing Work

- The user never has to enter his real password unlike other text-based schemes.
- Each time a new password is generated.
- Brute Force Attack: The attacker cannot go for exhaustive search to decode the password as we are using CAPTCHA, so he cannot use any computer software for this task. Moreover, the password is completely different each time.
- Shoulder Surfing, Spyware, and KeyLogger Attack: Unlike common text-based password our scheme is immune to these attacks because each time the password is completely different and random. It is impossible for the attacker to guess the correct password even after looking at previous login passwords.
- Dictionary Attacks and Guessing: The passwords are completely random and meaningless which cannot be found in a dictionary.
- Eavesdropping: Even if the attacker gets the password by intercepting the communication between the user and the server it would be completely useless as for the next time the password would be different.
- Our proposed scheme is not as easy as the conventional text-based password scheme but it is really simple and does not involve any complex mathematical calculations.

7 Conclusion

Based on the results, we conclude that the performance of the proposed method is good enough to be implemented. It can be used to protect confidential information like personal data, records, files, accounts. And can also be combined with other softwares and applications for user authentication. We can also incorporate two-step verification method in the scheme. If the user takes longer than usual to log in, we can send an one-time password (OTP) on his phone. The user can get through only if he enters his OTP correctly. That would significantly increase our password space.

References

1. Khandelwal, A., Singh, S., Satnalika, N.: User authentication by secured graphical password implementation. Int. J. Comput. Appl. (0975-8887) **1**(25), 22–24 (2008)
2. Bhanushaki, A., Mange, B., Vyas, H., Bhanushali, H., Bhogle, P.: Comparison of graphical authentication password technique. Int. J. Comput. Appl. **116**(1) (2015)
3. Kushwaha, B.K.: An approach for user authentication one time password (numeric and graphical) scheme. J. Global Res. Comput. Sci. **3**(11), 54–57 (2012)
4. Banue, S.S., Shedge, K.N.: CARP: CAPTCHA as a graphical password based authentication scheme. Int. J. Adv. Res. Comput. Commun. Eng. **5**(1), 14–18 (2016)
5. van Oorschot, P.C., Wan, T.: TwoStep: an authentication method combining text and graphical passwords. In: Babin, G., Kropf, P., Weiss, M. (eds.) E-Technologies: Innovation in an Open World. MCETECH 2009. Lecture Notes in Business Information Processing, vol. 26, pp. 233–239. Springer, Berlin, Heidelberg (2009)
6. http://searchsecurity.techtarget.com/definition/brute-force-cracking. Accessed 15 Nov 2017
7. https://www.sitepoint.com/simple-CAPTCHAs-php-gd/. Accessed 15 Nov 2017
8. https://en.wikipedia.org/wiki/Password_strength. Accessed 15 Nov 2017

Modeling and Control of a Six-Wheeled Mobile Robot

Jayprakash Rout, Subham Agrawal and B. B. Choudhury

Abstract This paper is based on the motion of a mobile robot which is capable of working in a challenging environment. It is a six-wheeled mobile robot with four legs at each corner and other two attached to the middle of the plate. The legs are essential assemblies of robot because they play an important role in a contact with the ground and carry heavy loads. The robot can change its position in any direction. As a main part of the robot, a microcontroller is used which transfers the signal to the relay board and it controls the wheel motion indirectly. An ultrasonic sensor is used in this robot to help in detecting obstacles by emitting sonic wave. Primary object is to detect the distance between the sensors and object. Robot under study is found to be complex and a bit costlier compared to other robots, but it has the capability to traverse different kinds of obstacles. The robot has been modeled in CATIA and analyzed using ANSYS. Different amount of force up to 8 N has been taken and observation for total deformation of the structure, directional deformation in x-axis and y-axis of the robot, maximum principal stress, and minimum principal stress has been made by using structural steel as the material for the robot.

Keywords Mobile robot · Microcontroller · ANSYS · Solid model

J. Rout · S. Agrawal · B. B. Choudhury (✉)
Indira Gandhi Institute of Technology, Sarang, Dhenkanal, Odisha, India
e-mail: bbcigit@gmail.com

J. Rout
e-mail: jayprakashrout420@gmail.com

S. Agrawal
e-mail: sagwl96@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_33

1 Introduction

1.1 Overview of Mobile Robot

A mobile robot can be called as an automated machine or device which shows different motion. It can be autonomous which means it can navigate an environment which is not suitable for it without the help of electrochemical and physical guidance devices. Mobile robot is more familiarly used in industrial and commercial application. A warehouse uses robotic system in order to shift materials from shelves used for stocking those materials to those areas or zones where order fulfillment occurs. Mobile robot is also considered to be an area under focus of current research, and every major university has been found to have one or more laboratories that focus on mobile robot research and development. Mobile robots are usually found in the industrial, security, and military settings. The domestic robot is also a consumer product that includes entertainment robot and performs household tasks such as gardening and vacuum cleaning. In a mobile robot, various components are used such as controller, software to control various parts, sensors to detect the environment, and actuators. The controller can be either a microprocessor or microcontroller. Mobile robot control software may be an assembly level language or high-level language.

1.2 Objective

The main aim is to study different techniques of mobile robot and observe their performances under different conditions. In this study, the aim is on the direction that the mobile robot can move in rough surface as well as smooth surface and it can move also in surface which is sloped. In previous studies, many researchers mainly concluded that the mobile robot can have rotation of only 180°. In this work, the main focus is that the robot can move in any direction and have a rotation of 360°. The mobile robot in this study is more stable, and its loading capacity is more. Different types of sensors are used like IR sensor, proximity sensor, optical sensor, humidity sensor. In this robot, ultrasonic sensor is used for detection.

1.3 Literature Review

Shaha et al. [1] presented a new and innovative design of the wheeled robot, which may be used for military, disaster response, and industrial function. The modification of actual height could be achieved by the use of heavy hydraulic and pneumatic systems which is the main subject in this robot design. Abdalla et al. [2] study a robot which is comprises a cell phone to which the robot is connected through GSM network and which is used as a remote controller also. The robot movement desired location is controlled by the DTMF technique. Diallo et al. [3] introduced a useful model robot whose efficiency is determined based on how well it confines itself around a certain place and interacts with the humans. They proposed that a cruise guide robot element with sonar and IR sensors for obstacle detection and avoidance and an RFID reader for localization. Mareka et al. [4] give out that of a home security system based on different wireless sensor network incorporated with a self-reconfigurable modular robot. Mobile robots these days find applications in security and surveillance. This robot can be controlled by mobile phone by the use of DTMF or RC system. Calderon et al. [5] extend that to expand a cost capable mobile base to be used for delivery in healthcare amenities. An automated mobile base will help appreciate the productivity of strategy tasks in the hospital, like collection and delivery. Yusoff et al. [6] presented a robot for wireless inspection of an air conditional line. The joystick can be used to control operation and speed of wireless inspection robot. Gautam et al. [7] give out that a human–machine interface that is multi-model to control a smart wheelchair using facial movements. A human–machine interface (HMI) based on head action whose purpose is that the left and right change of the head will lead to controlling the robot dynamics.

2 Design and Development of the Robot

2.1 Basic Concept

This paper consists of a robot with ultrasonic sensor taken for the observation. The mobile robot is modeled in the CATIA software and is simulated for different parts such as wheel, motor part, base part and analyzed in ANSYS software for total deformation. Directional deformations in three principal directions are also done. In this robot, the base plate is taken which is made of structural steel and aluminum alloy. Four-relay unit board is attached in Arduino board which provides a path to transfer data to various electronic parts. For power supply to the circuit, a 9 V battery is used. External power supply is provided by LAVSMF12 V-7.2 Ah. An ultrasonic sensor is used which generates the waves which have an important role in measuring the distance between obstacle from the robot. For moving of wheel in 360°, a motor is used. A supply of power is provided to motor of 3.5 rpm. A motor of 60 rpm is attached to it. As a result of which the wheel rotates in 360° and the direction of the movement of wheel is controlled by the signal from the micro-controller. Out of the six wheels attached to the base plate of the robot, four (telescopic leg) legs hold the entire weight of robot. Remaining two legs support the variable load applied. A variable load up to 8 N is applied on the base plate of the robot, and stress–strain relation is found in ANSYS software. To design creative qualities, strength, establishment, rigidity and safety, model and shape are

Fig. 1 Six-leg mobile robot

important. It should be kept in mind that perfect equality is desired between robot size, battery, and motor power to get efficient design and reduce the cost factor of the robot. Six-wheeled mobile robot is a battery-operated robot on wheels which is controlled by a remote, and whose primary role is to carry heavy loads. It is fully automated. It is able to navigate staircases and is able to negotiate and overcome steep slopes using its ultrasonic sensor which is used to measure the distances without being connected with the object, which is represented Fig. 1.

2.2 Design

CATIA software is used to design and inspect the assembly. In order to determine the ability to carry heavy loads, structural steel is used in the base part. Frames are made up of mild steel for the support of base plate. The base assembly weights about 8 kg. Dimension of base plate is (500 × 450) mm with the thickness of 5 mm which is represented in Fig. 2. The displacement analysis as well as stress analysis is carried out for all parts of the robot. Base assembly is constructed to bring 10 kg of load along with other accessories such as battery, motor, and electronics board.

In the leg assembly as shown in Fig. 3, four legs are attached to the corner and other two are attached to the middle of the plate.

The electric motor is a device or machine that has brought about significant advancement in the fields of engineering since the invention of electricity. A motor is an electromechanical device which converts electrical energy to mechanical energy. In controlling a robot, there is a necessity of many electronic parts. A microcontroller controls all the parts by processing the information from the sensors.

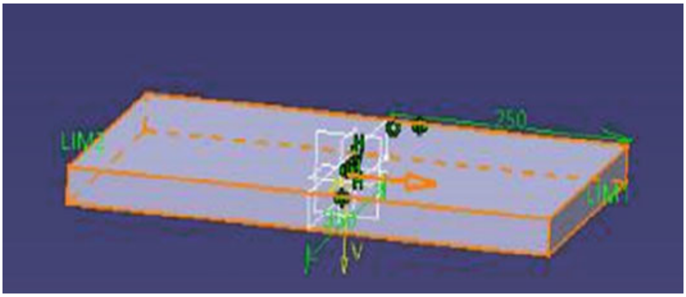


Fig. 2 Base assembly

Fig. 3 Leg assembly

The distance to an object is measured by an ultrasonic sensor by using sound waves. Sonic waves emitted via the transmitter are reflected by an object and received back in the transducer. Basic principle of distanced measurement is based on echolocation. As per this principle, when the sound wave travels in environment, it gets obstructed near the obstacle and returns back to origin as echo. The ultrasonic sensor can be used at several places such as water level measurement and distance measurement.

As per the requirement, the energy supplied is provided by using a battery pack. It contributes some weight to the robot which is required to be kept as low as possible. Here, as a battery pack, two batteries are used. A 9 V non-rechargeable battery is used for microcontroller, and 12 V rechargeable battery is used for driving the motors.

3 Results and Conclusion

3.1 Structural Steel Model

Deformation analysis of a structural steel plate is carried out in workbench of the ANSYS software. First, the robot model is imported to ANSYS by converting the file format extension of the CATIA file. Material property is defined after successful import of the converted file. The property which is necessary for analysis and which has been used is given in Tables 1 and 2.

In this section, the numerical study on deflection and stress developed in the six-wheeled mobile robot which is subjected to different loading is covered. Here the structural steel plate has been loaded with 8 N force.

Figure 4 shows the ANSYS numerical results on the plate of structural steel of the six-wheeled robot. It can be clearly observed that the static structural directional deformation in y-axis is maximum at the sides of four legs which are indicated in red color. The maximum value varies 0.00013353–0.00071859 mm. The static structural directional deformation in y-axis is minimum at middle leg in the wheel part, and its value ranges from −0.00042152 to −0.00035985 mm. The static structural directional deformation in z-axis is minimum at side leg of the telescopic arm. The directional deformation in x-axis is maximum at the middle leg of the wheel part. The maximum value varies 0.00050144–0.00044566 mm. The maximum value of static structural minimum principal stress in case of maximum deformation varies from 0.085621 to −0.68218 mPa. The static structural minimum principal stress is minimum at side of leg of the telescopic arm, and its value ranges between −6.8246 and −6.0568 mpa.

In Fig. 5, it can be clearly observed that static structural maximum shear stress in maximum deformation is in the spring of the four-leg part which is perceived in red color. The maximum value varies from 7.8689 to −0.69946 mPa. The static structural maximum shear stress is minimum at side of leg of the telescopic arm which is clearly noticed and pointed out in blue color, and the value ranges between 0 and −0.87432 mpa.

Table 1 Material property (structural steel)

Compressive ultimate strength	Compressive yield strength	Tensile yield strength	Tensile ultimate strength	Isotropic secant thermal coefficient thermal expansion
0	250	250	460	22°

Table 2 Isotropic elasticity (structural steel)

Young's modulus (MPa)	Poisson's ratio	Bulk modulus (MPa)	Shear modulus (MPa)
2.e+005	0.3	1.6667e+005	76,923

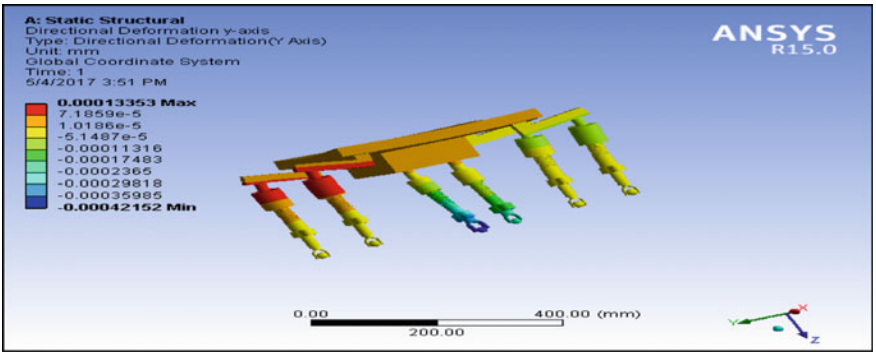


Fig. 4 Directional deformation (y-axis)

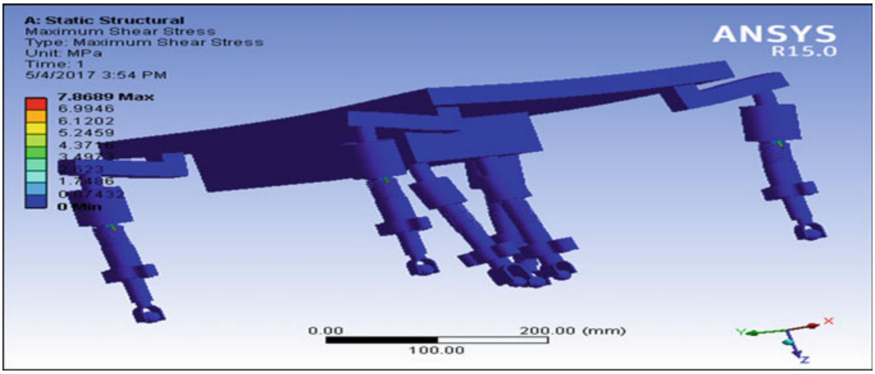


Fig. 5 Maximum shear stress

It is clearly observable in Fig. 6 that static structural equivalent (Von Mises) stress is maximum for deformation in spring of the four-leg part which is clearly observed in red color. The maximum value varies 13.69–12.169 mPa. The static structural equivalent (Von Mises) stress is minimum at side of leg of the telescopic arm which is clearly noticed and pointed out in blue color, and the value ranges between 0 and 1.5211 mpa. It can be clearly observed that static structural error is maximum for deformation in spring of the four-leg part which can be inferred from the red color. The maximum value varies 0.00062261–0.00055343 mJ. The static structural error is minimum at side of leg of the telescopic arm which is observed by the blue color, and the value ranges between 0 and 6.917×10^{-5} mJ. Also, the static structural maximum principal stress is maximum for deformation in the sides of four legs on spring part which is signified in red color. The maximum value varies 8.9181–7.915 mPa. The static structural minimum principal stress is minimum at the plate side of leg of the telescopic arm, and its value ranges between −0.1092 and 0.89319 mpa.

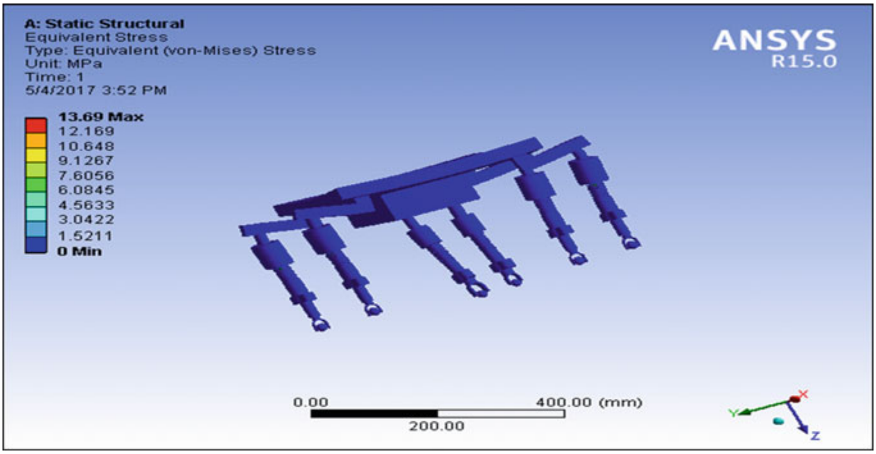


Fig. 6 Equivalent (Von Mises) stress

In Fig. 7, it can be inferred that static structural total deformation is maximum for deformation in wheel part which is indicated in red color. The maximum value varies 0.00074741–0.00066436 mm. The static structural minimum principal stress is minimum at the plate side of leg of the telescopic arm, and its value ranges between 0 and 8.3046×10^{-5} mm.

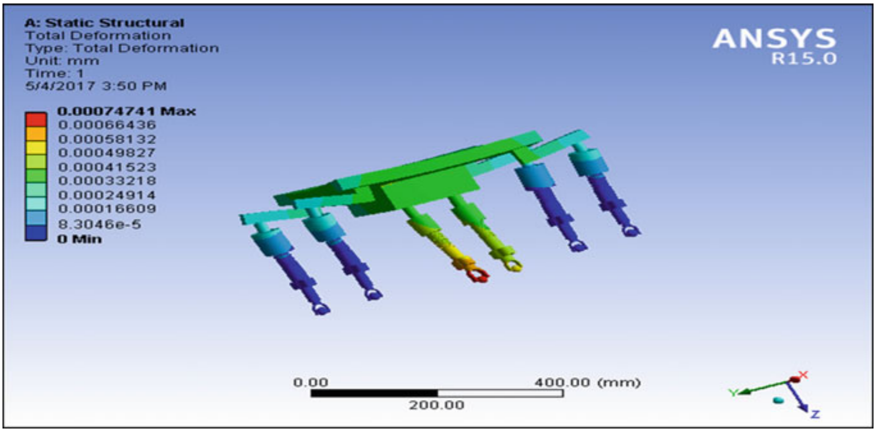


Fig. 7 Total deformation

3.2 Conclusion

Using parameters such as temperature, compressive yield strength, compressive ultimate strength, tensile ultimate strength, tensile yield strength, isotropic elasticity and analysis by ANSYS software, the results for minimum and maximum deformation over time in the categories of deformation like total deformation, directional deformation in x-axis, y-axis, z-axis, equivalent stress, structural error, maximum shear stress, minimum principal stress, maximum principal stress are found by taking the plate of structural steel. This robot can easily move in any environment and on any surface. In the robot, ultrasonic sensor has been used which detects the obstacles at different distances. It can detect obstacles in the range of 4 cm–40 m. We are capable of moving the robot in a desired way by the use of a simple method of controlling based on a mobile phone. The control signal is transmitted via a Bluetooth system. This system is able to perform essential task in areas or situations where there is threat to human life such as areas with radiation pollution or areas where dangerous tasks like disarming bombs and mines are performed. For controlling purpose, Bluetooth electronics applications are used.

References

1. Shaha, R., Ozcelika, S., Challoob, R.: Design of a highly maneuverable mobile robot. Proc. Comput. Sci. **12**, 170–175 (2012)
2. Abdallaa, A.M., Debnathb, N., Ahamed Khanc, M.K.A., Ismaild, H.: Mobile robot controlled through mobile communication. Proc. Comput. Sci. **76**, 283–289 (2015)
3. Dialloa, A.D., Gobeea, S., Durairajaha, V.: Autonomous tour guide robot using embedded system control. Proc. Comput. Sci. **76**,126–133 (2015)
4. Mareka, G., Peterb, Š.: Design the robot as security system in the home. Proc. Eng. **96**, 126–130 (2014)
5. Calderon, C.A.A., Rajesh, E., Mohan, B., Buck Sin, N.G.C.: Development of a hospital mobile platform for logistics tasks. Dig. Commun. Netw. **1**, 102–111 (2015)
6. Yusoff, M.F., Ibrahim, B.S.K.K., Hamzah, H., Kadir, H.A.: Development of air conditional route wireless inspection robot. Proc. Eng. **41**, 874–880 (2012)
7. Ambuj, K., Vasu, G.V., Raju, U.S.N.: Human machine interface for controlling a robot using image processing. Proc. Eng. **97**, 291–298 (2014)

FPGA Implementation of Parallel Transformative Approach in AES Algorithm

Padma Prasada, Sathisha, Ajay Prinston Pinto and H. D. Ranjith

Abstract In recent years, network security is the most critical component in information security. In this research, a new simple yet powerful and fast algorithm for AES is proposed. To have a secured data communication on network usage of an iterative symmetric key block, cipher-based AES is proposed widely. AES is implemented by adopting keys of 128, 192, or 256 bits for encryption/decryption of data in block of 128 bits. These include four transformations in AES: substitute bytes, shift rows, mix columns, and add round key. Here in this approach, parallel transformative method in these transformations mainly in mix columns is proposed. This research mainly focused on the designing of AES according to 192-bit key length in the Verilog language and implementation of it in Virtex6 ML605 FPGA evaluation platform using Xilinx ISE 14.4. To enhance the speed of operation of the algorithm, we followed parallel transformative approach, which achieved throughput of 5565.2173 Mbps with maximum frequency 564.972 MHz in latency of about 13 clock cycles.

Keywords Advanced encryption standard · Virtex6 · Xilinx ISE
Mix columns · Key expansion · Data security

P. Prasada (✉) · Sathisha · A. P. Pinto · H. D. Ranjith
MITE, Mangalore, India
e-mail: ppjain15@yahoo.com

Sathisha
e-mail: mailtosshetty@gmail.com

A. P. Pinto
e-mail: ajaypinto2@gmail.com

H. D. Ranjith
e-mail: hdranjith.web@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_34

1 Introduction

To protect information from intruders, a cryptography technique is used. The prime objective of cryptography is to keep data secure from attackers specially during data communication over Internet. Since 1997, Data Encryption Standard (DES) had been the US government standard.

AES consists of three major components, namely cipher, inverse cipher, and key expansion. A self-contained sequence of operation used to convert data into an unintelligible format is known as cipher, and its reverse operation is called inverse cipher to get the original data back. The key expansion part has special significance in generation of a key schedule for other components. In this proposed work, AES-128 algorithm is employed, and in this ciphertext is generated after ten rounds, where each round contains four transformations, substitute bytes, shift rows, mix columns, and add round key, as shown in Fig. 1. The round keys from the key expansion stage are applied in the add round key step followed by nine rounds of four transformations and last round excluding the mix column transformation.

The applications of AES algorithm are such as smart card and smartphones, Web servers, OFC networks, video recorders so there will be greatest importance in the designing of AES algorithm with simplified and high-performance approach.

2 Literature Review

In recent years, the wide use of handheld wireless devices causes greater amount of networking and wireless data exchange, and the issue of data security is being discussed from different points of views. The National Institute of Standards and

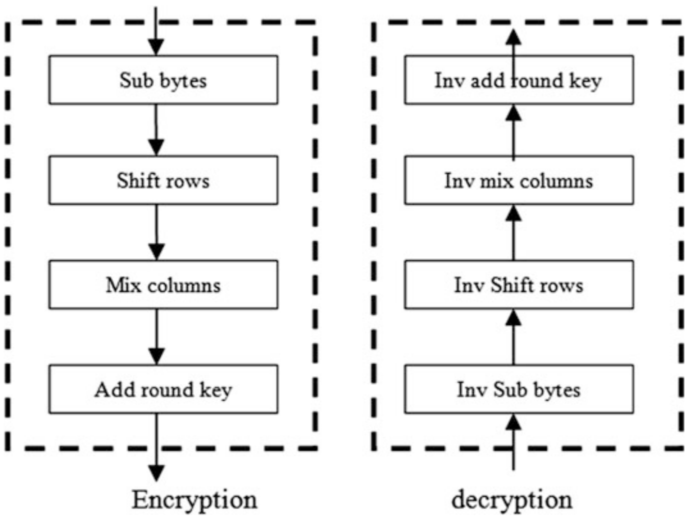


Fig. 1 Different stages in AES algorithm implementation

Technology (NIST) emphasized the use of Rijndael algorithm as a new Advanced Encryption Standard (AES) (FIPS PUB 197, 2001) [1]. This standard was first introduced for the secure data encryption/decryption for high-end applications. A lot of works related to AES also come into picture day by day. In 2001, Buchholz demonstrated AES algorithm in MATLAB [2, 3] and because AES acts upon state matrix, using MATLAB makes clear how algorithm works as the MATLAB programming language is based on matrix manipulation. Dr. Brian Gladman did large amount of research work into the AES algorithm [4]. Author included logarithm and antilogarithm tables to calculate the multiplication of bytes. In 2005, William Stallings shows how AES can be programmed in software or built-in hardware [5].

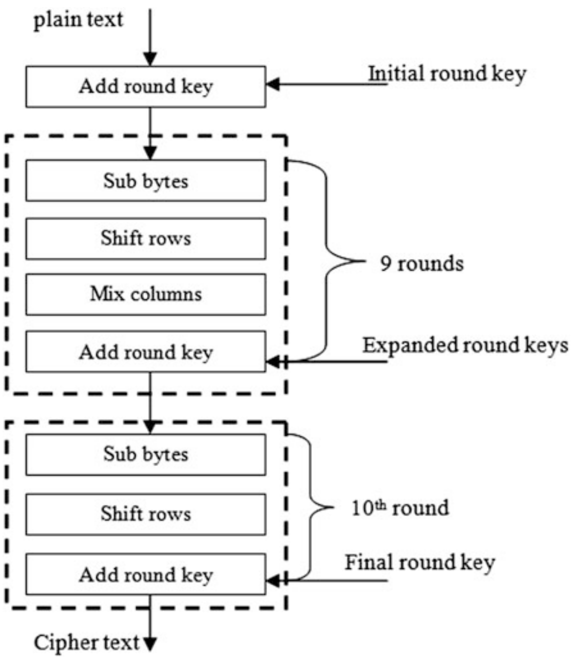
Our work is based on a novel approach to AES algorithm; it emphasizes normal function by two approaches and Strassen algorithm to get more improved matrix multiplication to increase speed of performance.

3 Advanced Encryption Standard

The length of the input, output block, and the state for the AES is 128 bits. This is delineated as $N_b = 4$ and exhibits the number of 32-bit words (number of columns) of the state (Fig. 2).

The length of the cipher key is denoted as K, and it is of 128, 192, or 256 bits. The key length is denoted by $Nk = 4$, 6, or 8, which represents the number of 32-bit

Fig. 2 AES-128 encryption flow diagram



words in the cipher key. The key size takes a significant role in deciding how many number of rounds need to be carried out during the execution of an algorithm. The notation Nr represents the number of rounds, where $Nr = 10$ in case $Nk = 4$, $Nr = 12$ in case $Nk = 6$, and $Nr = 14$ in case $Nk = 8$.

In proposed method, 128-bit data is chosen as the key for the encryption/decryption of the algorithm. For the encryption and decryption actions, there are ten rounds as portrayed in the flow diagram, so Nk value is 4 and Nr value is 10. For each operation, the input plaintext is taken as state matrix containing four rows and four columns.

4 Proposed Design Methodology and Implementation

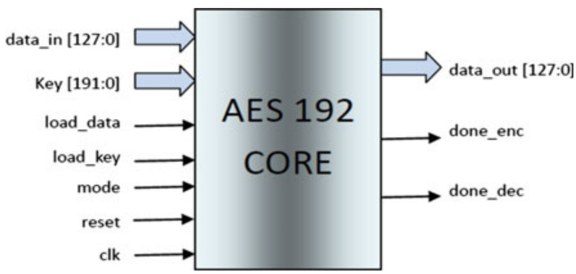
In this work, Verilog HDL was used for coding and the tool used is 'Mentor Graphics ModelSim PE Student Edition 10.1d' for simulation, 'Xilinx ISE 14.4' for synthesis and analysis, and 'Virtex6 ML605 FPGA' for hardware implementation.

The design of AES-192 version consists of input data of 128-bit and 192-bit key input and 128-bit output data. In the encryption, key input goes through key expansion and 192-bit (6 words) key transformed into total 52 words. The 128-bit (4 words) data input initially goes through 'add round key' stage with the first 4 words of the key in the key expansion, and 11 rounds include four transformations: 'sub-bytes,' 'shift rows,' 'mix columns,' and 'add round key' and final round excluding mix columns transformations.

The design additionally consists of mode selection input for the selection between encryption and decryption and indication bit at the output for showing which process (encryption or decryption) is being executed.

The core I/O diagram consists of seven inputs and three outputs as shown in Fig. 3. As normal case within all designs, it consists of clock input. An active low reset, mode input to select the encryption/decryption operation, load_data and load_key inputs to load the data_in and key inputs respectively. The three outputs consist of 128-bit data output which shows the AES-192 result and two indication bits to show that whether encryption or decryption is done.

Fig. 3 AES-192 core IO diagram



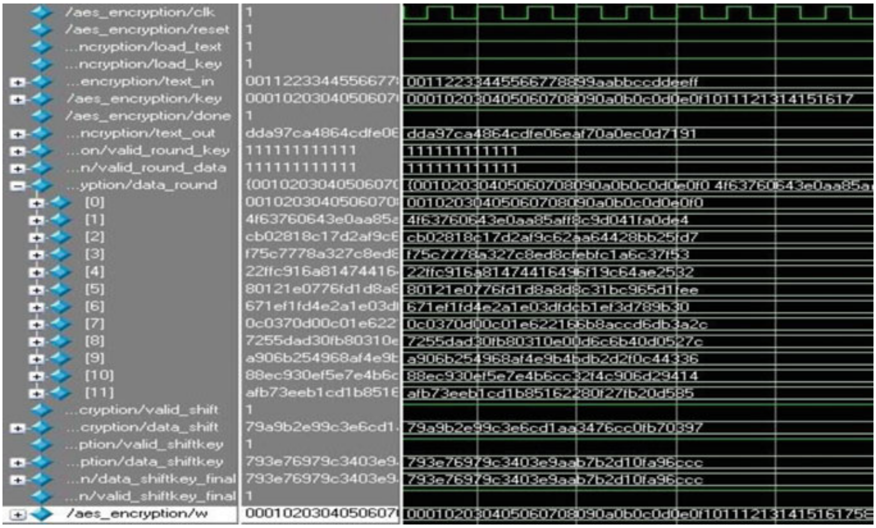


Fig. 4 AES-192 encryption simulation waveform

4.1 Simulation Results

The simulation-encrypted output waveform obtained in the ModelSim is shown in Fig. 4. We give 128-bit data input and 192-bit key input for encryption operation of the AES. Other active low reset is set to 1, to select encryption operation mode is set to 1, load_key and load_data are set to 1 to indicate data, and key is loaded. After 13 clock cycles, we got the resulted output with indication bit set to 1. 12 clock cycles for 12 rounds plus 1 clock cycle for initial key expansion operation.

Decryption is same as encryption except that mode input needs to be set as ‘0’ to select decryption operation and also ciphertext as data_in in order to get plaintext output. The previous encrypted output and same key are given in this stage, and final result got the same plaintext back after decryption. The resultant simulation output is given in Fig. 5.

4.2 Implementation Results

Implementation Results consisting of Synthesis Result obtained from the ‘Xilinx’, Hardware Test Result from ‘Chip Scope pro tool’ of the Xilinx after connecting the ‘Virtex6 FPGA’ and comparison of these results with the previous designs.

Device utilization is very small (Table 1), and as a result we can effectively add other features like fault detection scheme according to the new existent attacks to AES, additional features for special applications like smartphones, smart cards.

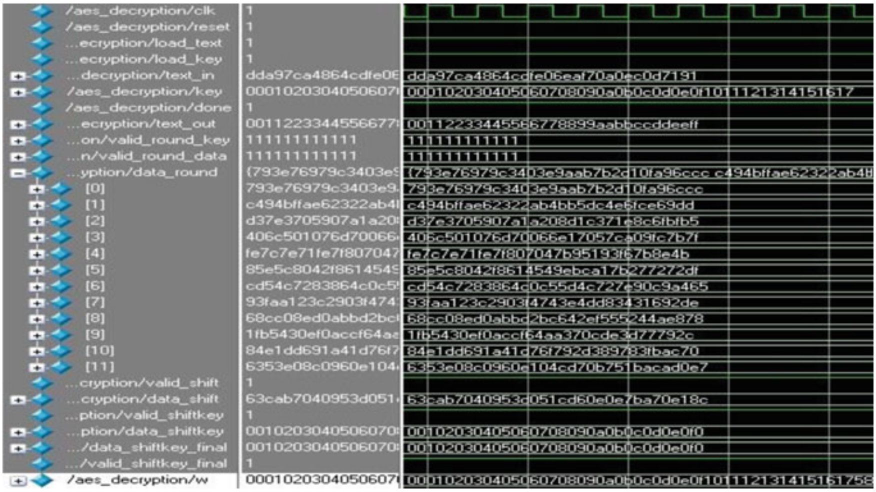


Fig. 5 AES-192 decryption simulation waveform

Table 1 Device utilization summary

Device utilization summary				
Slice logic utilization	Used	Available	Utilization	Note(s)
Slice registers used	2,449	301,440	1%	
Flip-flops count	2,449			
LUTs	841	150,720	1%	
Number used as logic	679	150,720	1%	
Memory	0	58,400	0%	
Average fan out	1.47			

4.3 *Timing Analysis*

The Virtex6 (ML605) FPGA implementation is done using Xilinx 14.4 tool. ModelSim IDE was used for simulating the design. The architecture requires 13 clock cycles to process 128-bit input blocks. The throughput was calculated using the equation

$$\text{Throughput} = \text{Data Block Size}/\text{Clock Time} * \text{Clock Cycles} \tag{1}$$

Table 2 FPGA implementation results comparison

Designs	Propped design	Kenney [6]	Selent [7]	Alaoui [8]	Adib [9]	Trang [10]	Krkljic [11]
FPGA vendor	Xilinx	Xilinx	Altera	Xilinx	Xilinx	Altera	Altera
FPGA chip	XC6YL X240T (Virtex6)	Virtex ii	EP2C35 (cyclone ii)	XC3S10 00 (Spartan iii)	XC5VL X50 (Virtex5)	EP2C20F 484C7 (cyclone ii)	APEX20 KC
Latency (clock cycles)	13	31	40	12	126	11	13
Throughput (Mbps)	5561.21	1570	2546	3220	32,035	1543	1188

Design verification is performed with the help of ‘chip scope pro tool’ of Xilinx. The maximum output frequency found was 564.972 MHz, and clock time is

$$\text{clock time} = 1/\text{maximum frequency} \tag{2}$$

Hence, throughput found is 5.56 Gbps.

4.4 Result Comparison

As compared to some previous designs in Table 2, our design established more efficient in terms of throughput and latency. In the design [9], the throughput obtained is 32.03 Gbps, which is greater than our achieved throughput of 5.56 Gbps. But there are 126 clock cycles used to complete the operation that means high latency results and a lot of time it takes to get the required output. After the analysis of this design, we found that a lot of lookup tables are also used in this design for total operation and include two tables made up of logarithm and antilogarithm is used in the design for the matrix multiplication, which consumes a lot of areas. If throughput is only matters at any cost this design is useful and all other ways most efficient design found is new approach in AES algorithm with parallel transformative approach.

5 Conclusion

In this paper, parallel transformative approach is used in AES algorithm. As technology improves, we can choose most appropriate parameter values for this algorithm. Today, AES is one of the main standards for encrypting all forms of electronic information in the symmetric block-level ciphers. Official specifications of AES are documented with respect to the mathematician's point of view, that the great difficulty software developer's facing during the implementation.

The proposed algorithm implementation followed key factors such as simple and less resource utilization and overhead. In this research, we followed parallel transformative approach in the design to speed up and decrease the delay between the transformations. Throughput of about 5.56 Gbps with maximum output frequency of 564.972 MHz in low latency of 13 clock cycles was obtained.

Acknowledgements We thank Abhijth R, Infosys, Mysuru, India, for the extended support in the preparation of this document.

References

1. National Institute of Standards and Technology (NIST), Information Technology Laboratory (ITL), Advanced Encryption Standard (AES), Federal Information Processing Standards (FIPS) publication 197, Nov 2001
2. Nechvatal, J. et al.: Report on the development of the advanced encryption standard. J. Res. NIST **106**(3), 511–577 (2001). https://doi.org/10.6028/jres.106.023
3. Stallings, W.: The advanced encryption standard. J. Cryptol. **26**(3), 165–188 (2002). https://doi.org/10.1080/0161-110291890876
4. Gladman, B.: A Specification for the AES Algorithm, vol. 3.6, pp. 1–23, 15 April 2003
5. Rouvroy, G., Standaert, F.X., Quisquater, J.J., Legat, J.D.: Compact and efficient encryption/decryption module for FPGA implementation of the AES Rijndael very well suited for small embedded applications. In: Proceedings of ITCC 2004 International Conference, vol. 2, pp. 583–587 (2004)
6. Kenney, D.: Energy efficient analysis and implementation of AES on an FPGA. University of Waterloo, Waterloo, Ontario, Canada (2008)
7. Selent, D.: Advanced encryption standard. Rivier Acad. J. **6**, pp. 1–14 (2010)
8. Alaoui, C.: New experimental results for AES-CCMP acceleration on cyclone-ii FPGA. IJCSNS Int. J. Comput. Sci. Netw. Security **10**(4) (2010)
9. Adib, S.E., Raissouni, N.: AES encryption algorithm hardware implementation: throughput and area comparison of 128, 192 and 256-bits key. Int. J. Reconfig. Embed. Syst. **1**(2) (2012)
10. Trang, H., Van Loi, N.: An efficient FPGA implementation of the advanced encryption standard algorithm. In: IC Design Research and Education Center, Vietnam. IEEE (2012). ISBN: 978-1-4673-0309-5/12
11. Krkljic, A., Dokic, B., Skobic, V.: FPGA implementation of AES algorithm. In: Proceedings of the 5th Small Systems Simulation Symposium 2014. Serbia, 12–14 Feb 2014

An Approach to Wireless Sensor Networks for Healthcare Scrutiny in Hospitals

K. Sujatha, K. SenthilKumar, V. Balaji, R. KrishnaKumar and Nallamilli P. G. Bhavani

Abstract In a hospital, a system by which a patient's physical condition and physiological parameters can be continuously monitored is essential. For example, parameters such as blood pressure (BP), cardiac rate, and fetal movements need to be taken in order to properly manage their condition. This chapter focuses on a scheme that has the ability to monitor physiological parameters from various patients. In the planned scheme, a director node is attached to the surface of the patient's skin to gather information from the unwired sensors and transmit it to the ground station. This scheme can sense anomalous conditions and can produce a corresponding alert signal for the patient. At the same time it can pass on a message to the mobile service or send an e-mail to the patient's general practitioner. In addition, the planned scheme contains a number of reliable, unwired relay nodes used to transmit the information passed on by the director node to the ground station. The key benefit achieved during assessment is a decrease in the energy used to extend the system's existence, speed up and the extent to which the communi-

K. Sujatha (✉)
Faculty of Electrical Engineering, Department of EEE, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. M.G.R. Educational & Research Institute, Vels University, Chennai, India
e-mail: drksujatha23@gmail.com

K. Sujatha
Dr. M.G.R. Educational & Research Institute, Chennai, India

K. SenthilKumar
Faculty of Electrical Engineering, Department of ECE, Dr. M.G.R. Educational & Research Institute, Chennai, India

V. Balaji
Faculty of Electrical Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia

R. KrishnaKumar
School of Electrical Engineering, Vel's University, Chennai, India

N. P. G. Bhavani
Department of EEE, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. M.G.R. Educational & Research Institute, Vels University, Chennai, India

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_35

cation exposure to boost the liberty for increasing the patient's life. Thus a multi-user design for infirmary healthcare has been developed and its performance weighed against existing accessible networks. It is supported by a multi-hop relay node in line with exposure to energy usage and speed.

Keywords Wireless body sensor network · Wireless sensor networks Healthcare and body sensor networks · Body sensor networks

1 Introduction

Body sensor network schemes improve the healthcare environment of the populace by offering healthcare services, such as therapeutic scrutiny, recall enrichment, the use of therapeutic statistics, and through contact with the health maintenance supplier in crisis situations via the mobile messaging system. Continuous monitoring via wearable, implanted transducers and body sensor networks will increase the detection of crisis situations and conditions. In addition, they offer value to the scheme, by remotely obtaining and examining physiological signals devoid of interruptions that can otherwise arise as result of the patient's lifestyle, thereby civilizing standards of existence [1, 2].

Even though current schemes allow continuous observation of a patient's vital signs, these schemes require sensors that are integrated into computing devices positioned at the patient's bedside for supervision purposes. At presently, the sensors and the bedside apparatus are somewhat unreliable due to unwired devices and unwired networks not being available at times. Such schemes do not require the patient to be restricted to their bed and permit them to go about their daily business to some degree, but they do entail the patient being no more than a specific distance from the bedside monitor. However, the exposure to the infrastructure-oriented network changes for every point in time or every different location. Occasionally, a wireless network is not accessible, or if it is accessible it may not be possible to contact the network owing to a deficit in bandwidth [3, 4]. So, in light of these problems and limitations, the constant monitoring of a patient's physical condition is extremely difficult to achieve and in a crisis signals may be blocked at a point between their source and their destination [5, 6].

2 Literature Survey

In a wireless sensor network, by organizing the healthcare services which depends upon the medical field analyzed. Depending on the hospital, many mobile sensor nodes are available which form the cluster group of each mobile sensor based on the coverage over a specific distance using the ZigBee network. Here, each group of clusters having the number of mobile sensor nodes like that is patients to be considered and use a clustering algorithm to elect the coordinator node from the mobile sensor nodes in the wireless sensor network based on the distance between base station and group of clusters. Each node transmits data that is considered to be energy or power to the coordinator node which then forwards it to the base station instantly. The base station collects all the mobile energy statuses according to details of the patient's specific health issues (blood pressure, diabetes, fever, malaria, cancer and so on) via the status of the related energy level to be considered. When health status of the patients is find out the symptom based on the diseases from signal range of wireless hospital network server via base station in the WSN. So, each mobile node utilizes the signal processing algorithm in order to detect the coverage network of the hospital server and to protect the power supply along with the voltage of the hospitalization with different specialist handled by remotely monitored in the wireless sensor network. Patient symptoms collected from the base station are then forwarded to the hospital server and placed into the backup of patient records. Each cluster of mobile patients can be switched to another remote cluster based on their mobility condition. The disease symptoms are filtered using machine learning techniques in the wireless sensor network [7, 8]. This scenario is used to implement the unique key for each mobile node assigned from the base station and to monitor the key changes when moving from one network to another network of patients and their symptoms. It is easy to detect the misuse node or the unknown node in the network, and also the relay node in order to forward data from the patient node to the hospital server via the base station. The hop count mechanism can be used due to the reduced need for power through the use of the sleep mode or idle mode of the mobile sensor.

If a patient requires a checkup or has a health issue update for the hospital then the base station only receives proper disease information in the active mode based on a patient's symptoms. The doctors can then provide medicine to treat the issues those patients are experiencing regarding their health. So, to save the energy level using Power ON or OFF methodology under energy status for forwarding the data to the base station location. Mobility tracking is easily achieved based on use of a localization algorithm which tracks the cluster group locations that were updated to the hospital server. Easy mobile user tracking along with a patient's unique random key allows symptoms of disease to be safely identified as quickly as possible using a fast randomized algorithm. Once the unique key of symptoms is used it automatically changes and updates a new key as soon as possible, dependent on location. So, a patient's energy is maintained via parameters such as energy consumption, power saving, data aggregation related to the symptoms of disease, end-to-end delay, latency, packet delivery ratio, throughput, bandwidth range, signal coverage, lifetime of network, sensor lifetime, and so on [9, 10].

3 Proposed Work

In the system proposed in this chapter, physiological signals are acquired by sensors attached to the patient's body and then broadcast to the remote base station. A personal computer is then used to collate and analyze the data. The signal strength of access points can be weakened by 30–90% as it passes through obstructions (for example, when the two wireless devices or relay nodes are in different rooms and the door which connects the rooms is closed). With an increasing number of obstructions between the nodes, we observe more packet loss and more dead spots that may cause a break in communication between the patient and the network. As a result we have to increase the number of relay nodes within indoor locations to cover the whole environment and enhance the reliability of wireless connectivity. In addition, an emergency alert service using the short message service (SMS) has also been added to the proposed system for emergency responses and rescues. In this chapter, the design of a universal healthcare prototype system for hospitals is discussed. The idea of this healthcare system is to place unobtrusive wireless sensors on a person's body to form a wireless network which can communicate the patient's health status to a base station connected to a monitoring personal computer. The architecture and application of the proposed system are shown in Fig. 1.

4 Efficient Energy Clustering Protocol (EECP)

Clustering routing provides high quality stability along with an investigation and utilization exploration capability with the minimum amount of recall requirements. A distance energy cluster configuration that takes both the distance and leftover energy of the nodes into consideration is presented in this chapter. This improves the procedure of cluster head selection and the process of data broadcast. In addition, it reduces energy consumption. The non-uniform allocation of nodes in the system shuns direct communication between the base station and the distant low-energy cluster head. The results of the simulation show that the improved algorithm efficiently balances energy consumption, prolongs lifetime, packet delivery, throughput, and energy harvesting, and shows a better performance when using the EECP as shown in Table 1.

In terms of wireless sensor nodes, this application uses a total of 44 nodes as presented in the simulation window. The cluster group has four nodes in addition to one for each base station and server of the node. The size of the network is 2000×1100 of environment designed for mobile nodes are performed and usage of implementation protocol is EECP for consume the energy status level and also to ensure that the predicted specified low status energy level can be achieved. Several different algorithms are used in this application with each one implemented in a specific module. The mobile sensor nodes are range, throughput, bandwidth,

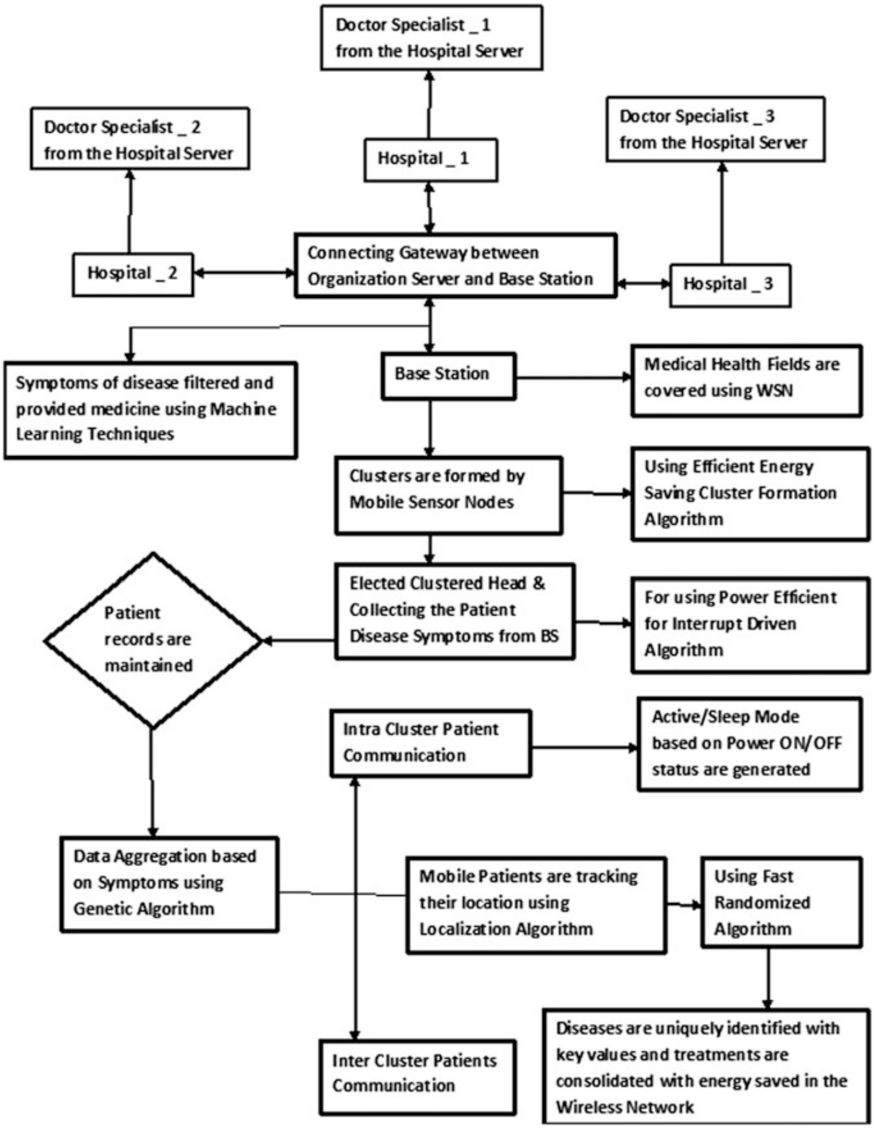


Fig. 1 Architecture of a wireless sensor network system for healthcare monitoring

frequency, packet rate and packet delivery ratio to be highly ensure for consume the energy status level via to predict the disease of such patients to be cured and well as soon in the wireless sensor network. There are end of the simulation time for entire progress of the application in the network simulation window.

Table 1 Parameter settings for EECP

S. no.	Network parameters	Value/Type
1	Total number of nodes	44
2	Number of cluster groups	4
3	Number of base stations	1
4	Number of servers	1
5	Network size	2000 $\times$ 1100
6	Protocol usage	Efficient energy clustering protocol
7	Algorithm design	EECSA, genetic, localization and fast randomized
8	Range	500 m (300–600 m)
9	Throughput	50 Mbps (9.0, 20 Mbps)
10	Bandwidth	400 Mbps
11	Frequency	50 Hz
12	Packet transmission	1000 Bytes
13	Packet rate	250 Packets per second (pps)
14	Request message interval	10–50 s
15	Energy status	100 J assigned
16	Node status	Active and sleep modes
17	Simulation time	5000 s

5 Results and Discussion

The received symptoms of disease medicine can be forwarded to the specific patient from the cluster head. Each cluster is parallel to the same scenario used to discover the symptoms of disease and provide an easy way of saving energy based on replication of symptoms that are then filtered in the medical field organization as illustrated in Fig. 2. A localization algorithm for intra- and inter-cluster communication is used in the organization for forwarding symptoms of disease level stored in the hospital server. It is then easy to identify a related specialist and obtain medicine from the organization as indicated in Fig. 3. Finally, again form the new cluster group with elect the new cluster head and it progresses by identifying the disease status based on the movements of mobile nodes from one group to another group located in the field as shown in Fig. 4.

The number of patients is shown on the X-axis and the lifetime of the cluster on the Y-axis. These are used to monitor the prediction of disease symptoms from the intra- and inter-cluster communication to be high level increased where lifetime of cluster to be high status performed in the network as shown in Fig. 5 where the number of patients is on the X-axis and energy consumption is on the Y-axis. The genetic algorithm is used to find the energy level and avoid the replication of disease symptoms in the network. So, easy to energy harvesting predicted based on replica disease to be filtered from the genetic algorithm as in Fig. 6.

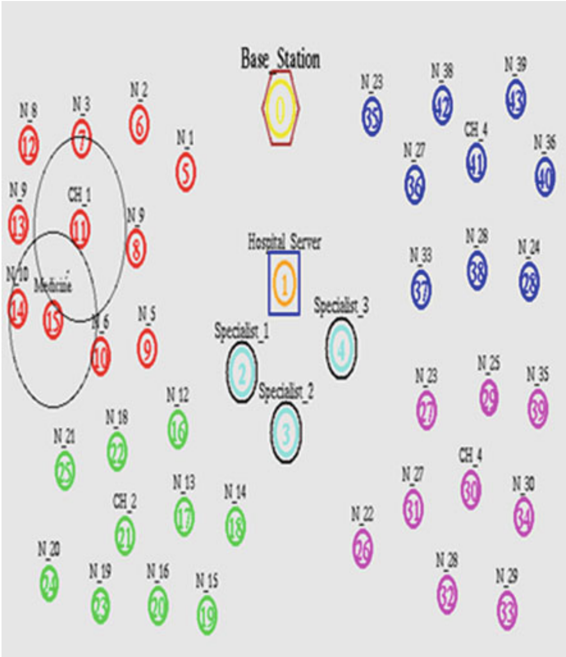


Fig. 2 Medicine received from cluster head

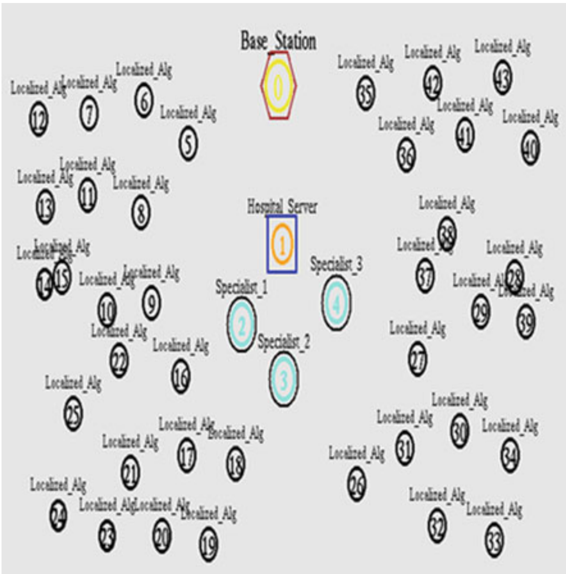


Fig. 3 Recycling cluster group

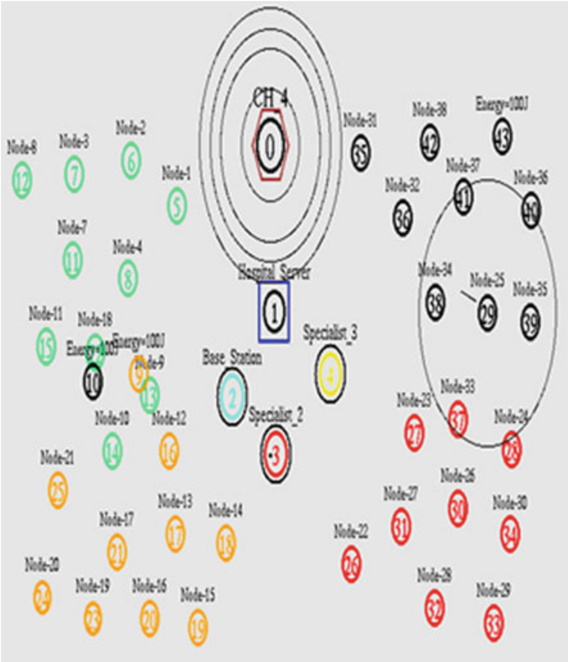


Fig. 4 New cluster group formation

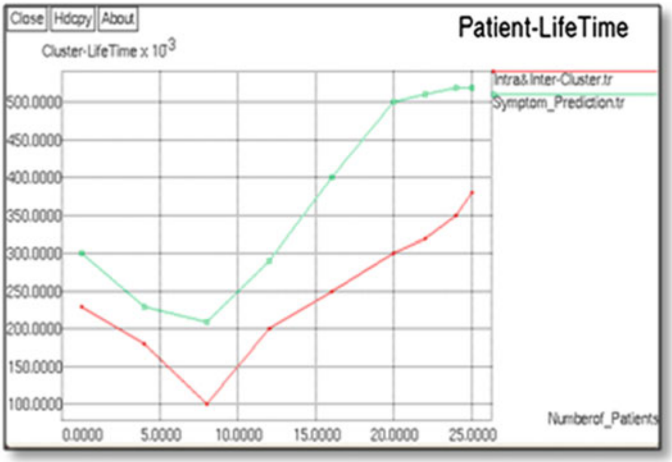


Fig. 5 Patient lifetime

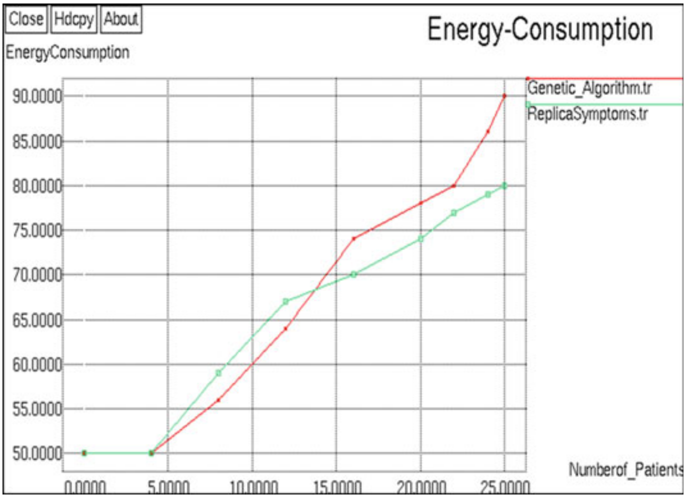


Fig. 6 Energy consumption

6 Conclusion

Wireless body sensor network (WSBN) expertise shows promise as a noteworthy constituent of the future generation of healthcare services. The whole scheme consists of a director node that is used to extract a patient's physiological data, a WMHRN to forward the data, and a base station to gather the data. The scheme is capable of long-term observation and is outfitted with an emergency rescue device using SMS/e-mail. In addition, the planned WBSN has been simulated in either a Network or OMNet++ simulator and evaluated against existing WBSN schemes in terms of coverage, energy consumption, and delay time. It was observed that the planned system showed improved performance compared to existing WBSN systems.

References

1. Stanford, V.: Using pervasive computing to deliver elder care. IEEE Perv. Comput. **1**, 10–13 (2002)
2. Mcfadden, T., Indulska, J.: Context-aware environments for independent living. In: Proceedings of the 3rd National Conference of Emerging Researchers in Ageing, Brisbane, Australia (2004)
3. Anliker, U., Ward, J.A., Lukowicz, P., Troster, G., Dolveck, F., et al.: AMON: a wearable multi-parameter medical monitoring and alert system. IEEE Trans. Inf Technol. Biomed. **8**, 415–427 (2004)
4. Cho, G., Yoo1, S.K.: Wearable ECG monitoring system using conductive fabrics and active electrodes. In: Proceedings of the 13th International Conference on Human-Computer Interaction, Berlin, Heidelberg (2009)

5. Darwish, A., Hassanien, A.E.: Wearable and implantable wireless sensor network solutions for healthcare monitoring. Sensors **12**, 12375–12376 (2012)
6. Shnayder, V., Chen, B., Lorincz, K., Fulford-Jones, T.R.F., Welsh, M.: Sensor networks for medical care. In: Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems, New York, USA (2005)
7. Alemdar, H., Ersoy, C.: Wireless sensor networks for healthcare: a survey. Comput. Netw. **54**, 2688–2710 (2010)
8. Varshney, U.: Using wireless networks for enhanced monitoring of patients. Int. J. Healthc. Technol. Manag. **6**, 489–499 (2005)
9. Varshney, U.: Pervasive healthcare. IEEE Comput. **12**, 138–140 (2003)
10. Lai, C.C., Lee, R.G., Hsiao, C.C., Liu, H.S., Chen, C.C.: A H-QoS-demand personalized home physiological monitoring system over a wireless multi-hop relay network for mobile home healthcare applications. J. Netw. Comput. Appl. **32**, 1229–1241 (2009)

Foetal Heartbeat and Volume Detection Using Spherical Harmonics Shape Model for Level Set Segmentation

K. Sujatha, V. Balaji, Nallamilli P. G. Bhavani and S. Jayalakshmi

Abstract The paper is aimed to estimate the heartbeat and volume of a foetus based on three-dimensional images from ultrasonic using image processing algorithms. Since human soft tissues show a minimal intensity difference, the ultrasonic produces poor quality image. Enhancement of the boundary is done to recover the deprived image which is the input image for segmentation. Simultaneously, the superfluous edges in the foetus image are removed to enhance the quality. Nonlinear dispersal techniques were incorporated for foetal image enhancement method. Foetus images are processed using segmentations based on iso-intensity and edge focusing to achieve apt and comprehensible edge image. The iso-intensity shape method is applied to sort the pixel values with equal intensity. Another shape-based method, which has been used, is edge focusing, and this method depends on the disparities of significance between dissimilar boundaries in the foetus picture. When outline is traced out as a series of the points in the three-dimensional volume, then the shape of the foetus can be fitted. The normal profile of anterior or posterior region resembles a sphere or an ellipse which needs to be fitted. As soon as an analytical figure is fit, the volume of the foetus and therefore the heartbeat and volume can be premeditated. These image processing algorithms are developed using image acquisition tool from MATLAB 2013a to acquire and process the foetal images. The results are finally validated using the error value.

K. Sujatha (✉) · N. P. G. Bhavani
Deptartment of EEE, Center for Electronics, Automation and Industrial Research (CEAIR), Dr. MGR Educational & Research Institute, Chennai, India
e-mail: drksujatha23@gmail.com

N. P. G. Bhavani
e-mail: sbreddy9999@gmail.com

V. Balaji
Faculty of Electrical Engineering, Bahir Dar Institute of Technology, Bahir Dar University, Bahir Dar, Ethiopia
e-mail: balajieee79@gmail.com

S. Jayalakshmi
Deptartment of EEE, Ramachandra College of Engineering, Eluru, Andhrapradesh, India
e-mail: jayasriram23@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_36

Keywords Ultrasound image · Speckle reduction · Nonlinear filter

1 Introduction

It is vital to monitor the physical condition of the mother and the foetus cautiously by a concerned specialist. State of the foetus is regularly found out from a foetal ECG, blood test, and ultrasound scans. The ultrasound image helps to identify the developmental stages including heaviness of the foetus [1], to find the foetal abnormalities, and to summarize about the resultant health care of the pregnant woman and foetus. Foetus assessment using image processing to find the status and to envisage the existence of the abnormalities is identified without human intervention during initial three months of pregnancy. Since images are distinct as compared with two-dimensional digital images, the replica in the form of multi-dimension may be avoided. The power of the manifestation finds the intensity of the image pixels, whilst their influx time determines their spatial setting in the image [2]. The traditional signal is subjected to interior processing and is then transformed into a video format (NTSC, RGB, component, etc.) for display on a normal monitor or for recording purposes.

Image enhancement is to visualize and assess the visual information contained in it with high precision [3]. Image segmentation is defined as the division of the image into distinct sections [4]. The major target of image separation is to divide into segments which are consistent with respect to one or more traits. Segmentation techniques can be broadly segregated into two sets: region-based and edge-based methods. For the foetus shape segmentation, the iso-intensity outline and edge-focusing segmentation methods are implemented. The obtained set of the points form a curve that fits the foetus [5]. The most important parts of the foetus are the head and body which have a form of the irregular sphere. It was decided to fit them with following geometrical figures: sphere and ellipsoid [6, 7]. After the figure fitting, the volume of these figures is calculated.

2 Literature Review

Falcon et al. propose methodologies to create the rapport amid the foetal trunk and head volume calculated by three-dimensional (3D) ultrasound and gestational age at 11 + 0 to 13 + 6 weeks of gestation. J. Weickert et al. propose in his work “Applications of nonlinear diffusion in Image processing and computer vision” nonlinear dispersal process which can be created by numerous new methods for image processing and computer vision.

The mathematical design is dependent on additive operator splitting (AOS). In contrast to traditional multiplicative splitting such as ADI, LOD, or D’yakonov splitting, all axes are treated in the same manner, and additional possibilities for proficient realizations on parallel and distributed architectures materialize the entire need. Geodesic active curves show the way to equations that reminds of the mean curvature motion. For this application, a new AOS plan is offered that employs harmonic averaging and does not need re-initializations of the distance function in each iteration step [8].

3 Related Works

The basic works were carried out based on image segmentation techniques. Some of the prominent segmentation techniques were as given below.

3.1 Preprocessing

As a part of preprocessing, histogram analysis and surface plotting are carried out to identify the variation between the normal and abnormal images for further processing. Figures 1 and 2 show the histogram and surface plot analysis.

3.2 Region Segmentation Techniques

The division of an image into smaller parts is for resolving the boundary of smaller region called region segmentation. There are quite a few techniques used for region segmentation such as thresholding [9].

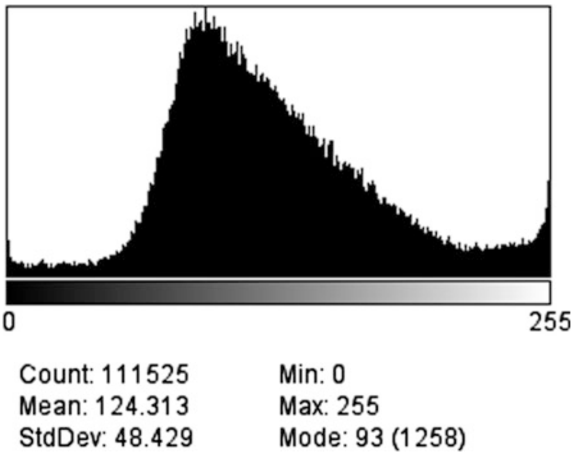


Fig. 1 Histogram analysis

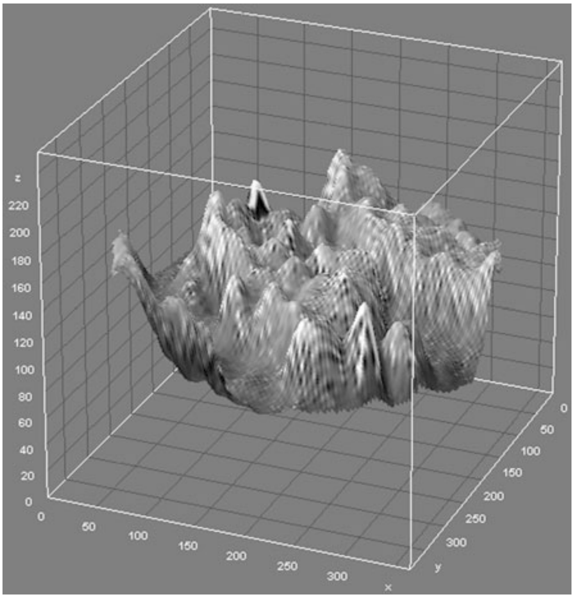


Fig. 2 Surface plot analysis

3.3 Thresholding Technique

A few of the thresholding methods depend on the image histogram. Other thresholding methods are reliant on confined properties such as local mean and standard deviation, or the confined gradient. If the local thresholds are selected independently for each pixel, it is called as dynamic or adaptive type [10]. The sample database is displayed in Fig. 3.

3.4 Edge-Based Segmentation Techniques

The periphery is denoted by the local pixel intensity gradient. A rise is the rough calculation of the first-order derivative. Both magnitude and directions of the gradient can be displayed as images. The magnitude of the grey levels that is comparative to the amount of the confined intensity changes, whilst the trail images will have grey levels, representing the trend of maximum confined gradient in the original image. The results for edge detection and denoising are shown in Fig. 4.

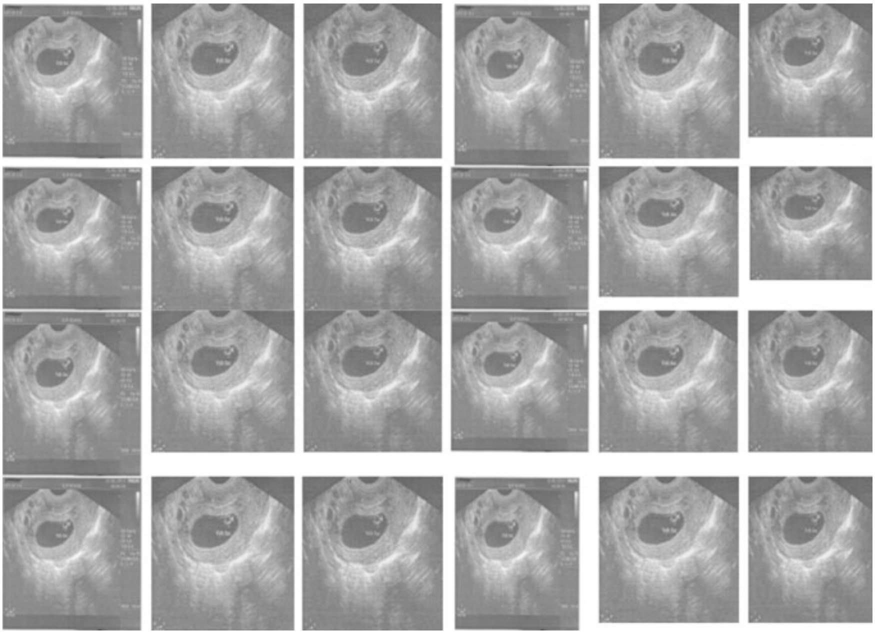
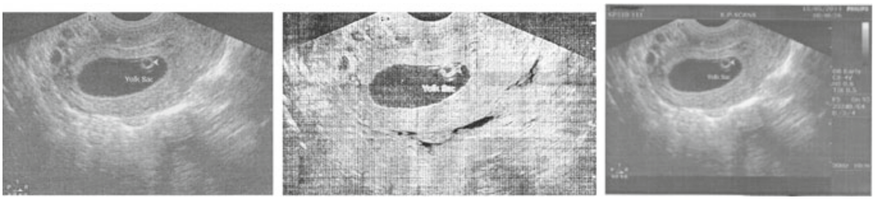


Fig. 3 Images of the foetus used for assessment of foetus volume and heartbeat



(a) Original Image (b) Edge detected image (c). De-noised Image

Fig. 4 Thresholding

4 Proposed Work

Estimation of foetal dimensions was identified and examined to achieve the most favourable solution. The error in volume identification was estimated. To compute the volume of the foetus in given ultrasound image, the following steps are carried out:

I. Civilizing the image feature.
II. Identifying the significant boundaries.
III. Evolving the foetus shape.
IV. Fitting the shape matching to the obtained foetus shape.
V. Scheming the volume of the fitted form.

5 Results and Discussion

The algorithm that is suggested for the automatic volume calculation is promising, but further improvements are necessary. The algorithm consists of the following steps: image enhancement, segmentation, form fitting, and volume calculation. In this study, many image enhancement techniques are described from pixel to global operations. The nonlinear diffusion techniques give the best image enhancement technique. It appeared that the nonlinear Perona and Malik technique works better when the image is blurred first. Maybe the enhancement can be further improved by using more techniques consecutively. Figure 5 shows the heartbeat obtained by monitoring the foetus. The effectiveness of the enhancement technique used is illustrated in Table 1.

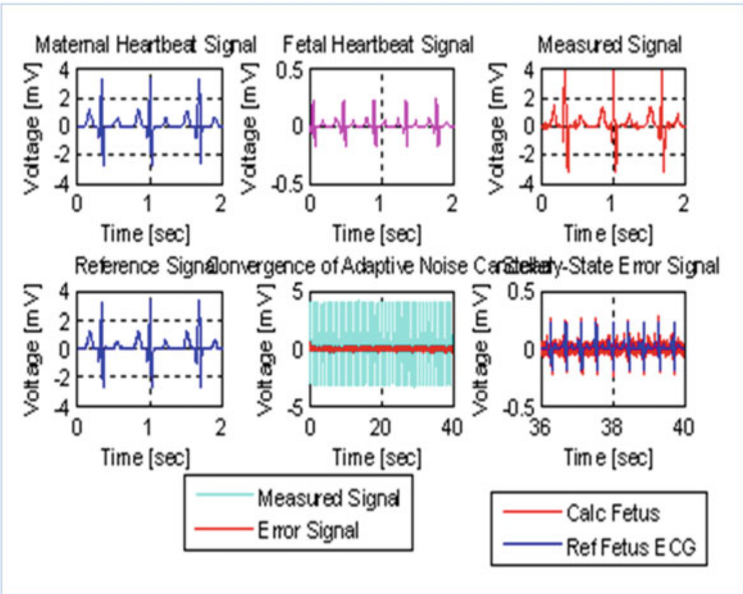


Fig. 5 Results for monitoring the foetal status

Table 1 Calculation of volume error for different algorithms

Image processing techniques		No. of voxels	Volume voxel (cm^3)	Volume of ellipsoid (cm^3)	Absolute error (%)
Enhancement technique	Segmentation technique				
Perona and Malik diffusion	ISO-intensity contour	1,22,425	1.74×10^4	21.32	6.60
Euclidian shortening flow		1,22,230	1.74×10^{-4}	21.29	6.45
Perona and Malik diffusion	Edge focusing	91,887	1.74×10^{-4}	16.00	20.00
Euclidian shortening flow		92,809	1.74×10^{-4}	16.16	19.20

6 Conclusion

The measurement of the volume of a foetus based on three-dimensional ultrasonic sound image using image processing techniques was carried out. Iso-intensity segmentation and edge dependant division methods are implemented on an image to acquire correct and lucid boundary image. By iso-intensity shape method, the pixels with same intensity are chosen. The first two techniques have the disadvantage that they are susceptible to noise. The last technique requires many input parameters. The deformable models might give a better result, but then a closer look at the input parameters is necessary. These techniques need to be further tested on well-defined objects to determine the errors that are made. The error of the fit of the ellipsoid through the points should also be determined. The experiments on a well-defined object should be repeated enough times in order to obtain a sufficient number of output data for statistics. The original form of body or head is like a sphere or an ellipse for which these figures have been fitted. As soon as an analytical figure is fit, the volume of the foetus can be calculated. These image processing methods are implemented for the foetal image using image acquisition tool from MATLAB.

References

1. Benoit, B., Hafner, T., Kurjak, A., et al.: Three-dimensional sonoembryology. J. Perinat. Med. 30 (2002)
2. Abd-Elmoniem, K.Z., Student member. Abou-Bakr, M.Y., Kadah, Y.M., Member: Real-Time Speckle Reduction and Coherence Enhancement in Ultrasound Imaging via Nonlinear Anisotropic Diffusion, vol. 6, no. 3. IEEE (2009)
3. Greenspan, H.: Image enhancement by nonlinear extrapolation in frequency space. IEEE Trans. Image Process. **9**(6), 1035–1048 (2000)

4. Ciurte, A., Rueda, S.: Ultrasound Image Segmentation of the Foetal Abdomen: A Semi-Supervised Patch-Based Approach, Department of Computer Science, Technical University of Cluj-Napoca, Romania
5. Florack, L.M.J., Romeny, B.M.H., Koenderink, J.J., Viergever, M.A.: Scale and the differential structure of images. Image Vis. Comput. **10**, Roven Press, New York (1987); in the third trimester, Obstetr. Gynecol. **95**(4), (2000)
6. J´er´emieAnquez, Angelini, E.D., Senior Member, IEEE, Gilles, G., Isabelle, B., Member, IEEE Automatic segmentation of antenatal 3-D ultrasound images. IEEE Trans. Biomed. Eng. **60**(5) (2013)
7. Klingensmith, J.D., Shekhar, R., Geoffrey Vince, D.: Evaluation of three-dimensional segmentation algorithms for the identification of luminal and medial–adventitial borders in intravascular ultrasound images. IEEE Trans. Med. Imag. **19**(10) (2000)
8. Sujatha, K., Pappa, N.: Combustion quality monitoring in PS boilers using discriminant RBF. ISA Trans. **2**(7), 2623–2631 (2011)
9. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intelligen. (1990)
10. Florack, L.M.J., Romeny, B.M.H., Koenderink, J.J., Viergever, M.A.: Scale and the differential structure of images. Image Vis. Comput. **10** (1992), Roven Press, New York, (1987) in the third trimester, Obstetr. Gynecol. **95**(4) (2000)

An Efficient Approach for Clustering US Census Data Based on Cluster Similarity Using Rough Entropy on Categorical Data

G. Sreenivasulu, S. Viswanadha Raju and N. Sambasiva Rao

Abstract In the field of data mining, clustering is one of the major issues. In the categorical clustering, data labeling has been acknowledged as an important method. The grouping of all the similar data points together is called as clustering. Those points which are not labeled earlier go through the data labeling process. For categorical data, very limited algorithms are applied, although there are many approaches in the numerical domain. In categorical domain, the most challenging issue is to allocate all the unlabeled data points into proper clusters. In this paper, a method is anticipated for labeling and maintaining the similar data points into proper clusters. We have a data set named US Census, where the data was collected as part of the 1990 census. There are 68 categorical attributes. This data set was derived from the US Census 1990 raw data set. The new proposal is to allocate each unlabeled data point into the equivalent proper cluster with data labeling also. It is much useful to understand the demographic survey of the public. This method has two rewards: (1) The proposed algorithm exhibits high execution efficiency. (2) This algorithm can achieve superiority clusters. The proposed algorithm is empirically validated on US Census data set, and it is shown considerably more efficient than previous works while attaining results of high quality.

Keywords Clustering · Categorical data · Clustering · Data labeling
Outlier · Entropy · Rough set

G. Sreenivasulu (✉)
Department of CSE, ACEEC, Hyderabad, India
e-mail: gvsreenu@gmail.com

S. Viswanadha Raju
JNTU Jagityala, Karimnagar, India
e-mail: svraju.jntu@gmail.com

N. Sambasiva Rao
VCE, Hyderabad, India
e-mail: snandam@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_37

1 Introduction

Clustering is a major challenge in data mining [1]. Clustering is the method which is used to group objects together [2, 3]. The scope of grouping mechanism has been complete in Pattern Recognition, Information Retrieval Systems, Medical Diagnosis, Machine Learning, Statistics, etc. There are various types [3] in which the complete point on clustering practice can be originating. In data set, numeric, mixed, and categorical data are the three different types. When the three data types [4, 5] are compared, in numeric data there are greater types of procedures. Data clustering is a complicated task in categorical data since the distance between the data points is not accurate whenever the data is increased on time. When it comes to clustering a large data set, it becomes difficult because of the time it takes for the completion of the process [6, 7]. Examining is another strategy in grouping which is utilized to get the ability of bunching by haphazardly choosing the information focusing on early grouping and identify the unnamed to decide for traditions and intends to allocate them into reasonable groups (unlabeled refers to those which are not sampled and are not clustered). This is nothing but cluster tagging [7–9]. In finding the class field, the numerical field is not that much straightforward in categorical field. The concept drift in data mining is time consuming [10]. The time maturing information in the numerical field for grouping [2, 4, 5, 9, 11] was investigated in the last review writing; however, very little was mentioned in the specific area. In this way, despite everything it states it still has a primary trouble in getting the information straight out. Therefore, our exploration is in changing the procedure of Ming-Syan Chen structure 2009 [7] by utilizing any bunching strategy to discover the floating practice. This paper chiefly manages the point-by-point clarification about the technique and working outcomes on an extensive information collection for information naming in harsh set hypothesis. The unpleasant set hypothesis is said to be a compelling scientific hypothesis in Machine Learning, dimensionality decrease, Internet of Things (IoT), design acknowledgment, wireless ad hoc systems, and so forth, as it has been of extraordinary utilize. Whatever remains of the paper is sorted out as takes after: Thrash out in Sect. 2 is about examination of related writing; Basic definitions in Sect. 3 are talking about the entropy show inside harsh set hypothesis, Labeling of Data in Sect. 4 arrangement for straight-out grouping, in Sect. 5 new outcomes are uncovered and the conclusions and future work are talked about in Sect. 6.

2 Review of Related Literature

We will offer entire portrayal of grouping strategy on all out information for information naming by bunch gathering in this part [9, 10, 12]. Bunching large chunks of data is a painstaking procedure and is not an effective undertaking. To audit and delineate the bunching BIRCH, BRICH is a speed bunching technique which handles "commotion" capably as composed by Parmer et al. [13]. BRICH discovers grouping, and by utilizing a tad bit, it improves the predominance of more output bunches in a first round sweep of information. CLARAN [14] is consistently substandard compared to BRICH which is a grouping strategy plausible for enormous data sets. The best various-leveled grouping approach which maintains a strategic distance from the issues with non-uniform figured or molded bunches is CURE [15]. In CURE, they are contracted toward the mean of the group by segments subsequent to recognizing a settled number and very much stamped objects of a bunch, and these focuses are known as gathering of the group. Since it is agglomerative and divisive (various leveled), the groups with the adjoining close agony of committee are intensified in each progression. It makes it in part touchy to exceptions since it legitimately recognizes the bunches. In grouping move toward K-modes [16], the superiority virtue esteem in both question area of a bunch gathering the most successive for that bunch. Proclamation mode might be exquisite, yet the course of utilizing just a single quality incentive in each credit field to allude a bunch is indeterminate.

The following technique is Robust Clustering algorithm for categorical data. It is a procedure of top-down grouping calculation. For gathering of downright information esteems, (CACTUS) Categorical grouping utilizing outlines is a summation-based approach [17]. The advancement CACTUS at the back is there, "The entire database gross is sufficient to gauge an arrangement of jogger bunches that can be accepted to verify the substance agreement of groups."

There are two extra grouping calculations that are very much loved in straight-out space named COOLCAT and LIMBO. COOLCAT, the ideal entropy information benchmarks is irritated such that the ordinary entropy of the total arrangements is limited. In any case, these strategies complete grouping in view of plunging or most extremely the factual objective capacity, and the bunching committee in these techniques are not legitimately indicated. In this way, the rundown and quality data of the grouping focuses cannot be taken by utilizing these calculations. A disparate technique in view of unpleasant set model in order to get down to earth results is the goal of the paper.

The harsh hypothesis is pointed on the theory that with each purpose of the world, there is connected a persuaded sum regarding data, explained by methods for a few articles which are utilized for information point representation. Objects having a similar clarification resemble with yielding to the available data. The similarity connection in this manner produced constitutes a scientific model of the unpleasant hypothesis.

3 Entropy Model in Rough Sets

D is a data set accumulation of n information esteems, where every information esteem is a variety of q characteristic esteems as xj = (xj1, xj2, … xjq). For instance, A = (A1, A2 … Aq), where Aa is the *a*th category question, N be sliding

Table 1 A data set D with 14 data points divided into two equal size sliding windows S^1 and S^2

S1							
	x_1	x_2	x_3	x_4	x_5	x_6	x_7
F_1	F	H	F	F	R	S	R
F_2	N	N	N	N	N	N	L
F_3	I	I	J	J	Q	Q	Q
S2							
	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
F_1	F	H	F	S	H	I	S
F_2	N	E	L	L	L	E	N
F_3	I	G	J	Q	J	G	Q

window described measure. Defeat the n data centers into proportionate width windows and call this social affair as St, at time between time t. Thusly, first N data reasons for data set D are orchestrated in the primary social occasion S1, and next sliding window data motivations behind D are arranged in the second subset S2, etc. The point of our proposed technique is to take St + 1, as an obscure informational index and speak to these information focuses into the groups which are picked up from St.

For example, consider the going with D as a Data set = $\{x_1, x_2, \ldots x_{14}\}$ of 14 centers showed up underneath, and the sliding window gauge N is 7; then S1 addresses initial 7 data centers, and next sliding window S2 contains next 7 centers anticipated underneath Table 1. Apply any one packing approach on S1 to break the data centers into two gatherings showed up in underneath Table 2. The concentrations that are clustered are called assembled data centers or stamped or requested data centers, and the remaining are called unlabeled core interests. Our technique is to make whatever remains of the 7 unlabeled data centers keeping them into fitting clusters which have a place with next sliding window S2.

By utilizing K-means bunching and K-modes calculation on sliding window (S1) partition information focuses into groups appeared in the beneath table. The focuses that are kept under one gathering are called assembled information focuses or marked focuses, and the left finished are called unlabeled one. Our motivation is to mark the left finished unlabeled focuses into next sliding window in which they have a place with.

4 Rough Entropy-Based Data Labeling

In harsh set hypothesis, unpleasant threshold is an instability data gauge. Be that as it may, in unpleasant set group there are few worries about the issue of bunching the information in light of information marking. Going with subsection, some crucial definition to perform data naming using obnoxious entropy figuring and how to name the data using this measure are analyzed.

Table 2 Two clusters C_1^1 and C_2^1 after performing a clustering method on S^1

C_1^1				C_2^1		
x_1	x_2	x_3	x_4	x_5	x_6	x_7
F	H	F	F	R	S	R
N	N	N	N	N	N	L
I	I	J	J	Q	Q	Q

Example 1: The Table 1 is representing two windows initially. Subsequent to apply a clustering method on window S^1, the clusters c_1^1 and c_2^1 are fashioned as shown in Table 2. In this setting perspective of the information purposes of window S2, data marking of every datum focuses is talked about in this segment. Every U with all properties an A1 is the partition is ascertained by utilizing recipe (1) as

$$U/IND\ (a) = \{\{x_1, x_3, x_4, x_5\}\{x_2\}\}$$
$$U/IND\ (b) = \{\{x_1\}\{x_3, x_5\}\{x_2, x_4\}\}$$
$$U/IND\ (c) = \{\{x_1\}\{x_3\}\{x_5\}\{x_2, x_4\}\}$$

The rough purity for c_i with esteem points $a \in A_1$ is considered by using prescription (3). By applying this logic on $c_{1,}$ meaning of each feature is as below.

$$RE(C_1^1, X_6) = \frac{1}{2}\left(\frac{3}{4} \times \frac{4}{2} + \frac{2}{2} \times \frac{2}{3} + \frac{1}{1} \times \frac{1}{4}\right) = 1.208$$

$$RE(C_1^1, X_7) = \frac{1}{2}\left(\frac{3}{4} \times \frac{4}{2} + \frac{0}{2} \times \frac{0}{3} + \frac{1}{1} \times \frac{1}{4}\right) = 0.875$$

$$RE(C_1^1, X_8) = \frac{1}{2}\left(\frac{0}{0} \times \frac{0}{2} + \frac{0}{2} \times \frac{2}{3} + \frac{0}{0} \times \frac{0}{4}\right) = 0$$

$$RE(C_1^1, X_9) = \frac{1}{2}(\frac{0}{1} \times \frac{1}{2} + \frac{0}{0} \times \frac{0}{3} + \frac{0}{0} \times \frac{0}{4}) = 0$$

$$RE(C_1^1, X_8) = \frac{1}{2}\left(\frac{0}{0} \times \frac{0}{2} + \frac{0}{0} \times \frac{0}{3} + \frac{0}{0} \times \frac{0}{4}\right) = 0$$

In the similar way, concern this progression on cluster c_2.

$$RE(C_1^2, X_6) = 0.25$$
$$RE(C_1^2, X_7) = 0.25$$
$$RE(C_1^2, X_8) = 0.33$$
$$RE(C_1^2, X_9) = 1.58$$
$$RE(C_1^2, X_{10}) = 0$$

Table 3 Data points representing belonging clusters

Object	Cluster label
X_6	C_1
X_7	C_1
X_8	Outlier
X_9	C_2
X_{10}	Outlier

The table comprises objects and respective cluster label based on the above calculation of Rough Entropy (Table 3).

4.1 *Algorithm for Data Labeling Based Using Data Labeling*

Algorithm: Rough Entropy-based Data Labeling Algorithm (REDLA)

Input: Data set D with n data objects, window size N.

Output: No. of outliers.

Method:

Stage 1: Data set D will be separated into rise to estimate windows in light of given sliding window measure N, say those sliding windows are S1, S2 … Sn
Stage 2: Use any gathering measure on sliding window S1 to get introductory grouping result C1 with different packs. Data frameworks will come about Ct for t = 1, 2 … Where t is a time stamp.
Stage 3: where out = 0
Stage 4: For all unlabeled in grouping point Pj St + 1 window, begin with introductory t = 1 begin.
Stage 5: For each gathering cti begin
Stage 6: For each an Ai begin
Stage 7: Find the segment U/IND ({a}) use (1)
Stage 8: Find the repulsive entropy RE ({a}). (2)
Stage 9: end
Stage 10: Calculate frightful entropy (3)
Stage 11: Set edge to 0.5.
Stage 12: End Step
Stage 13: Consider the object of the next data points and move the objects into proper clusters
Stage 14: return out.

5 Experimental Results

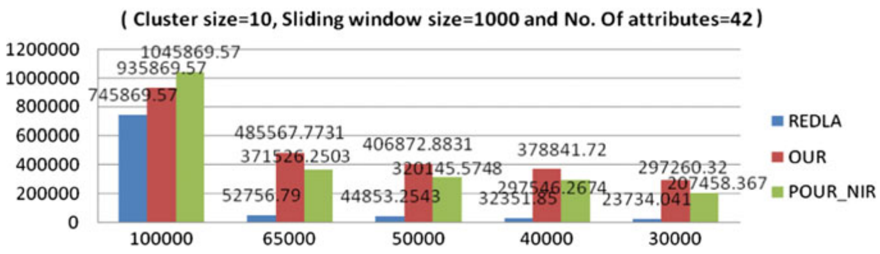
In this portion, we display the show of the anticipated create bunching all out information by a cautious exploratory investigation on the first Dataset. In Sect. 6, nature and the data set utilized are said. In next area manages the creating procedures of grouping comes about which are envision on the first data set.

$$\text{Cluster size} = 10,\ \text{Sliding window size} = 1000\,\text{and}\ \text{No. of attributes} = 42$$

The above explains various running times against cluster size, sliding window size, and no. of attributes. We took into consideration top three categorical data clustering algorithms, namely Our-NIR, Pour-NIR, and REDLA. In this approach, cluster size = 10, sliding window size = 1000, and no. of attributes = 42 against three algorithms running times were recorded. This helps us to understand the efficiency of REDLA better (Graph 1).

The Table 4 compares the running time of three different algorithms against three different parameters. The three parameters are cluster size, sliding window size, and no. of attributes. The value of parameters taken for analysis are cluster size = 10, sliding window size = 2000, and no. of attributes = 42. The running time is recorded for three different algorithms with variable data size, keeping the above-mentioned parameters as constant. From the table, the clustering algorithm REDLA is most efficient than OUR and Pour-NIR algorithms (Graph 2).

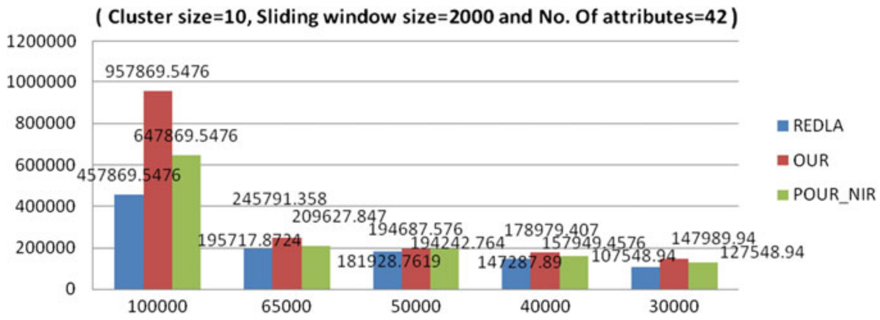
The Table 5 summarizes the running time of REDLA, OUR, and Pour-NIR clustering algorithms. The running time is evaluated for the above-mentioned algorithms against cluster size, sliding window size, and no. of attributes. The parameter values chosen for evaluation are cluster size = 10, sliding window size = 3000, and no. of attributes = 42. By observing the running time, we can



Graph 1 Cluster size = 10, sliding window size = 1000 and no. of attributes = 42

Table 4

.	.
.	.



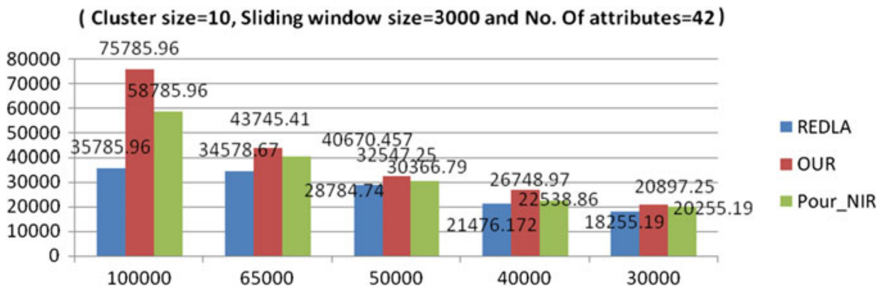
Graph 2 Cluster size = 10, sliding window size = 2000, and no. of attributes = 42

Table 5

.	.
.	.

conclude that REDLA is better in performance than OUR and Pour-NIR (Graph 3 and Table 6).

The Table 7 shows the results of REDLA and two other clustering algorithms, namely OUR and Pour-NIR, which are found for a data set. The best performance is evaluated in terms of running time. The running time is calculated by taking parameters, i.e., cluster size, sliding window size, and no. of attributes into consideration. The values chosen for parameters are cluster size = 5, sliding window size = 1000, and no. of attributes = 30. The performance of REDLA is statistically better than OUR and Pour-NIR algorithms (Graph 4).



Graph 3 Cluster size = 10, sliding window size = 3000, and no. of attributes = 42

Table 6 Cluster size = 10, sliding window size = 3000, and no. of attributes = 42

.	.
.	.

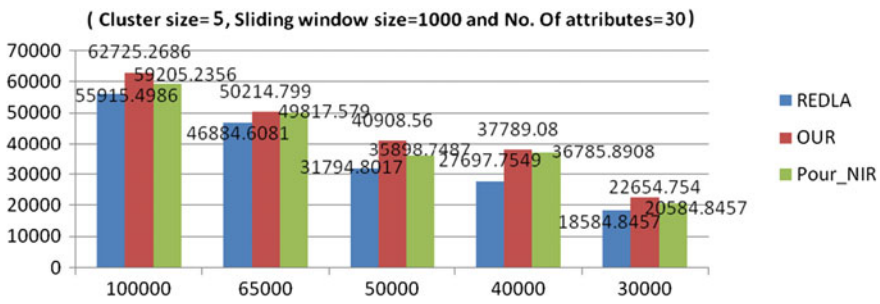
Graph 4 Cluster size = 10, sliding window size = 3000, and no. of attributes = 40

Table 7

.	.
.	.

The Table 8 reports the running time of three categorical clustering algorithms, namely OUR, Pour-NIR, and REDLA calculated against cluster size, sliding window size, and no. of attributes. The values chosen for parameters are cluster size = 5, sliding window size = 2000, and no. of attributes = 30. The result shows that REDLA algorithm ranks first in terms of running time when compared to OUR and Pour-NIR algorithms (Graph 5).

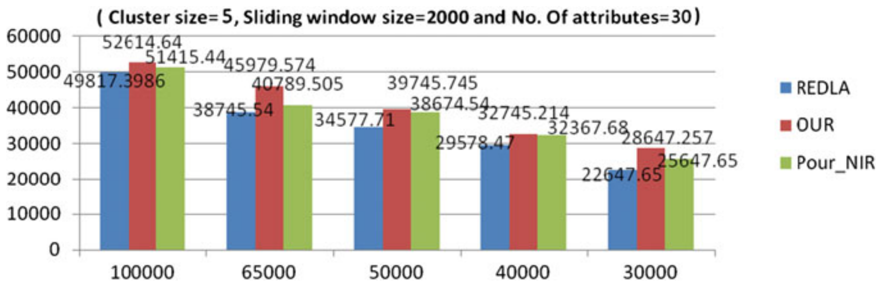
In this Table 9, we understand how the parameters, i.e., cluster size, sliding window size, and no. of attributes effect the running time for different categorical clustering algorithms. The values chosen for parameters are cluster size = 5, sliding window size = 3000, and no. of attributes = 30. As the data size decreases, the running time decreases for a data set. REDLA algorithm's running time is lower when compared to OUR and Pour-NIR. Hence, we can conclude that REDLA algorithm is the best categorical clustering algorithm (Graph 6).



Graph 5 Cluster size = 5, sliding window size = 1000, and no. of attributes = 30

Table 8 Cluster size = 5, sliding window size = 1000, and no. of attributes = 30

.	.
.	.



Graph 6 Cluster size = 5, sliding window size = 2000, and no. of attributes = 30

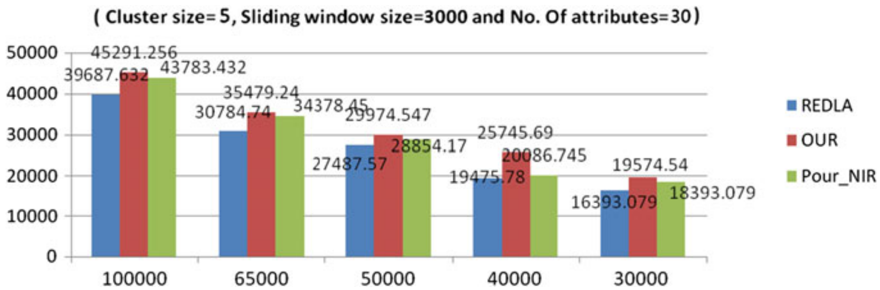
Table 9 Cluster size = 5, sliding window size = 2000, and no. of attributes = 30

Method name	Data size	C. S	S.W. S	Attribute size	Time
Cluster size = 5, sliding window size = 2000, and no. of attributes = 30					
REDLA	100,000	5	2000	30	49,817
OUR	100,000	5	2000	30	52,615
Pour-NIR	100,000	5	2000	30	51,415
REDLA	65,000	5	2000	30	38,746
OUR	65,000	5	2000	30	45,980
Pour-NIR	65,000	5	2000	30	40,790
REDLA	50,000	5	2000	30	34,578
OUR	50,000	5	2000	30	39,746
Pour-NIR	50,000	5	2000	30	38,675
REDLA	40,000	5	2000	30	29,578
OUR	40,000	5	2000	30	32,745
Pour-NIR	40,000	5	2000	30	32,368
REDLA	30,000	5	2000	30	22,648
OUR	30,000	5	2000	30	28,647
Pour-NIR	30,000	5	2000	30	25,648

Table 10 lists the data size, cluster size, sliding window size, no. of attributes, running time for three categorical clustering algorithms, namely REDLA, OUR, and Pour-NIR. The running time is calculated against three parameters, namely cluster size, sliding window size, and no. of attributes. Running time is calculated by choosing the values of parameters as cluster size = 3, sliding window size = 1000, and no. of attributes = 20. REDLA clustering algorithm exhibits lower running time than OUR and Pour-NIR. This shows that REDLA algorithm succeeds in proving that it has better efficiency than OUR and Pour-NIR algorithms (Graph 7 and Table 11).

Table 10 Cluster size = 5, sliding window size = 3000, and no. of attributes = 30

Method name	Data size	C.S	S.W.S	Attribute size	Time
Cluster size = 5, sliding window size = 3000, and no. of attributes = 30					
REDLA	100,000	5	3000	30	39,688
OUR	100,000	5	3000	30	45,291
Pour-NIR	100,000	5	3000	30	43,783
REDLA	65,000	5	3000	30	30,785
OUR	65,000	5	3000	30	35,479
Pour-NIR	65,000	5	3000	30	34,378
REDLA	50,000	5	3000	30	27,488
OUR	50,000	5	3000	30	29,975
Pour-NIR	50,000	5	3000	30	28,854
REDLA	40,000	5	3000	30	19,476
OUR	40,000	5	3000	30	25,746
Pour-NIR	40,000	5	3000	30	20,087
REDLA	30,000	5	3000	30	16,393
OUR	30,000	5	3000	30	19,575
Pour-NIR	30,000	5	3000	30	18,393



Graph 7 Cluster size = 5, sliding window size = 3000, and no. of attributes = 30

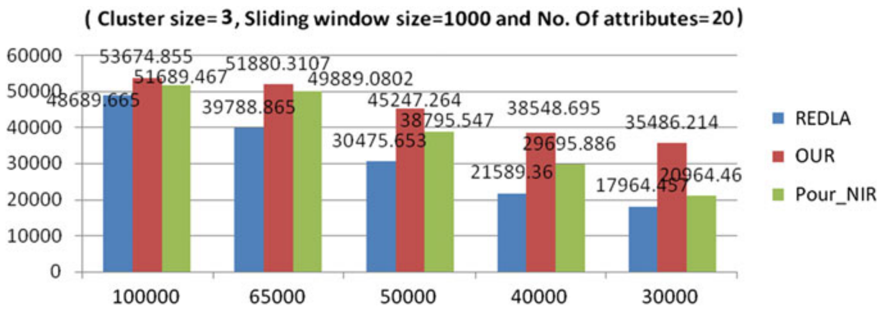
Table 12 lists the running time (milliseconds) to compare the performance of clustering algorithms, namely REDLA, OUR, Pour-NIR. The parameters taken into consideration for calculation of performance are cluster size, sliding window size, and no. of attributes. The values chosen for parameters are Cluster size = 3, Sliding (Graphs 8 and 9).

Table 11 Cluster size = 3, sliding window size = 1000, and no. of attributes = 20

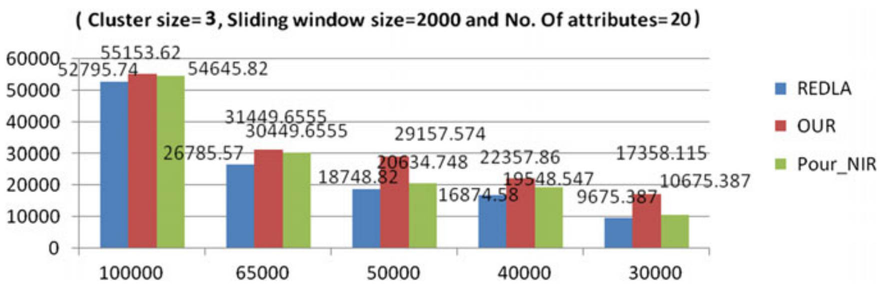
Method name	Data size	C.S	S.W.S	Attribute size	Time
Cluster size = 3, sliding window size = 1000, and no. of attributes = 20					
REDLA	100,000	3	1000	20	48,690
OUR	100,000	3	1000	20	53,675
Pour-NIR	100,000	3	1000	20	51,689
REDLA	65,000	3	1000	20	39,789
OUR	65,000	3	1000	20	51,880
Pour-NIR	65,000	3	1000	20	49,889
REDLA	50,000	3	1000	20	30,476
OUR	50,000	3	1000	20	45,247
Pour-NIR	50,000	3	1000	20	38,796
REDLA	40,000	3	1000	20	21,589
OUR	40,000	3	1000	20	38,549
Pour-NIR	40,000	3	1000	20	29,696
REDLA	30,000	3	1000	20	17,964
OUR	30,000	3	1000	20	35,486
Pour-NIR	30,000	3	1000	20	20,964

Table 12 Cluster size = 3, sliding window size = 2000 and no. of attributes = 20

Method name	Data size	C.S	S.W.S	Attribute size	Time
Cluster size = 3, sliding window size = 2000, and no. of attributes = 20					
REDLA	100,000	3	2000	20	52,796
OUR	100,000	3	2000	20	55,154
Pour-NIR	100,000	3	2000	20	54,646
REDLA	65,000	3	2000	20	26,786
OUR	65,000	3	2000	20	31,450
Pour-NIR	65,000	3	2000	20	30,450
REDLA	50,000	3	2000	20	18,749
OUR	50,000	3	2000	20	29,158
Pour-NIR	50,000	3	2000	20	20,635
REDLA	40,000	3	2000	20	16,875
OUR	40,000	3	2000	20	22,358
Pour-NIR	40,000	3	2000	20	19,549
REDLA	30,000	3	2000	20	9675.4
OUR	30,000	3	2000	20	17,358
Pour-NIR	30,000	3	2000	20	10,675



Graph 8 Cluster size = 3, sliding window size = 1000 and no. of attributes = 20



Graph 9 Cluster size = 3, sliding window size = 2000 and no. of attributes = 20

6 Environment and Data Set

The trials are done on a PC with an Intel Core i5 processor with 8 GB memory and the Windows 8.1 expert working framework. In the investigation, the K-modes [5] bunching calculation is done the underlying grouping and reclustering on the data sets. As the K-modes calculation is subject to the determination of starting group focuses, we misuse an instatement technique, which was proposed in [6], to acquire beginning bunch focuses before execution of the K-modes. For building up this paper, we utilized .NET dialect as front end and back end as MySQL. We have used an educational accumulation named US specification, where the data was assembled as an element of the 1990 assessment. There are 68 clear-cut traits. This informational index was gotten from the US Census 1990 crude informational collection. The new proposition is to distribute each unlabeled information point into the proportionate appropriate bunch with information marking moreover. It is much valuable to comprehend the statistic overview of general society. The proposed calculation in this way utilizes cases to inductively refine prior information, which has been utilized before to evaluate a few stream-gathering calculations and DCDAs, in our examination. Along these lines, this data set is time-advancing and is appropriate to inspect our calculations. We used the 10% subset rendition, which is given from the KDD document information for our analyses. In US Statistics data set, there are 24,58,285 records, and each record contains 68 fields (class name is incorporated, e.g., the span of the association, we acknowledge indistinguishable quantization on those numerical properties where each property is assembled into five straight-out qualities.

7 Evaluating Scalability

To test the adaptability of the wk-modes calculation, we utilize an engineered information generator [18] to produce data sets with various numbers of items and properties. The quantity of items fluctuates from 10,000 to 100,000, and the dimensionality is in the scope of 10–50. In all engineered data sets, each measurement has five diverse quality esteems as the distinctive bunching results will be gotten on the same data set when we select distinctive beginning group. Table 10 demonstrates the execution time of records by applying the bunching calculation, through Ming Chen strategy and proposed technique. Along these lines, each incentive in Table 10 is the normal of ten tests (Fig. 1).

This examination settled the dimensionality to 10, the bunch number to 3, and the information estimate shifts from 10,000 to 100,000. It can be seen that the proposed calculation is straight as for the information estimate. The execution time of the proposed strategy is particularly compelling than the K-modes calculation and Ming-Chen Method. In this manner, the wk-modes calculation can guarantee proficient execution when the information estimate is substantial (Table 13).

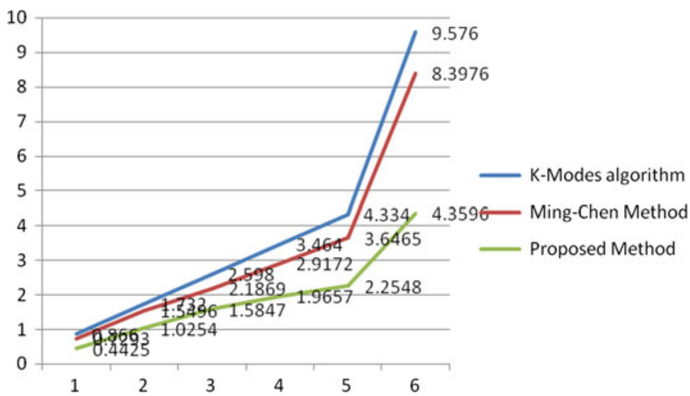


Fig. 1 Representation of efficiency (Time in milliseconds)

Table 13 Performance of various algorithms

Data records	K-modes algorithm	Ming-Chen method	Proposed method
10,000	0.866	0.7293	0.4425
20,000	1.732	1.5496	1.0254
30,000	2.598	2.1869	1.5847
40,000	3.464	2.9172	1.9657
50,000	4.334	3.6465	2.2548
100,000	9.576	8.3976	4.3596

8 Conclusion

In recent research papers in data mining/clustering, the problem of assigning the unlabeled data points into suitable clusters has not been fully explored in categorical domain. Besides, for the data which changes eventually, clustering not only decreases the excellence of clusters, but also disregards the potential of users. This paper fundamentally bargains the technique in light of Rough Entropy correspondence measure for apportioning the unlabeled information point into fitting group. Anomaly identification or grouping marking is done in light of variety in bunch similitude limit utilizing Rough Entropy. In future work, the idea float can be identified utilizing the above strategy whether it is happened or not.

References

1. Clustering Categorical Data Using Summaries, ACM SIGKDD (1999)
2. Jain, A.K., Dubes, R.C.: Algorithms for clustering data. Prentice-Hall International (1988)
3. Jain, A.K., Murthy, M.N., Flyn, P.J.: Data clustering: a review. ACM Comput. Surv. (1999)
4. Kaufman, L., Rousseuw, P.: Finding groups in data-an introduction to cluster analysis. In: Wiley Series in Probability and Mathematical Sciences (1990)
5. Guha, S., Rastogi, R., Shim, K.: CURE an efficient clustering algorithm for large databases. In: ACM SIGMOD International Conference on Management of Data, pp. 73–84 (1998)
6. http://archive.ics.uci.edu/ml/machine-learning-databases/molecular-biology/splice-junction-gene-sequences/splice.names
7. Sreenivasuluy, G., Viswanadha Raju, S., et al.: A threshold for clustering concept—drifting categorical data. In: IEEE 3rd International Conference on Machine Learning and Computing (ICMLC), vol. 3, pp. 383–387. IEEE, Feb 2011. ISBN: 978-1-4244-9253/11
8. Han, J., Kamber, M.: Data mining concepts and techniques. Morgan Kaufmann (2001)
9. Ganti, V., Gehrke, J., Ramakrishnan, R.: CACTUS. In: Vapnik, V.N, The Nature of Statistical Learning Theory. Springer (1995)
10. Gibson, D., Kleinberg, J.M., Raghavan, P.: Clustering categorical data an approach based on dynamical systems. VLDB, pp. 3–4, 222–236 (2000)
11. Sreenivasulu, G., Viswanadha Raju, S., et al.: Data labeling method based on rough entropy for categorical data clustering. In: International Conference on Electronics, Communication and Computational Engineering—ICECCE 2014, pp. 383–387 (2014). ISBN: 978-1-1170-1175/11/ IEEE

12. Li, X., Rao, F.: An rough entropy based approach to outlier detection. J. Comput. Inf. Syst. **8** (24), 10501–10508 (2012)
13. Parmer, D., Wu, T., Blackhurst, J.: MMR: an algorithm for clustering data using rough set theory. Data Knowl. Eng. **63**(3), 879–893 (2007)
14. Klinkenberg, R.: Using labeled and unlabeled data to learn drifting concepts. In: IJCAI-01Workshop on Learning from Temporal and Spatial Data, pp. 16–24 (2001)
15. Sreenivasulu, G., Viswanadha Raju, S., et al.: A review of clustering techniques. In: International Conference on Data Engineering and Communication Technology (ICDECT). Springer, March 2016. ISSN: 2250-3439
16. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical (1948)
17. Venkateswara Reddy, H., Viswanadha Raju, S.: A study in employing rough set based approach for clustering on categorical time-evolving data. IOSR J. Comput. Eng. (IOSRJCE) **3**(5), 44–51 (2012). ISSN: 2278-0661, https://doi.org/10.9790/0661-0354451
18. Andritsos, P., Tsaparas, P., Miller, R.J., Sevcik, K.C.: LIMBO: scalable clustering of categorical data. Extend. Database Technol. (EDBT) (2004)
19. Farnstrom, F., Lewis, J., Elkan, C.: Scalability for clustering algorithms revisited., ACM SIGKDD, pp. 51–57 (2000)
20. Barbara, D., Li, Y., Couto, J.: COOLCAT: an entropy-based algorithm for categorical clustering, ACM International
21. Sreenivasulu, G., Viswanadha Raju, S., et al.: A Comparative study of node importance in categorical clustering. Int. J. Adv. Eng. Global Technol. (IJAEGT) **1**(1), 784–788 (2013). ISSN: ISSN No:2309-4893 (print)|ISSN: 0975-397(online)
22. Bradley, P.S., Fayyad, U., Reina, C.: Scaling clustering algorithms to large databases. In: Fourth International Conference on Knowledge Discovery and Data Mining (1998)
23. Ghosh, J.: Scalable clustering methods for data mining. In: Ye, N. (ed.), Handbook of Data Mining, Chap. 10, pp. 247–277. Lawrence Ealbaum Assoc (2003)
24. Chen, H.L., Chuang, K.T., Chen, M.S.: On data labeling for clustering categorical data. IEEE Trans. Knowl. Data Eng. **20**(2011), 1458–1471 (2008)
25. Cao, Fuyuan, Liang, Jiye: A Data Labeling method for clustering categorical data. Elsevier Expert systems with applications **38**, 2381–2385 (2011)
26. Chen, H.L., Chuang, K.T., Chen, M.S.: Labeling un clustered categorical data into clusters based on the important attribute values. In: IEEE International Conference. Data Mining (ICDM) (2005)
27. Pawlak, W.: Rough sets. Int. J. Comput. Inf. Sci. **11**, 341–356 (1982)
28. Gluck, M.A., Corter, J.E.: Information uncertainty and the utility of categories. Cognit. Sci. Soc. 283–287 (1985)
29. Chen, C.-B., Wang, L.-Y.: Rough set-based clustering with refinement using Shannon's entropy theory. ELSEVIER Comput. Math. Appl. **52**, 1563–1576 (2006)
30. Liang, J.Y., Wang, J.H., Qian, Y.H.: A new measure of uncertainty based on knowledge granulation for rough sets. Inf. Sci. **179**(4), 458–470 (2009)
31. Jiang, F., Sui, Y.F., Cao, C.G.: A rough set approach to outlier detection. Int. J. Gen Syst **37** (5), 519–536 (2008)
32. Sreenivasulu, G., Viswanadha Raju, S.: A proficient approach for clustering of large categorical data cataloguing. In: International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). IEEE, March 2016. ISSN 978-1-4673-9939
33. Sreenivasulu, G., Venkateswara Reddy, H., Viswanadha Raju, S.: A threshold for clustering concept—drifting categorical data. IEEE Comput. Soc. ICMLC (2011)
34. Zhang, T., Ramakrishnan, R., Livny, M.: BIRCH: an efficient data clustering method for very large databases. In: ACM SIGMOD International Conference on Management of Data (1996)
35. Ng, R.T., Han, J.: CLARANS: a method for clustering objects for spatial data mining. Knowl. Data Eng., IEEE Transactions (2002)
36. Huang, Z., Ng, M.K.: A fuzzy k-modes algorithm for clustering categorical data. IEEE Fuzzy Syst. (1999)

37. Guha, S., Rastogi, R., Shim, K.: ROCK: a robust clustering algorithm for categorical attributes. In: International Conference On Data Eng. (ICDE) (1999)
38. Conf. Information and Knowledge Management (CIKM) (2002)
39. Sreenivasulu, G., Viswanadha Raju, S., et al.: Graph Based approach for clustering categorical data. Int. J. Adv. Comput. (IJAC) 117–125. ISBN: ISSN: 0975-7686

An Algorithm for Load Balancing in Computational Grid Using Activity-Based and CPU Usage Approach

Ramesh T. Prajapati, Tejas Kadiya, Bhavesh Jain and RajeshKumar

Abstract Grid computing is a collection of networked computers which work together to solve large problems. Grid becomes complex if resources are geographically spread over wide area, dynamic, and heterogeneous network. This makes load adjusting and adaptation to internal failure more critical in the event of processing network and determining which job requires resource to be assigned and how the load allocated among different machines to improve performance of grid computing. We introduce Dynamic Load Balancing algorithm which takes the resources to be assigned to the job according to the queue length, CPU usage, and free memory. In this research paper we reduced the execution time of application for completion of particular task. When system performing task in between system fail then machine of load transfer from one machine to another machine. We found that Dynamic Load Balancing algorithm is used for matrix users which aims to execute their program speedily using minimum execution time and minimize the idle time of processors.

Keywords Grid computing · Load balancing · Computational grid
Resource allocation

R. T. Prajapati (✉) · T. Kadiya · B. Jain · RajeshKumar
Indrashil Institute of Science and Technology, Rajpur, Kadi, Gujarat, India
e-mail: ramesh.9998590139@gmail.com

T. Kadiya
e-mail: kadiya.tejas@gmail.com

B. Jain
e-mail: bmjain007@gmail.com

RajeshKumar
e-mail: rajeshk542@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_38

1 Introduction

Grid Computing is defined as applying resources from thousands or many computers in a network to a single problem, usually one that requires to access large amounts of data. Network [1] is sort of parallel and circulated framework that empowers the sharing, determination, and collection of geologically conveyed assets progressively at run time contingent upon their accessibility, ability, and execution cost and client nature of administrations. Load adjusting is a system to upgrade assets, using parallelism, abusing throughput ad lib, and to slice reaction time through a proper dissemination of the applications [2]. Work movement is the main proficient approach to ensure that submitted employments are finished dependably and productively if there should be an occurrence of process disappointment, processor disappointment, hub crash, arrange disappointment, framework execution corruption, correspondence delay; expansion of new machines progressively despite the fact that an asset disappointment happens which changes the dispersed environment [3].

These are critical issues in Load Balancing: A surprising pinnacle can be steered to generally sit out of gear machines in the Grid. On the off chance that the grid is as of now completely used, the least need job being completed on the matrix can be incidentally postponed or even scratched off and carried out further next to prepare space for the big need job.

Lattice Computing ought to empower the work being referred to be kept running on a sit still machine somewhere else on the system [4]. Networks practically unite all inclusive conveyed PCs and data frameworks for making an all inclusive wellspring of figuring force and data [5]. The main feature of matrix is that service allocated between numbers of applications, and therefore, of utilizations, and along these lines, the measure of services accessible to any specific program exceptionally changes with extra duration. Load adjust is a strategy to upgrade assets, using correspondence, abusing flow and to decrease minimum duration through a proper dissemination of the program [6].

There are basically three important parameters which determine that which heap adjusting strategy will be employed:

- Who takes the decision for heap adjusting?
- What type of information is required for making the heap adjusting decision?
- Where the decision about heap adjusting is made?

2 Literature Survey

Load adjusting is a procedure to upgrade assets, using parallelism, misusing throughput act of spontaneity, and to slice reaction time through a suitable appropriation of the application [5].

Pal Nilsson and Michal Pioro have talked about Fair assignment that directing issue for correspondence network [7]. Hans Jorgen Bang, Torbjorn Ekman and David Gesbert have suggested opportunistic scheduling has recently attracted much attention as a means to increase the spectrum efficiency of wireless data networks. By allocating the common radio resource to the users that are currently best able to utilize it the inherent multiuser selection diversity is exploited [2]. Daphne Lopez and S. V. Kasmir raja have reported and contrasted Fair organize calculation and First Come First Serve (FCFS) and Round Robin (RR) schemes [8].

Stuff adjust is the most important problem in matrix computing [3, 9]. Grosu and Chronopoulos [10], Penmatsa and Chronopoulos [1] examined static matrix adjusting framework with masters and PCs where masters adjust stack between all PCs in a Round Robin design.

Qin Zheng, Chen-Khong Tham, and Bharadwaj Veeravalli directed characteristic of figuring out of which bunch a coming employment ought to be assigned and how heap is circulated between PCs in the gathering to improve the execution and furthermore suggest calculations which ensure searching a heap dissemination beyond PCs in a gathering that prompts the base reaction time [9]. M. Kamarunisha, S. Ranichandra, and T. K. P. Rajagopal discussed Load adjusting algorithm sorts, and three arrangements used are Information Strategy, Triggering Policy, and Selection Policy in grid environment [11, 12].

In [13], an expected load data planning calculation (ELISA) and Perfect Information Algorithm (PIA) are proposed. In PIA, when an occupation arrives, a processor registers the activity's complete time on all mate processors utilizing definite data about the present heap of a mate processor, its landing rate, and administration rate. The source processor chooses a pal processor with the base complete time and promptly moves an occupation on that mate processor, in the event that it can complete the activity sooner than this processor.

The decentralized load adjusting calculation is proposed in [14] for a grid domain. In spite of the fact that this work endeavors to incorporate the correspondence inertness between two hubs amid the activating procedure on their model, it did not consider the genuine cost for an occupation exchange. In [15, 16], a sender processor gathers status data about neighboring processors by imparting each heap adjusting minute. This can prompt regular message moves that outcome in huge correspondence overhead which is unwanted. Preemptive and non-preemptive process relocation methods are proposed in [17]. In this procedure, relocation happens proficiently by considering memory use and load on processor.

2.1 Load Balancing Categories

2.1.1 Static Algorithms

Static load adjusting calculations allot errands of a parallel program to terminals in light of either normal load of terminal group or the heap at the time hubs are doled out to some assignment. The benefit of calculation is effortlessness implies both execution and overhead, since there is no compelling reason to frequently screen the workstations for execution measurements.

2.1.2 Dynamic Algorithms

The dynamic load adjusting calculations take choice at run time. Allotment and reallocation of assets are designated at run time in view of need in framework environment which characterize when and which assignment to be moved [18].

2.2 Load Balancing Policies

Stack adjusting calculations characterized by their execution of the accompanying arrangements [19]:

- Information arrangement: determines what workload data to be gathered, when it is to be gathered and from where.
- Triggering arrangement: decides the proper period to begin a heap adjusting operation.
- Resource sort strategy: characterizes an asset as server or recipient of errands as indicated by its accessibility status.
- Location arrangement: utilizes the aftereffects of the asset sort strategy to locate an appropriate accomplice for a server or beneficiary.
- Selection arrangement: characterizes the assignments that ought to be relocated from overburden assets (source) to most sit out of gear assets (recipient).

The fundamental goal of load adjusting strategies is to accelerate the execution of applications on assets whose workload changes at run time in unusual way. Subsequently, it is huge to characterize measurements to gauge the asset workload. Each dynamic load adjusting technique must gauge the opportune workload data of each asset. This is entering data in a heap adjusting framework where reactions are given to taking after inquiries:

- How to quantify asset workload?
- What criteria are holding to characterize this workload?
- How to maintain a strategic distance from the negative impacts of assets dynamicity on the workload?
- How to consider the assets heterogeneity keeping in mind the end goal to acquire a quick normal workload illustrative of the framework?

A few load files have been proposed in the writing, similar to CPU line length, normal CPU line length, CPU use, and so on. The achievement of a heap adjusting calculation depends on strength of the amount of messages, reinforce condition, ease overhaul of the workload, and short mean response time which is a noteworthy estimation for a client. It is additionally fundamental to gauge the correspondence price actuated by a heap adjusting action.

2.3 Research Gaps

Based on Literature review, I found some research gaps. It takes amount of time that elapses between the job arrival time and the time at which the job is finally accepted by a user. Existing method used five policies so it creates overhead during balance from intensely stacked hub to daintily stacked hub. If failure occur during the execution of job at that time, it takes more time to execute.

3 Problem Statements

The nature of resources is dynamic and shared which affects the performance of application. There are two important functions: Management of resource and Workload which are provided at the usage level of the matrix. The main objective of load adjusting is to examine problems due to which Load Balancing technique is required in grid computing and also to transfer workload from heavy loaded nodes to lightly loaded nodes based on migration policy. This research evaluates the existing Load Balancing algorithm and finds out the performance of grid. There are five policies for implementation of Load Balancing algorithms [20].
The usage of Information Policy in current heap adjust calculation utilizes intermittent approach which is tedious.
The Triggering approach use Queue length which is not decided how to Load Balancing process is done.
Selection policy algorithm uses CPU job length as a parameter, which can be used for decision about selection of task for migration from heavily loaded node to lightly loaded node.

4 Proposed Algorithm for Grid Load Balancing

4.1 Load Balancing Algorithms

As indicated by the name, dynamic load adjusting calculations take choice at run time and utilize present or late load data when settling on dissemination choices. Job monitoring is responsible for detecting alert situations and information about the systems that would be helpful to initiate a migration. This information is reported to the re-scheduler, which evaluates if it is worth migrating the job, and in that case, decides a new allocation for the job.

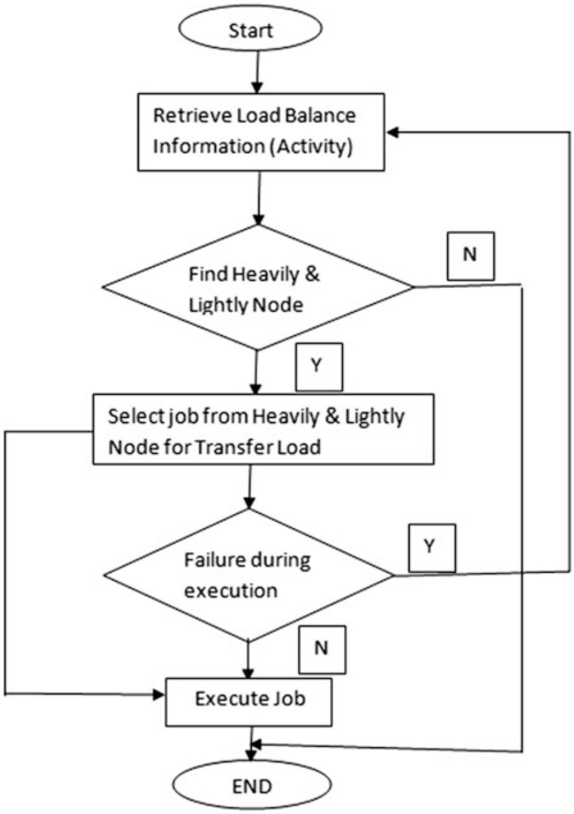


Fig. 1 Proposed system architecture for load balancing

Scheduler schedules the job to all processors for balancing load. Using some activities, we can change the load configuration in grid environment. We are using following activities.

In proposed stack adjusting calculation the exercises can be ordered as taking after arrival of any new employment and lining of that occupation to a specific hub, completion of execution of any occupation, arrival of any new asset, withdrawal of any current asset (Fig. 1).

Segment of code related to the Proposed Algorithm:

1. Construct a grid using Alchemi.Net
2. Create prime number application and submit the application to central node.
3. Check run time load on each resource.
4. Check if the nodes are heavily loaded or lightly loaded or failure node (CPU usage, free memory, and queue length).
5. If the load of particular node is heavily loaded or failure node, then share the load of node to lightly loaded node.

6. Analyze lightly loaded node and heavily loaded node and check the execution time of application.

In this algorithm, we have created grid and created one application of prime number allocated to central node. Central node allocated this application to number of machine in which it is connected to central node. Existing algorithm used five policies with period approach; queue length, threshold value, and proposed algorithm used three policies with activity approach, and CPU usage for balancing load. We checked how many nodes have heavy loaded and lightly loaded. If some nodes are lightly loaded, then we assign load and balance the load in prime number application.

5 Implementation

A Load Balancing implementation has been done which executes in grid environment; this application has been developed using VB.NET 3.5 and SQL Server 2008 R2 database server (Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11).

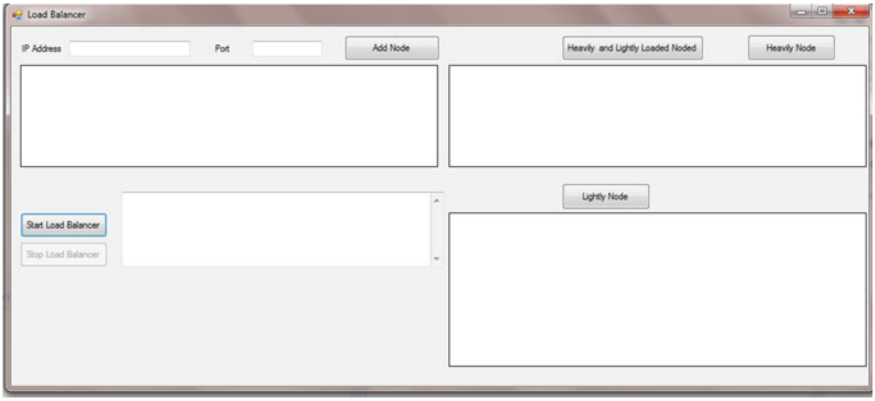


Fig. 2 Load balancer screenshot

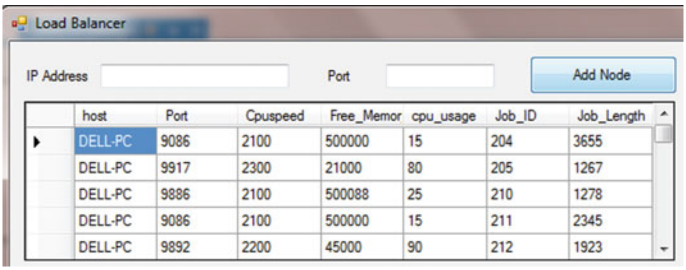


Fig. 3 Resource page with job

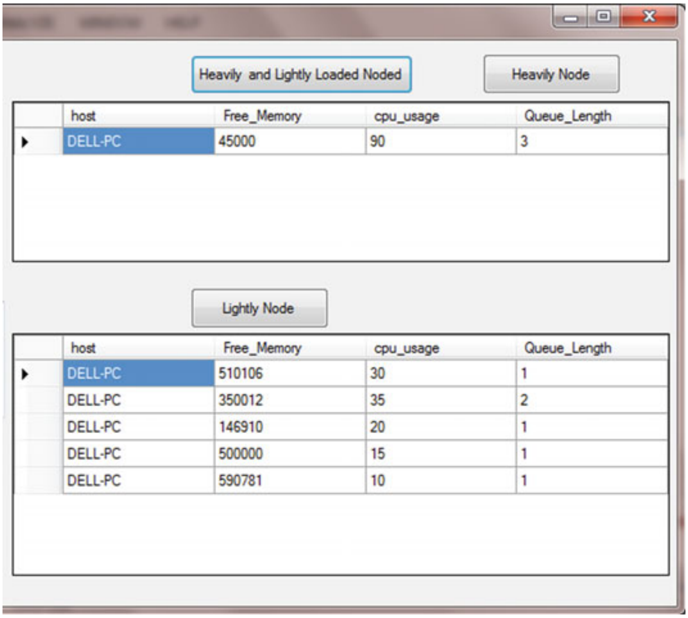


Fig. 4 Heavily loaded node and lightly loaded node

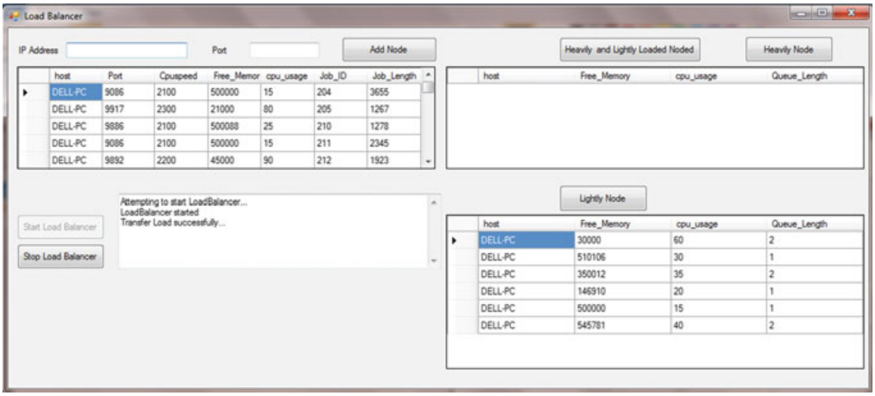


Fig. 5 Transfer load from fully stacked host to lightly stacked host

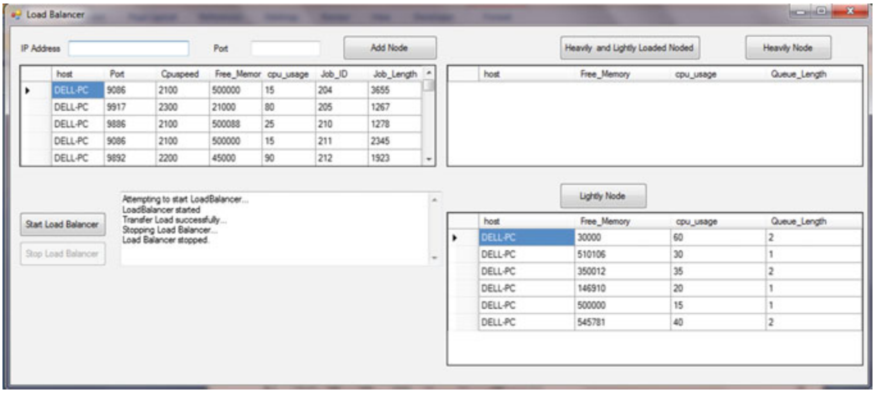


Fig. 6 Load balancer stopped successfully

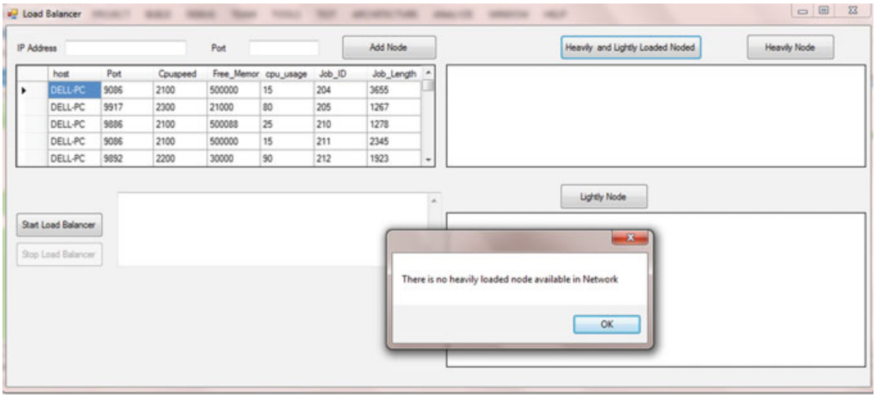


Fig. 7 No heavily loaded node available

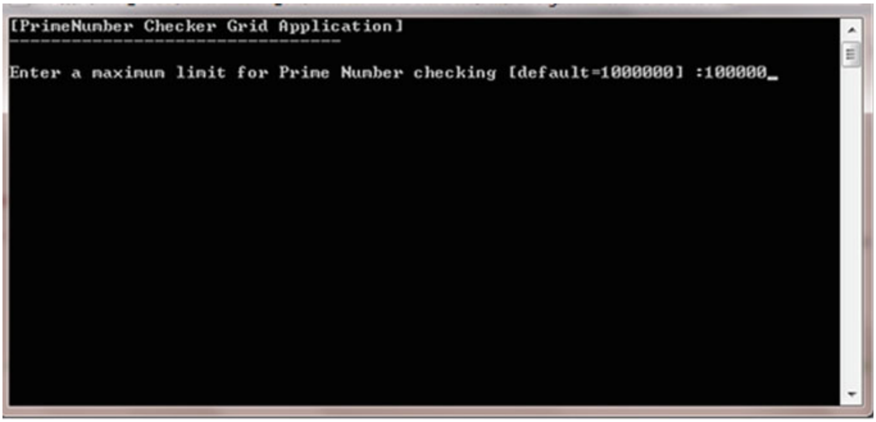


Fig. 8 Prime number generator application

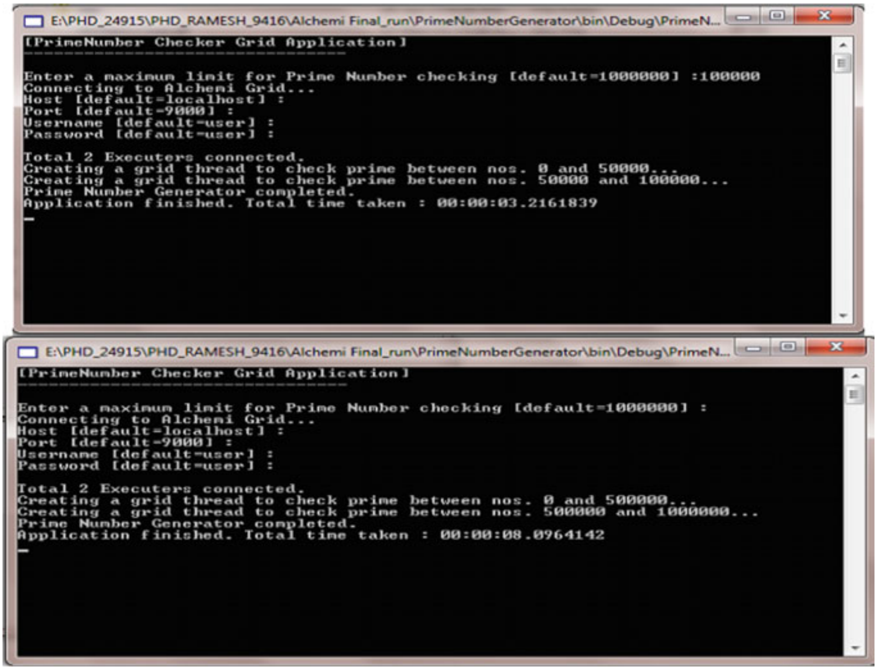


Fig. 9 Prime number generator new and existing application executions with two executors

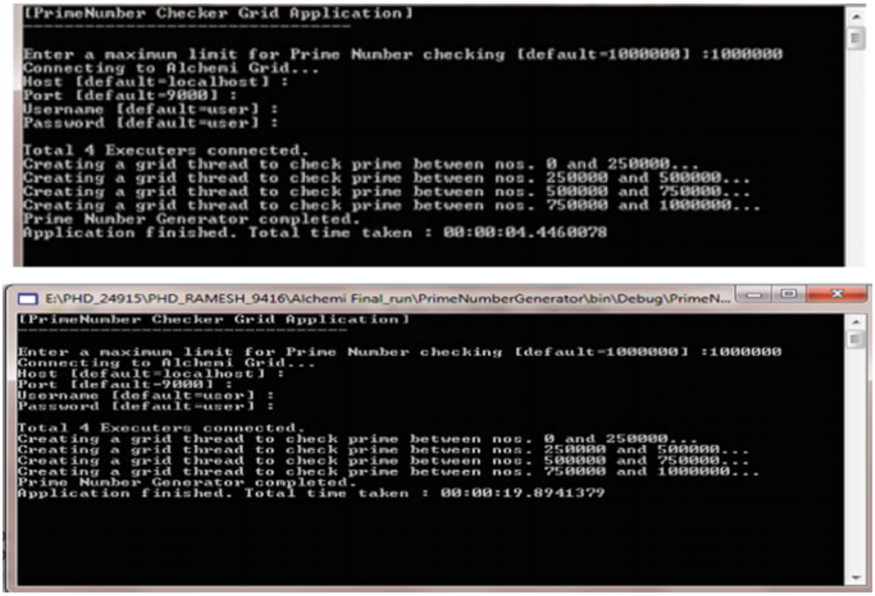


Fig. 10 Prime number generator new and existing application executions with four executors

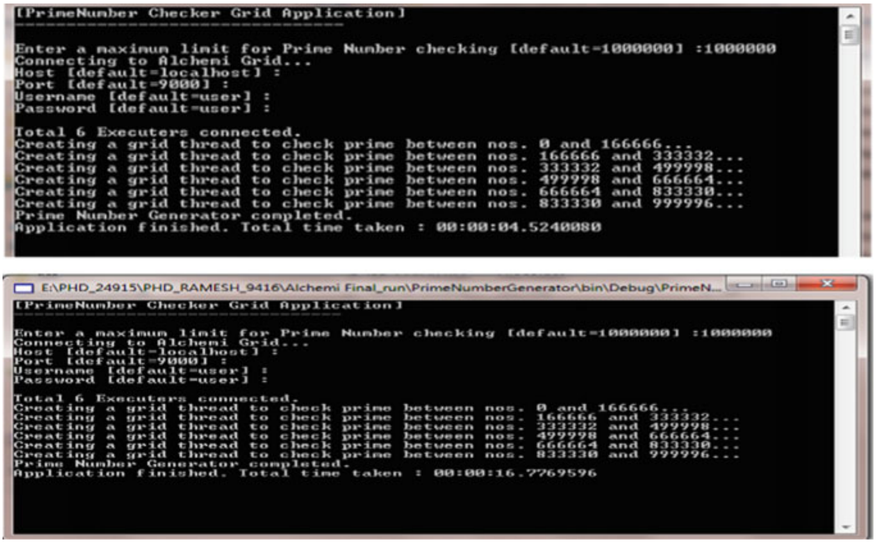


Fig. 11 Prime number generator new and existing application executions with eight executors

6 Comparison of Existing and Proposed Load Balancing Algorithm

In the following Table 1, we compared existing algorithm with three policy and parameter and proposed algorithm with three policy and parameter. If we consider find the prime number of 10,000,000 at that time we checked how many nodes are lightly loaded node and heavily loaded node. If some nodes are lightly loaded node, then we assign the prime number application to that node and execute application. Same thing is done after the failure of particular node; we assign load to other machine. Then, we checked the execution time of prime number application.

Total Execution Time: Differentiation of dissimilar no of executer with dissimilar input area with regard to time and parameters.

In Fig. 12, differentiation of Load Balancing algorithm for execution of application time is minimum. And Fig. 13 shows differentiation of Load Balancing after failure occurs during execution of application.

Table 1 Differentiation of dissimilar no of executer with dissimilar input area with regard to time and parameters

Input range to find out prime number	Total no of executer Running	Total time to complete execution	
		Existing method (in second)	Proposed method (in second)
		Periodic/job length	Activity/CPU usage
Prime of 10,000,000	2	06.1561	03.2161
	3	14.1274	07.9716
	4	19.2463	04.4460
	6	16.7769	04.5240
	8	13.8060	07.4601
	10	12.9067	06.4523
	15	15.3215	05.8967
Failure occurred during execution of prime number generator application			
Prime of 10,000,000	2	08.0964	03.2161
	4	12.5424	05.4444
	6	16.1616	06.7292
	8	12:1992	07.1616
	10	11:7612	08.4608
	15	15.4096	09.9961

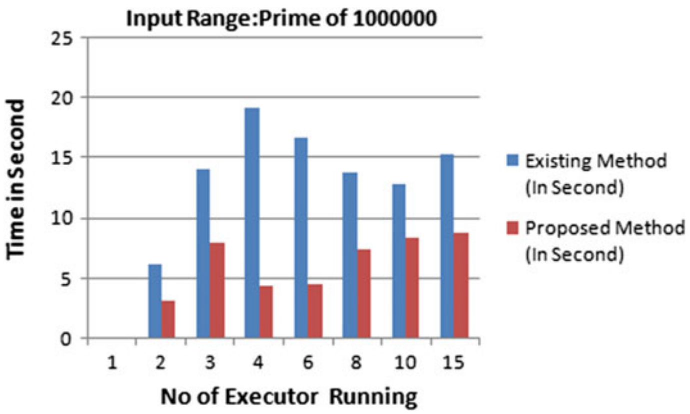


Fig. 12 Differentiation of no of executor running with regard to time

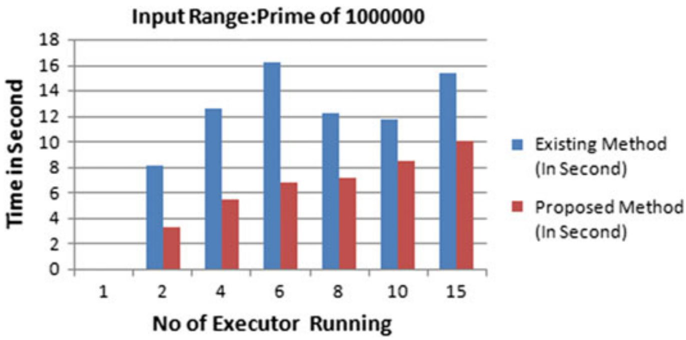


Fig. 13 Differentiation of no of executor running with regard to time after the failure of node

7 Conclusion and Future Enhancements

Grid Computing is unquestionably a promising propensity to understand high requesting applications and a wide range of issues. Goal of the grid condition is to accomplish superior registering by ideal use of topographically conveyed and heterogeneous assets. In this exploration, we broke down existing Load Balancing calculation and proposed an upgraded calculation which all the more effectively actualizes three out of five strategies executed in the existing load balancing calculation. With the help of proposed Load Balancing algorithm to minimize the execution time of application. In future, we will implement Load Balancing in the Internet of things and balance the load.

References

1. Penmatsa, S., Chronopoulos, A.T.: Job allocation schemes in computational grids based on cost optimization. In: Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium. Denver (2005)
2. Jorgen Bang, H., Ekman, T., Gesbert, D.: A channel predictive proportional fair scheduling algorithm
3. Zheng, Q., Tham, C.-K., Veeravalli, B.: Dynamic load balancing and pricing in grid computing with communication delay. J. Grid Comput. (2008)
4. Kaur, P., Singh, H.: Adaptive dynamic load balancing in grid computing an approach, [Ijesat]. Int. J. Eng. Sci. Adv. Technol. **2**(3), 625–632 (2012). ISSN: 2250–3676
5. Prajapati, R., Khanna, S.: A review paper on grid computing. IJEDR **1**(2), 67–72
6. IBM. Redbooks paper
7. Nilssonl, P., Pi´oro, M.: Unsplittable max-min demand allocation—a routing problem
8. Nguyen-Tuong, A.: Integrating fault-tolerance techniques in grid applications. In: PhD Thesis. University of Virginia, Aug 2000

9. Schamberger, S.: A shape optimizing load distribution heuristic for parallel adaptive FEM computations. Springer, Berlin, Heidelberg (2005)
10. Grosu, D., Chronopoulos, A.T.: Noncooperative load balancing in distributed systems. J. Parallel Distrib. Comput. **65**(9), 1022–1034 (2005)
11. Kamarunisha, M., Ranichandra, S., Rajagopal, T.K.P.: Recitation of load balancing algorithms in grid computing environment using policies and strategies an approach. Int. J. Sci. Eng. Res. **2**(3) (2011)
12. Srivastava, P.K., Gupta, S., Yadav, D.S.: Improving performance in load balancing problem on the grid computing system. Int. J. Comput. Appl. (0975–8887) **16**(1) (2011)
13. Wilson, L.F., Shen, W.: Experiments in load migration and dynamic load balancing in SPEEDS. In: Proceedings of the Winter Simulation Conference (1998)
14. Kanungo, P., Chandwani, M.: A process migration methodology for distributed computing environment. Ind. J. Comput. Technol. (2006)
15. Arora, M., Das, S.K., Biswas, R.: A decentralized scheduling and load balancing algorithm for heterogeneous grid environment. In: Proceedings International Conference of Parallel Processing Workshops (ICPPW '02)
16. Shan, H., Oliker, L., Biswas, R.: Job super scheduler architecture and performance in computational grid environments. In: Proceedings ACM/IEEE Conference of Super Computing, Nov 2003
17. Anand, L., Ghose, D., Mani, V.: ELISA: an estimated load information scheduling algorithm for distributed computing systems. Int. J. Comput. Math. Appl. (1999)
18. Saravanakumar, E., Prathima, G.: A novel load balancing algorithm for computational grid. Int. J. Comput. Intel. Tech. **1**(1), 20–26 (2010). ISSN: 0976–0466 & E-ISSN: 0976–0474
19. Smith, J.M.: Computer Science Department Columbia University New York, NY 10027 "A Survey of Process Migration Mechanisms". Technical Report CUCS 324-88
20. GRMS. http://www.gridworkflow.org/snips/gridworkflow/space/GRMS

ICT's Role in Building and Understanding Indian Telemedicine Environment: A Study

S. Syed Thouheed Ahmed, M. Sandhya and S. Shankar

Abstract The India is a home for 1.2 billion people and recognized as fastest growing economy on globe. In India, 68% (8.84 million) population contribute towards rural community. Information and communication technology (ICT) plays a vital role in development of telemedicine in India. In this study, a detailed analysis and observation reviews are conducted on the role of ICT in building telemedicine environment in India with major challenges and scenarios on implementation. Article also tracks major setbacks in storage systems, and need for the design of economic infrastructure and optimistic usage of available resources as India is a developing country.

Keywords Telemedicine · Big data · ICT · Telemedicine challenges
Space optimization

1 Introduction

India is marked among fastest growing economy and has made firm steps under urbanization enhancement. Indian government has also considered various measures to improvise current health conditions; Ministry of Health has declared schemes to add value for below poverty line (BLP) community. In India, there are 58,000 villages let to be recognized and reached. Today India has reached 100th rank in hungry countries with Bhutan and China still leading ahead. According to a report from DNA of Zee News, Indian economy state had its infrastructure condition cum management system are running under massive distress and major revolving in policies is needed. India has reached its peak to aid and treat its citizens. UNICEF study reports that India to be a slow progressive country under healthcare infrastructure development since two decades.

S. Syed Thouheed Ahmed (✉) · M. Sandhya · S. Shankar
School of Computer, Information and Mathematical Sciences,
BSA Crescent University, Chennai, India
e-mail: thouheed_cse_phd_17@bsauniv.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_39

In India, 4.1% of GDP has been drafted under national health policy (NHP) of 2011. However, actual spending was estimated to be 1.04% of allocated share, later 4.9% of GDP shared in 2005. In 2017, NHP has declared 2.5% of GDP under a time bond manner for development [1]. This inclination towards spending versus allocating has a huge gap to be covered. The report also targets to achieve a rise in mortality rate to 23 by 2025 and maternal mortality ratio (MMR) from current level to 100 by 2020 with improvising states towards infant mortality rate to 28 by 2019 [1]. From past observation study, the ratio of out-of-pocket (OOP) spending is resultantly increased with an influential damage towards poverty [2].

Rural health improvisation and maintenance have major challenging aspects such as infrastructure and basic amenities, lack of professionals, inaccessibility of location, lack of knowledge, orthodox preaching, planning and finance are few to mention. From various studies, no straightforward solution is proposed on these trivial methodologies. However, a hope of reformation is sighted with ICT for delivering a promising health care for rural community. ICT can resolve major setbacks on imbalanced infrastructure, remote professionals or experts and knowledge sharing with wide spacing new hopes in Indian Telemedicine system.

2 Telemedicine in India Today

Indian Telemedicine services (ITS) are blooming field for researcher and experts in exploring new and innovative approaches to resolve and tackle trivial methodologies in health care. In India, Telemedicine was primarily initiated by National Aeronautics and Space Administration (NASA) in 1975 with Application Technology Satellite-6 (ATS-6) to provide backbone support with S-band television, education, agriculture and health care [3]. In early 2000, government of India in association with Indian Space Research Organization (ISRO) and Department of Information and Technology (DIT) has marched progressive initiatives on Telemedicine infrastructure development [4].

In early 2002, on progressive steps with government and government-recognized institution towards telemedicine, many medical colleges and private healthcare hospitals became a part of executive and implementing team. In 2002, Narayana Hrudayalaya executed its first pilot project of cardiac surgery with a remote distance of 250 kms from Bengaluru in Karnataka [5]. On future isolation with NASA, Indian Telemedicine gained its shape under Dr. Ghanapati, a neurosurgeon and recounted as father of Indian Telemedicine with Apollo Hospitals. Today, Apollo Hospitals provide 125 plus centres on telemedicine under connection with peripheral or remote line rural healthcare units.

In developing countries, project and information score are vibrantly unaligned for any telemedicine initiative [6]. Since many initiatives of Telemedicine is carried out by private organizations, government-affiliated organizations or NGOs, thus standard reporting and charting of data are missed in scientific literature sites and archives. According to this report, a major factor influencing Telemedicine growth in developing countries when compared to developed countries is the allocation of GDP, and developing countries are still striking hard to increase the per capita income. In such a case, the resources' mobilization is a challenging risk for balancing National Health Service [6].

According to Combi et al. [6], a majority of Indian population stands out from the coverage region of aided healthcare units and urban cities. An estimated 770 million rural population is attended by inexperienced professionals with face-off under 80% of healthcare units in urban cities which is accessible by 30% of Indian population alone. In 2016, 65% of Indian population and major cities such as Bengaluru, Chennai, Mumbai, Madurai, Delhi, Hyderabad, Cochin and Calcutta are covered with Telemedicine healthcare units and expected to treat and consult 80% of population (overall) which in itself has created a major gap in resources and quality of treatment.

In Telemedicine report from World Health Organization (WHO) in 2010, spotlight was focused on inclusion criteria's such as participants, outcomes, interventions and language and publication dates [7]. In this report, it future discuss on reliability and benefits on shared telemedicine infrastructure in developing countries. The economic cost for patient on consultation is reduced by 7 times in metropolitan cities. WHO has discussed a model of social accountability as shown in Fig. 1. This clearly states that Telemedicine network development is possible on collaboration, partnership and knowledge sharing among academic institutions, government bodies, hospitals and policy-makers to retain balanced and user-friendly infrastructure.

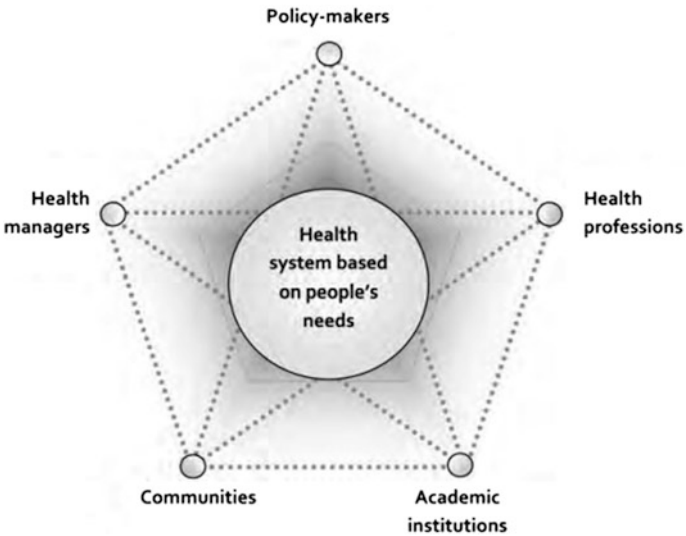


Fig. 1 Social accountability on telemedicine infrastructure balancing. *Source* WHO 2000

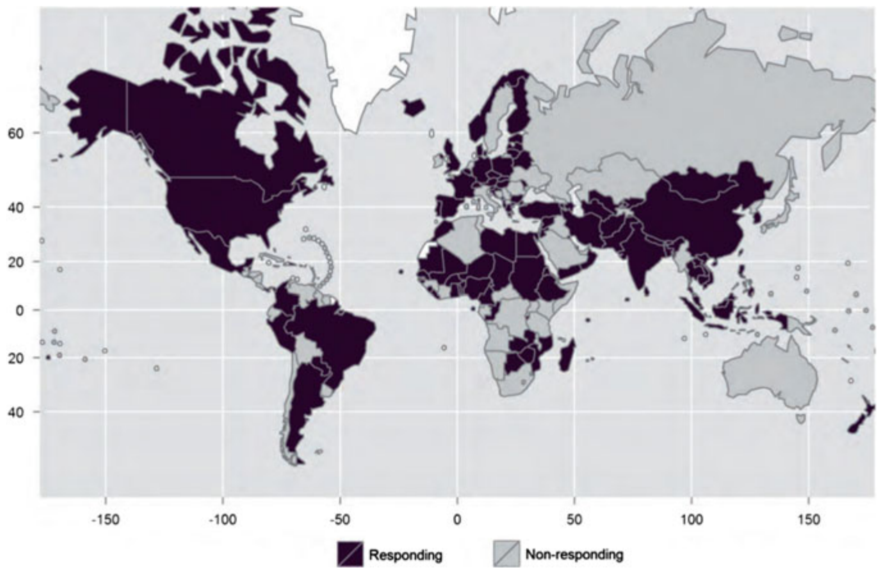


Fig. 2 Responding WHO member countries for funding under telemedicine. *Source* WHO Response Rate

WHO seed a resolution to adopt eHealth under Telemedicine infrastructure in 2005 with primary objectives to connect private–public under ICT development and health sectors under GOe policy, and its successive survey is conducted on 2009 [7]. In these policies, the objectives and trends were identified as mHealth (Mobile Health), Telemedicine, Management of Patients information, Legal front, policies on eHealth, eLearning (Knowledge Sharing), Foundations (NGO's or Third Party) and Country Profile. Based on the attributes extracted, the decision of selection and processing is projected with respect to World Bank funds and commerce as shown in Fig. 2.

3 Understanding Indian Telemedicine Infrastructure

India is a land of villages, culture, heritage and civilization since ancient times and has gained popularity worldwide with YOGA and Ayurveda in early twentieth century with its parallel proofs from modern science and hybrid treatments. As a fact, 20% of Indian population is inaccessible and declared tribal. These communities are influenced with ancient practices and thus creating awareness, and trust is a major challenge. Many surveys and literature focus on designing modern system for Telemedicine but a lack of its understanding on grounds makes the mathematically proven system fail in India.

In our study, Telemedicine is understood in two major fronts, Economy Front and Technology Front. The Indian population and its economy are in mixed culture. Today India is leader in manpower and shall retain its status under global recursion and paralleled fact; India is among largest unemployed populated country. In designing a eHealth infrastructure, one needs to understand both fronts of Indian society.

3.1 Economy Front

A large number of indexing and scientific literature with journals and publication articles are found discussing economy front on Telemedicine in India and world-wide. In this process, major challenging scenarios for understanding are identifies as policies updates, infrastructure design, political influence and language. Today in India, public–private sector collaborates such as Apollo Hospitals, Narayana Hrudayalaya and ISRO have designed policies and set-up models for establishing telemedicine base units, remote units and controlling units. Apollo Hospital's Aragonda's Telemedicine project is considered for major telemedicine sets as benchmark.

According to WHO's report [7], 25% of member states with active responses as shown in Fig. 2 have private National Telemedicine Policies and updated by DIT, India, has list of standards and policies for designing and setting up Telemedicine centres. The report continues as only 20% of these countries have started implementation on internal influential factors, and India in this has a mixed response with active research pending at scientific institutions.

3.2 Technology Front

Many primary health researchers and medical researchers are unaware on technological aspects and its influence on Indian Telemedicine. India is second largest contributor to be connected on social media such as Facebook, Twitter and LinkedIn. Reaching the population via dedicated telemedicine network is no more required. Indeed Telemedicine infrastructure can be designed on regular networking protocols such as LTE and 3G. According to Indian Telecom Survey, 4G (LTE) and 5G network developments are progressed and have covered 25% of India land area and followed by 3G with 58% and 2G or basic OFC networks of 90% and above. With such connectivity, the major challenges are identified and discussed in Table 1.

Table 1 Challenges identified in technological front

Challenges	Description
Storage systems	India is yet findings its place in storage system management, and 70% of Indian storage networks are outsourced and managed by international clients
Optimizing algorithms	Optimizing algorithms are a major challenge for soft transferring of large, unmodified and untampered biomedical files. Indian researchers and scientific and academic institutions are still in progress
Responsive interfaces	About 70% of Indian population is rural and in an average 45% of overall population is unfamiliar with English and computers. Hence, resolved solutions are required in graphical notations for easy understanding in minimal space and memory bandwidth
Low network coverage	Villages and downtowns are not assured with high-speed network connectivity as shown in paper. Designing an overall protocol of telemedicine to work, collect and transmit the files in such scenarios is challenging
Artificial Intelligence (AI)	Today, AI is most researched and expanded. Using such technologies under analysis for prioritizing the cases via telemedicine is required in India

4 Discussion

In developing country as India, delivering healthcare is a challenging task towards rural and inaccessible areas. Researchers are well contributing towards modernization of infrastructure [8]. Spotlights are now focused towards technological front in understanding the reasons for the unsuccessful implementation of telemedicine on ground. New challenges such as shown in Table 1 are recent updates in strengthen the infrastructure. Researchers are now focused to implement optimizing algorithms for faster downloading on host system under minimal bandwidth.

World Bank, WHO, ISRO and other private organizations such as Apollo Hospitals, Narayana Hrudayalaya are few to invest for the development of telemedicine skeleton in Indian economy. Telemedicine is already established in many developed countries and Indian administration should consider the real-time observations in deploying such successful models in India on considering the WHO's report on Telemedicine challenges in developing countries.

References

1. Home: http://www.thehindu.com/news/national/centre-cleared-the-long-awaited-national-health-policy-2017/article17487845.ece. Accessed 26 Oct 2017
2. Ghosh, S.: Catastrophic Payments and Impoverishment Due to Out of Pocket Health Spending, pp. 63–70 (2011)
3. Ferguson, E.W., et.al.: Survey of global telemedicine. J. Med. Syst. **19**(1) (1995)

4. Home: http://www.telemedindia.org/dit.html. Accessed 26 Oct 2017
5. Home: http://narayanahealth.org/telemedicine. Accessed 28 Oct 2017
6. Combi, C., Pozzani, G., Pozzi, G.: Telemedicine for developing countries a survey and some design issues. Appl. Cilincal Inform. **7**(4), 1025–1050 (2016)
7. Telemedicine Opportunities and developments in Member States, Report on the Second Global Survey on eHealth WHO (2010). ISSN 2220-5462
8. Patil, K.K., Ahmed, S.T.: Digital tele mammography services for rural India, software components and design protocol. In: International Conference on Advances in Electronics, Computers and Communications (ICAECC), pp. 1–5 (2014)
9. Home: ONC. http://onconet.nic.in. Accessed 26 Oct 2017
10. India—National Health Policy (2015)
11. International Telemedicine Disaster Medicine Conference, December 1991. NASA Publications NP-207, May 1993

Age Classification with LPCC Features Using SVM and ANN

Gaurav Aggarwal and Latika Singh

Abstract For humans, speech is one of the vital communication channel used for interchanging information, knowledge, and thoughts. Identifying the age of a person based on his/her speech is an essential part of speech therapy and many telecommunication applications. Many speech-related disorders can be diagnosed and cured using age identification at early ages. Depending on the age group, particular speech therapy can be given to a child. In this research, typical speech sentences were used to identify the age of 200 Indian children from the age group of 4–8 years. Linear predictive cepstral coefficients (LPCC) (formant frequencies) was applied to extract 128 acoustic features using sustained phonation, reading and imitation tasks. Artificial neural network (ANN) and support vector machine (SVM) were used to build two classification models. Comparisons were made on classification accuracy. Classification results were substantially higher between the age group of 4 and 8 years. This work will further be extended to gender classification with more robust features and algorithms.

Keywords Linear predictive cepstral coefficients · Formant frequency
Support vector machine · Artificial neural network

1 Introduction

The primary goal of this study is to identify the children's age and classify them into different age groups of 4–8 years using ordinary speech sentences uttered in Hindi. LPCC was employed to extract the acoustic features from the speech samples of the children. Early age identification using speech can help in detection of

G. Aggarwal (✉)
Manipal University Jaipur, VPO Dehmikalan, Ajmer Highway, Jaipur, India
e-mail: mtech.gaurav@gmail.com

L. Singh
The NorthCap University, Sector-23A, Gurugram, India
e-mail: latikasingh@ncuindia.edu

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_40

several speech-related neurodevelopment diseases. Other applications like identification for access control and surveillance, building user models using age estimation, human–computer interfaces and many recognition tasks are also comes under it. Effect of age over LPCC is also explained in detail. In previous studies, the numerical value of cepstral coefficients which are uttered by children from different countries decreases as the age progresses and also it has been observed that the formant frequencies of girls are higher than boys at the same age.

The research [1] worked on fundamental frequencies and first two cepstral coefficients frequencies F0 and F1 of speech of 84 subjects between 3 and 13 years of age. Their outcomes showed that the cepstral coefficients frequencies decreased with age. They also presented that, children with higher age group, the cepstral coefficients values for girls inclined to be higher than boys. It was investigated that at the same age the formant frequencies in prepubescent boys were about 10% lower than those for girls [2]. The study [3] showed the differences associated with age in cepstral coefficient values of speech articulated by preteen boys and girls. They concentrated only on first four cepstral values F0, F1, F2, and F3 of Australian English children and also reported that the formant frequencies decreased with increasing age and girl's formant frequencies were greater as compared to those in boys of similar age group.

A dataset comprises 435 children with age between 5 and 17 years and 56 adults were used by Lee et al. in 1999 [4] and investigated a significant decrement in cepstral coefficients with increasing age. Also, they reported that these frequency values were higher in females to those in males because male speakers have consistent axial growth of vocal tract.

The study [5] presented the classification of utterers in different age pool and shown that as the age progressed there was a consistent effect on format frequency values. Moreover, it also shows the similar decreasing values of formant frequency with age [6]. In the research [7], it was investigated that the speech production is directly dependent on personal characteristics of the speakers and linguistic information. The speaker characteristics consist of the physical properties of utterer vocal organs, including age range and gender of the speaker.

Most of the studies based on the English vowels that uttered by British children or other nationalities. Also, many of these researchers focussed on vowel instead of a complete sentence. So there is an urgent requirement to analyse the Hindi sentences uttered by Indian children and build a classifier which can correctly classify between different age group using LPCC. The primary aim of this study is to classify the age of the children using linear predictive cepstral coefficients from their simple speech sentences in their early age.

2 Methodology

2.1 Subject

In this study, 200 typically developed Indian children were considered in six different age groups: 4, 5, 6, 7, and 8 years. For the advanced research and investigation, we included the 15 adult speech samples in the dataset. Table 1 shows the speech sample database.

Before performing the recordings, approvals from the ethical committee of The NorthCap University and the concerned school were taken. Additionally, consent from the parent/guardians of the children who participated in the study and adults was taken. Children were identified from the primary school around Delhi/NCR. The gender ratio within the groups was similar. No one is suffering from any speech disorder, respiratory diseases, oral pathology, flu, allergies on any of the subject.

2.2 Recording Procedure

Speech samples were obtained from all the speakers divided into six groups, age from 4- to 8-year-old and adult. Children were asked to do three tasks: sustained phonation, reading, and imitating task for recording purpose. In the first task, children were instructed to utter Hindi and English vowels with a constant frequency. In the second task, they were invited to pronounce two phonetically rich nursery rhymes for approximately 2 min at a uniform pitch and loudness. In the third task, they were imitated the author's voice. This task was primarily done to examine the articulatory correctness and for 4-year-old children, as they are too small to pronounce rhymes. Children were well instructed about the recording task to avoid the hesitation.

A head-fitted SONY ICD-UX533F microphone with automatic gain control was used to record the speech with a sampling rate of 22 kHz 16-bit PCM, in a typical room environment. We used MATLAB signal processing toolbox for extracting 128 LPC formant frequencies' values. SVM and ANN were used to build classification models. The outcome was a mechanical stress during articulation with a

Table 1 Speech sample database

Age (years)	Participants	Male	Female
4 ± 0.3	40	13	17
5 ± 0.5	40	14	16
6 ± 0.4	40	13	17
7 ± 0.4	40	16	14
8 ± 0.3	40	15	15
21–28 (adult)	15	11	4

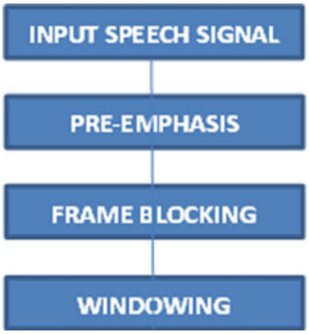


Fig. 1 Block diagram of speech signal preprocessing

natural vibration of the vocal chords. So there were some significant changes in LPC values in different age groups.

2.3 Speech Preprocessing

Speech samples were recorded in a soundproof room; still some unwanted noises like mike adjustments and whispering sounds were recorded with the signal. Goldwave noise reduction filter function was employed with 12-point FFT and 90% scaling to reduce the noise from the speech sample. For proper parameterization, the recorded sample was downsampled to 16 from 44 kHz [8]. For removing DC component, speech sample was pre-emphasized with the first-order high-pass filter shown in Eq. 1.

$$H(z) = 1 - \propto * z^{-1} \tag{1}$$

Lastly, Hamming window was used to divide the speech sample into frame segments. Speech preprocessing has been shown in Fig. 1.

3 Feature Extraction—Linear Predictive Cepstral Coefficients

LPCC is the representation of linear predictive coefficients (LPC) in the cepstral region. LPCC can directly be obtained from the LPC by applying recursion method rather than using inverse Fourier transform. The computational expenses are less in LPCC as it can be calculated without using Fourier transforms to alter signal from time domain to frequency domain. LPCC has all the advantages of LPC. LPC provides an ideal and approximate speech production model to the vocal tract spectral envelope. The basic idea of LPC, an assumed acoustic sample at time n, ŝ(n), can be calculated as the linear combination of past p acoustic samples, where p represent the order of the LPC [9]. Calculation for ŝ(n) is as shown in Eq. 2.

$$\hat{\mathrm{s}}(\mathrm{n}) = \sum_{n=1}^{p} a_m s(n-m) \tag{2}$$

The difference between the ideal and the calculated sample value gives rise to prediction error e(n), which is defined by Eq. 3.

$$\mathrm{e}(\mathrm{n}) = \mathrm{s}(\mathrm{n}) - \hat{\mathrm{s}}(\mathrm{n}) = \mathrm{s}(\mathrm{n}) - \sum_{n=1}^{p} a_m s(n-m) \tag{3}$$

where a_m is linear predictive coefficients.

It was computed by minimizing the mean squared error over a frame of an acoustic signal. Autocorrelation technique was applied to each frame of the windowed signal explained in Eq. 4.

$$(\mathrm{m}) = \sum_{n=0}^{N-1-m} x(n)x(n+m), m = 0, 1, .., p \tag{4}$$

where autocorrelation function is symmetric, and thus the equation for the linear predictive function is shown in Eq. 5.

$$\sum_{m=1}^{p} r(|m-k|)a_m = \mathrm{r}(\mathrm{m}), \mathrm{p} \geq \mathrm{m} \geq 1 \tag{5}$$

In this research, the number of cepstral coefficients used to depict each frame was computed by implementing Q $\approx \frac{3}{2}p$ with Q > p.

Using recursion method, LPCC is straight away obtained using LPC explained in Eqs. (6)–(8).

$$c_0 = ln\sigma^2 \tag{6}$$

$$c_m = a_m + \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} \quad 1 \leq \mathrm{m} \leq p, \tag{7}$$

$$c_m = \sum_{k=1}^{m-1} \frac{k}{m} c_k a_{m-k} \quad \mathrm{m} > p, \tag{8}$$

where σ^2 is Gain of LPC model, c_m is LPCC, a_m is LPC and P is order.. The block diagram representation of LPCC in shown in (Fig. 2).

In this study, 128 LPCC speech features were extracted to distinguish between the children's speech with different age groups. For better analysis, various parameters of LPCC like frame length and value of α were varied with

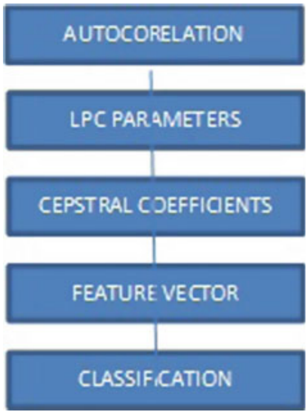


Fig. 2 LPCC feature extraction using speech

non-overlapping window. The result was calculated from the LPCC features with frame length = 20 ms, $\alpha = 0.97$, order = 12 with 256-point FFT. The outcomes of the study were conferred in the Result section.

4 Classification Algorithms

Classification between ages of children from 4 to 8 years and adult was completed using SVM and ANN techniques, previously employed [10]. The primary goal of the algorithms is to frame a model intelligent enough to train itself from extracted cepstral coefficients, build a classifier, and apply it for correct classification between children with different age groups. In this research, classifier contains six classes (4–8 years and adult) for the training phase. Tenfold cross-validation is used for testing and validating the classifier. Cepstral features are used to classify between children ages.

4.1 Support Vector Machine

The support vector machine is a useful and convenient statistic machine learning algorithm that has been applied in the area of pattern recognition [11, 12]. SVM is used for classification of both linear and nonlinear data. A nonlinear classifier will be used when the data is nonlinearly separable and linearly non-separable. It uses a nonlinear mapping to translate the training data into a higher dimension using a nonlinear transformation ϕ and then apply a linear separation.

To build a nonlinear SVM classifier, the intermediate product (X, Y) is substituted by a kernel function K(X, Y):

$$\mathrm{f(x)} = \mathrm{sign}(\sum_{i=1}^{L} \alpha_i y_i K(x_i, x) + \mathrm{b}) \tag{9}$$

SVM consists of two layers. In the learning phase, the first layer opts for the base $K(x_i, x)$ where I = 1, 2, …, N, from the bases defined by the kernel. The second layer builds linear function in the feature space. It is same to construct the optimal hyperplane in the analogous feature space. Hyperplane with the larger margin considered to be more accurate than the smaller margin for classifying feature data. The distance between the centre and the boundaries of the hyperplane is interpreted as the maximum margins on both sides. Equation 10 defined the separating hyperplane.

$$W \cdot X + b = 0 \tag{10}$$

where W is the weight vector W = $(w_1, w_2, \ldots., w_n)$, n is the number of attributes, and b is the scalar.

4.2 *Artificial Neural Network*

In artificial neural network (ANN), the thinking process of a human brain can be stimulated using an interconnected structure of artificial neurons. ANN is trained to take intelligent decisions depending upon the input and output information base. Thus, there is no requirement for a specific algorithm for age classification. An ANN using Radial Base Function (RBF) Network is used to derive probabilistic neural network (PNN). In [13], it is stated that the PNN is chosen over RBF due to its many times faster speed. It has a simple structure and robust in noise. An ANN is composed of processing elements called or perceptron, designed in different ways to form the network's structure. Every perceptron takes inputs, processes inputs, and sends a single output. ANN uses a sigmoid function. An activation function is shown in Eq. 11.

$$\mathrm{f(x)} = \frac{1}{1 + e^x} \tag{11}$$

In this paper, SVM and ANN were used for age classification between age 4–8 years and adult.

5 Results

The 128 cepstral coefficients of all the children of age group 4–8 year and 15 adults were extracted using MATLAB. For LPCC feature extraction, various combinations of frame size, α, non-overlapping window were taken into account. The experiment was performed with 20 ms window size, 0.97α, and non-overlapping Hamming windows. This combination generated 128 LPCC features for good classification accuracy. Figure 3 shows the LPCC results for one child from each group. Each figure consists of four plots which show the complete information about the cepstral and the error value.

According to the results, cepstral coefficient values of Indian children decreased as the age progresses. However, there were some exceptions at different coefficients in different age groups. The difference in length of the vocal tract was the only reason for the differences between their cepstral frequencies. This result was in sync with the previous researches done [3, 14, 15] in other languages.

The classification accuracy of both the classifier SVM and ANN with different age group of children and adults was discussed in Table 2. In the context of accuracy, ANN outperformed SVM.

Although an extraordinary difference between age 4 and 5 years could not be found, the usual trend is still similar with earlier researches. On clubbing the age

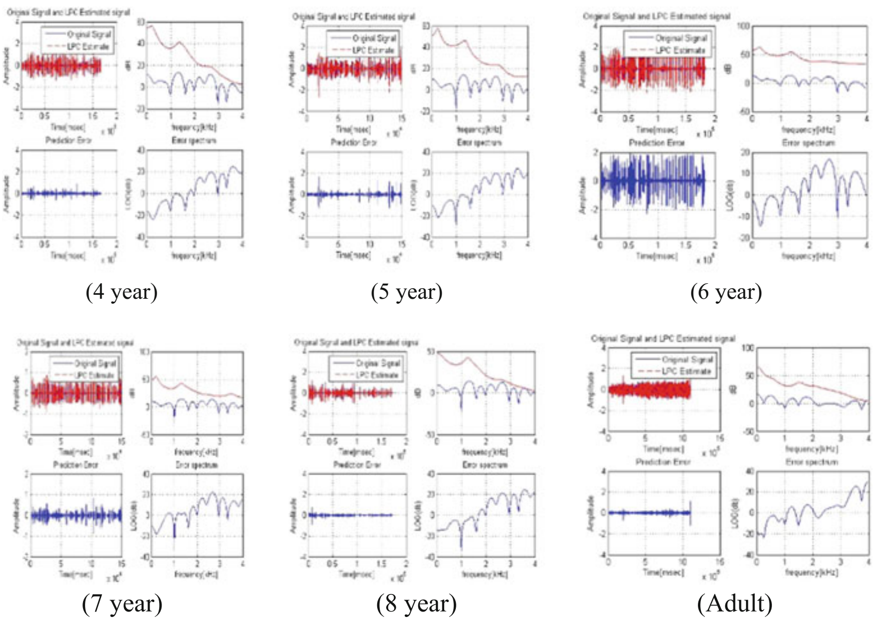


Fig. 3 LPCC of the speech signals of one child from each age group from 4 to 8 year and adult with four plots which show the complete information about the cepstral and the error value

Table 2 Classification results using ANN and SVM. (1) Between all the children groups (2) and (3) different combinations of children's groups (4) adults versus children from all years

S. no.	Classification (age group)	No. of participants	Accuracy (%)	
			ANN	SVM
1	4 year	40	68.0	66.3
	5 year	40		
	6 year	40		
	7 year	40		
	8 year	40		
2	4 + 5 years (combined as 4 year)	80	82.1	77.3
	6 + 7 years (combined as 7 year)	80		
	8 year	40		
3	4 + 5 years (combined as 4 year)	80	89.5	85.2
	6 year	40		
	7 + 8 years (combined as 8 year)	80		
4	Adult	15	94.2	87.3
	Children (8 children from each group)	40		

groups, we found remarkable results with SVM and ANN. Significant classification accuracy was achieved between adult and children with all the age groups. Accuracy reached 94.2 using ANN and 87.3 using SVM classifier.

6 Conclusion and Discussion

The objective of the paper is to develop an automated classifier which can distinguish between the ages using acoustic parameters. Linear predictive cepstral coefficients are used to categorize voice of different age groups which can help in many language development applications. The research can also be used in gender recognition technique by incorporating more acoustic features with higher accuracy. Using LPCC, a significant difference has been noticed between a speech of a child and an adult. A remarkable difference has also been noticed in the acoustic features of children with age group of 4 and 8 years. No remarkable difference was investigated between 4- and 5-year age group because of the similar articulatory structure. As this work is our first step to classify age using acoustic features, we are optimistic to improve it to another level of speech analysis.

References

1. Eguchi, S., Hirsh, I.J.: Development of speech sounds in children. Acta Oto-Laryngol. Suppl. 257, 1–51 (1968)
2. Bennett, S.: Vowel formant frequency characteristics of preadolescent males and females. J. Acoust. Soc. Am. **69**, 231–238 (1981)
3. Busby, P.A., Plant, G.L.: Formant frequency values of vowels produced by preadolescent boys and grils. J. Acoust. Soc. Am. **97**(4), 2603–2606 (1995)
4. Lee, S., Potamianos, A., Narayanan, S.: Acoustics of children's speech: developmental changes of temporal and spectral parameters. J. Acoust. Soc. Am. **105**, 1455–1468 (1999)
5. Xue, W.R., Hao, G.J.: Changes in the human vocal tract due to aging and acoustic correlates of speech production: a pilot study. J. Speech Lang. Hear. Res. **46**, 689–701 (2003)
6. Harrington, J., Palethorpe, S., Watson, C.I.: Age-related changes in fundamental frequency and formants: a longitudinal study of four speakers. In: INTERSPEECH-2007, pp. 2753–2756 (2007)
7. Reubold, U., Harrimgton, J., Kleber, F.: Vocal aging effect on F0 and the first formant: a longitudinal analysis in adult speakers. Speech Commun. **52**, 638–651 (2010)
8. Huang, X., Acero, A., Hon, H.: Spoken Language Processing: A Guide to Theory, Algorithm, and System Development. Prentice Hall PTR, Upper Saddle River, NJ, USA (2001)
9. Rabiner, L., Juang, B.: Fundamentals of Speech Recognition. Prentice Hall (1993)
10. Deng, L., Xiao, L.: Machine learning paradigms for speech recognition: an overview. IEEE Trans. Audio Speech Lang. Process. **21**(5) (2013)
11. Jiang, H., Bai, J., Zhang, S., Xu, B.: SVM-based audio scene classification. In: Natural Language Processing and Knowledge Engineering, IEEE NLP-KE'05, pp. 131–136 (2005)
12. Guo, G., Li, S.Z.: Content-based audio classification and retrieval by support vector machines. IEEE Trans. Neural Netw. **14**(1), 209–215 (2003)
13. Master, T.: Practical Neural Network Recipes. Wiley, New York (1993)
14. White, P.: Formant Frequency Analysis of Children's Spoken and Sung Vowels Using Sweeping Fundamental Frequency Production (1999)
15. Hillenbrand, J., Getty, L.A., Clark, M.J., Wheeler, K.: Acoustic characteristics of American English Vowel. J. Acoust. Soc. Am. **97**(5), 3099–3111 (1995)

Analyzing Challenges to Transportation for Successful Sustainable Food Supply Chain Management Implementation in Indian Dairy Industry

Yogesh Kumar Sharma, Sachin Kumar Mangla and Pravin P. Patil

Abstract Food Corporation of India (FCI) and Comptroller and Auditor General of India (CAG) clearly mentioned in their reports that transportation is a critical issue nowadays in Indian dairy industries. Increased population in India forced the companies to add sustainability in their food processing and manufacturing to meet the increasing demands. Thus, sustainable food supply chain management (SFSCM) helps to reduce the wastage of food throughout the food supply chain. For successful implementation of SFSCM, the present work is mainly focused on challenges in transportation. The challenges were analyzed by new method, i.e., best worst method (BWM). This method is generally used for solving (MCDM) multi-criteria decision-making problems. In MCDM problem, the best alternative is selected by comparing number of alternatives with respect to selected criteria. In BWM, the policy maker identifies the best (very important) and the worst (less important) criteria. Eight challenges were identified from the literature review and experts' opinion. The important factors of the BWM method include less comparison data and more reliable comparisons.

Keywords SFSCM · BWM · MCDM · Transportation · Dairy industry
Food wastage · Environment

Y. K. Sharma (✉) · P. P. Patil
Department of Mechanical Engineering, Graphic Era (Deemed To be University), Dehradun, Uttarakhand, India
e-mail: yogeshsharmame355@gmail.com

P. P. Patil
e-mail: deanresearch4geu@gmail.com

S. K. Mangla
Knowledge Management and Business Decision Making, Plymouth Business School (PBS), University of Plymouth, Devon, Drake Circus, Plymouth PL4 8AA, United Kingdom
e-mail: sachinmangl@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_41

1 Introduction

A supply chain consists of organizations, people, resources, and information involved in moving a product or service from supplier to customer. But the increasing demand for food globally also increases the challenges for the supply chain to make them sustainable [1]. Transportation, communication technology, and advances in production are the key drivers for the dairy development and enhance the dairy supply chain. In many developing countries like India, dairy growth is constrained by transportation, refrigeration, promotion, processing, and dietary issues [2]. To overcome these issues, sustainable food supply chain management (SFSCM) in dairy industries plays an important role. SFSCM is defined as "The integration of sustainability in food supply chain at each and every step during manufacturing, transportation, distribution and collection" [3]. Transportation is one of the most important challenges in SFSCM implementation in Indian dairy industries [4]. The shortage of road and rail network, and lack of financial support from government has been the root of problems in front of food sector [5, 6]. Companies dealing with consumer packaged goods (CPG) and fast-moving consumer goods (FMCG) also face the problem of transportation as to travel a product from one place to another. Truck and driver's availability, environmental factors, and miscommunication are the issues for the cancelation of product supply, and these factors also increase the level of cancelation risks every day [7, 8]. In the present study, a novel approach called best worst method (BWM) is projected to solve (MCDM) problem where a number of alternatives are evaluated with respect to a number of criteria in order to select the best alternative(s). In BWM, the best (e.g., most desirable, most important) and the worst (e.g., least desirable, least important) criteria are recognized first by the experts. Pair-wise comparisons are then conducted between each of these two criteria (best and worst) and the other criteria [9–12]. The remaining research is as follows: Section 2 presents literature review related to transportation challenges. Sections 3 and 4 explain about the objective and research gap in the present study. The solution methodology is specified in Sect. 5. Discussions of results and managerial implications are discussed in Sect. 6. Finally in Sect. 7 conclusions and limitations of the work are specified.

2 Literature Review

The literature review mainly focuses on the transportation challenges for the successful implementation SFSCM in the Indian dairy industry.

2.1 Transportation in SFSCM

Usually transportation and logistics problems have been expressed through operations research, mathematical modeling, and simulation. But these approaches always face criticism about their success to deal with real-life problems [13]. Transportation has many challenges, and these challenges were identified on the basis of the literature review and decision makers.

Challenges in Transportation to Implement SFSCM in Indian Dairy Industry
Transportation in SFSCM is a big issue nowadays, so mostly companies are giving more stress on the challenges to minimize its effect. These challenges are mentioned below.

Environment Pollution and Environmental Factors (CH1)
Due to the environmental factors such as floods, heavy rains in the hilly areas will affect the transportation system. As a result, environmental problems are correlated with transportation sector [14, 15]. In addition, environment pollution and environmental factors are hurdles in implementation of sustainable transportation of food products.

Trucks and Drivers Availability (CH2)
Trucks and drivers availability when they required in the firm is a big issue in transportation mainly in dairy/food industry. Usually, they go on strike and there are some technical problems in the trucks [7]. Skilled truck drivers and well-equipped trucks are important for sustainable practice of transportation. Turn off the refrigerators when they are not required, to save the fuel costs [6].

Network Structure (CH3)
In transportation, network structure is an issue because of route patterns that differed every day. Scheduler duty is to find the best route for delivery. For this type of complications, scheduler takes help from computational software [16]. It is clearly mentioned in the past studies that network obstruction is a big issue in transportation [17]. For getting higher efficiency in the food supply chain, larger vehicles are required to transport heavier loads.

Loading and Unloading Practice (CH4)
Loading and unloading is very straight and simple process in trucks, and it will happen thousands of time every day throughout the country [18, 19]. Loading and unloading of food products plays an important role for the economic growth of the company, but still it is a challenge in front of companies during transportation. SFSCM technique is useful to reduce all the barriers in loading and unloading in a sustainable manner.

Poor Pest Control (CH5)
In food safety, pest control plays an important role. Insects like flies and cockroaches are responsible to spread the food-borne disease by contaminating food at any point of manufacturing. Stored product insects can damage and contaminate food during transport and storage [20]. Monitoring, detection, and investing in pest control can help the firms to stop the entering of insects in food during processing, and it is useful for food safety also. Food production and preparation areas must be clean and free from pest [21].

Maintenance for Transportation Units (CH6)
Transportation vehicles, accessories, and connections in dairy industry should be kept clean and free from dust, remains, and any other substance or smell that may infect the food product. Cleaning and sanitation process should be specified in writing. For different types of food products that are transported, it may be required to have different cleaning and sanitation procedures [22].

Good Communication Between Supplier, Transporter, and Receiver (CH7)
Communication between shipper, transporter, and receiver in dairy industry must be good, and it is essential for the betterment of the industry economic growth. They are responsible for food safety in food supply chain at each and every step. Good communication will help a supplier build a rapport line that could see them getting more work from a buyer and in turn provide them a steady stream of revenue [6, 23].

Government Certification (CH8)
The food and drug administration (FDA) passes a law FSMA (Food Safety Modernization Act) which is based on the safely transportation of human and animal food, which is effort made by FDA for protecting the food from any contamination and keeping the food safe during transportation. This act is one out of seven foundational rules which had been proposed since January 2013 for ensuring food safety. The main aim of this rule is to reduce practices during transportation which creates risks of food safety like cleaning of vehicles between loads, refrigeration of food, and improper ways of handling. Thus, the government must ensure that food should be transported in a safe manner [24].

3 Objective of the Study

Transportation is a big issue in Indian dairy industry from the past few decades. The best and worst challenge in transportation is identified by BWM method. BWM method gives very promising results as compared to other methods for MCDM problem. In the current research, we mainly focused on the successful implementation of SFSCM approach in dairy industry. Eight such challenges were identified from literature survey and experts opinion

4 Research Gap and Research Problem

In the current scenario, several industries were benefited by transportation but one industry where it is critical is the dairy and food industry. On a daily basis, it is very complex for seller, packager, and farm to bring the dairy products to the market without any damage or spoilage. Dairy and food industry is the time-sensitive industry where delaying in time is not acceptable. Seifu et al. [25] investigate the significance of the lactoperoxidase system in the dairy industry and its potential applications during the transportation of milk. He found that most of milk is contaminated during the transportation. Wakeland et al. [26] clearly mentioned that transportation is responsible for greenhouse emission in many developed or developing countries as the dairy or food products travels for a long distance to reach customers. Kumar and Prabhakar [27] found that dairy products facing many problems to meet the extended demands of high-quality dairy products because of the high costs of transportation. Interpretive structure modeling (ISM) approach is used for risk analysis and mitigation for perishable food supply chain: a case of dairy industry by used analytical hierarchy process (AHP) approach and sensitivity analysis to understand the complex nature of the influencing factors in dairy industries [28]. Haq and Boddu [29] used (QFD) quality function deployment approach incorporated with AHP. At the last main points are as follows:

1. It is found that transportation is a big challenge in Indian dairy industry.
2. BWM method is yet not used by any researcher in Indian dairy industry.

Transportation has many challenges which are hurdle for the implementation of SFSCM in Indian dairy industry.

5 Solution Methodology

To demonstrate the projected method and to evaluate its performance, we contacted and collect data from two dairy product-based firms in India [30]. The firms were situated and serving the northern part of the country. They collect raw materials from the nearby villages which can increase the rural economy of country.

Company 1: It is situated in Roorkee, Uttarakhand, and it mainly deals with liquid milk, curd, paneer, cream, butter, ghee, etc. It is also capable of processing 100,000 L of milk into other milk products. They have capacity to collect about 40,000 L of milk from nearby villages. It has annual turnover of two thousand lakhs with around 100 employees.

Company 2: It was established in 2009 and situated in Haridwar, Uttarkhand; it mainly deals in manufacturing and supplying of ghee, cow ghee, fresh lassi, etc. They have capacity to collect about 20,000 L of milk from nearby villages. They follow international standards during their processes. It has annual turnover of one thousand lakhs with around 90–100 employees.

5.1 Best Worst Method (BWM)

Decision making is a part of day-to-day life-related that are related directly or indirectly to our work or individual life. The values and preferences of the experts help to determine evaluate and select the best alternative for decision making. These methodologies require pair-wise comparison among several criteria. A study was conducted by Rezaei in 2015a, and he found that most of the methodologies having problem of consistency during pair-wise comparison. In order to minimize this problem, a new approach of MCDM called the best worst multi-criteria decision-making method (BWM) was developed by Rezaei in 2015a. This approach mainly used to reduce the problem of uncertainty during the pair-wise comparison and effectively implemented by Rezaei et al. in 2015 and 2016. To obtain the weight of the criteria, the following steps involved in BWM method are described below [9, 10]:

Step one: The experts determine a set of decision criteria (C_1, C_2, …, C_n)

Step two: The best and the worst criteria are chosen by experts. Experts choose the best and the worst criteria from the identified criteria in the step one according to their own point of view.

Step three: A pair-wise comparison was conducted by experts between the best criterion and the other criteria.

Step four: A pair-wise comparison was conducted by experts between the other criteria and the worst criterion.

Step five: Obtain the optimal weights: $(W_1, W_2^*, \ldots, W_n^*)$: The main purpose is to find out the optimal weight, i.e., the maximum total differences $\left|W_B - a_{Bj}W_j\right|$ and $\left|W_j - a_{jW}W_w\right|$ for all j is minimized, which is expressed by the model given below:

Min max$\left\{\left|W_B - a_{Bj}W_j\right|, \left|W_j - a_{jW}W_W\right|\right\}$ such that

$$\begin{aligned} &\sum_{\mathrm{j}} W_j = 1 \\ &W_j > = 0, \quad \text{for all } j. \end{aligned} \tag{1}$$

When we convert this equation into linear programming problem, i.e., given below

$$\text{Min } \xi^{\mathrm{B}}$$

Such that,

$$\begin{aligned} |W_B - a_{Bj}W_j| \le \xi^{\mathrm{B}}, \quad \text{for all } j. \\ |W_j - a_{jW}W_W| \le \xi^{\mathrm{B}}, \quad \text{for all } j \end{aligned} \tag{2}$$

$$\begin{aligned} &\sum_{\mathrm{j}} W_j = 1 \\ &W_j > = 0, \quad \text{for all } j. \end{aligned}$$

Solving the linear model (2), we will get optimal weights $(W_1, W_2^*, \ldots, W_n^*)$ and optimal value ξ^{B}.

Table 1 Consistency index [9]

BW[a]	1	2	3	4	5	6	7	8	9
Consistency index	0.00	0.44	1.00	1.63	2.30	3.73	4.47	3.00	5.23

[a] Consistency ratio = ξ^B/Consistency index

Step six: Check the consistency of comparisons:

After completing the step five, the next step is to check the consistency level of the comparisons. Consistency of the comparison depends on the value of ξ^B. In which, a value closer to 0 indicates higher consistency. All the values below 1 for ξ^B indicate consistent comparisons [10]. Table 1 explains about consistency index.

6 Result and Discussions

6.1 Selecting Most Desirable and Least Desirable Challenge Using BWM

After selecting the eight most suitable challenges, decision maker was asked to choose the most desirable and least desirable challenge. Based on their comment, environmental pollution and environmental factors (CH1) and government certification (CH8) are represented as the most desirable challenge (best challenge) and least desirable challenge (worst challenge), respectively. In addition, decision maker was asked for their choices on best-to-other and other-to-worst challenges. Pair-wise comparison is constructed between the challenges. As a result, the table for all comparisons for the best-to-other and other-to-worst challenges is shown in Tables 2 and 3.

Optimal weights and optimal value ξ^B were calculated after identifying the pair-wise comparison between the challenges by Eq. (2). Optimal weights $w_1^*, w_2^*, w_3^*, w_4^*, w_5^*$ and ξ^B were determined after solving the linear programming problem for the challenges shown in Table 4 respectively.

ξ^B value 1.130 indicates a high consistency of 0.3766. It shows the reliability of MCDM problems. According to the weights, CH1 has got first position with the value of 0.177, and CH8 comes at the last place in the list with value of 0.036. Therefore, the ranking of remaining main challenges is CH1 > CH3 > CH4 > CH5 > CH6 > CH2 > CH8.

Table 2 Pair-wise comparison between best-to-other (BO) challenge

Best to other (BO)	CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8
Best challenge (CH1)	1	6	7	6	5	7	7	8

Table 3 Pair-wise comparison between other-to-worst (OW) challenge

Other-to-worst challenges (OW)	Worst challenge (CH8)
CH1	8
CH2	3
CH3	2
CH4	3
CH5	2
CH6	2
CH7	2
CH8	1

Table 4 Optimal weights of the challenges

Criteria	Weights	ξ^B
CH1	0.177	
CH2	0.079	
CH3	0.165	
CH4	0.159	1.130
CH5	0.157	
CH6	0.138	
CH7	0.089	
CH8	0.036	

6.2 Managerial Implications

Handling of transportation is a big challenge in front of decision makers and higher authorities in Indian dairy industry. The present study helps the decision makers to overcome the transportation challenges in Indian dairy industries. This study helps managers in many ways such as:

1. To find out the best (most important) challenge and the worst (less important) challenge among the identified challenges.
2. Understanding of SFSCM-related challenges will help industries for better management of their resources in Indian dairy sector.

7 Conclusion/Limitations

Safety and security of food is directly affected by transportation. Transportation is mainly done by means of roads and railways and also increases the chances of contamination. In present study, challenges in transportation were analyzed by BWM method and it helps the decision makers to implement SFSCM successfully in dairy industry. Environment and its factors attain the highest attention and government certification attains less attention in the study. CH1 has got first position with the value of 0.177, and CH8 comes at the last place in the list with value of 0.036. Therefore, the ranking of remaining main challenges is CH1 > CH3 > CH4 > CH5 > CH6 > CH2 > CH8. This study helps the industries to implement SFSCM successfully. The SFSCM-related challenges were modeled according to the expert's views. The outcomes of the study can be modified further and can be used in the real world. Implementation of SFSCM in dairy industry is not an easy task for decision makers. Adding of sustainability in their processes also increases the costs of manufactured product. In future these challenges were further analyzed by various MCDM approaches like ISM, DEMATEL, GRA and GTMA.

Acknowledgements The authors acknowledge and express the gratitude for the support of the research facilities and funds provided by the Department of Mechanical and Automobile Engineering, Graphic Era University, Dehradun, India.

References

1. Brindley, C. (ed.): Supply Chain Risk. Taylor & Francis (2017)
2. Eshel, G., Shepon, A., Makov, T., Milo, R.: Land, irrigation water, greenhouse gas, and reactive nitrogen burdens of meat, eggs, and dairy production in the United States. Proc. Natl. Acad. Sci. **111**(33), 11996–12001 (2014)
3. Li, D., Wang, X., Chan, H.K., Manzini, R.: Sustainable food supply chain management. Int. J. Prod. Econ. **152**, 1–8 (2014)
4. Ramundo, L., Taisch, M., Terzi, S.: State of the art of technology in the food sector value chain towards the IoT. In: 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI), pp. 1–6. IEEE (2016)
5. Mogale, D.G., Kumar, S.K., Márquez, F.P.G., Tiwari, M.K.: Bulk wheat transportation and storage problem of public distribution system. Comput. Ind. Eng. **104**, 80–97 (2017)
6. Ryan, J.M.: Guide to Food Safety and Quality During Transportation: Controls, Standards and Practices. Academic Press (2017)
7. Dellino, G., Mari, R., Meloni, C.: Waste reduction in fresh food supply chains: an operations research approach. In: Food Waste Reduction and Valorisation, pp. 235–259. Springer International Publishing (2017)
8. La Scalia, G., Nasca, A., Corona, O., Settanni, L., Micale, R.: An innovative shelf life model based on smart logistic unit for an efficient management of the perishable food supply chain. J. Food Process Eng. **40**(1) (2017)
9. Rezaei, J.: Best-worst multi-criteria decision-making method. Omega **53**, 49–57 (2015)
10. Rezaei, J., Nispeling, T., Sarkis, J., Tavasszy, L.: A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. J. Clean. Prod. **135**, 577–588 (2016)
11. Rezaei, J., Wang, J., Tavasszy, L.: Linking supplier development to supplier segmentation using best worst method. Expert Syst. Appl. **42**(23), 9152–9164 (2015)
12. Gupta, H., Barua, M.K.: Identifying enablers of technological innovation for Indian MSMEs using best–worst multi criteria decision making method. Technol. Forecast. Soc. Chang. **107**, 69–79 (2016)
13. Villarreal, B., Garza-Reyes, J.A., Kumar, V., Lim, M.K.: Improving road transport operations through lean thinking: a case study. Int. J. Logist. Res. Appl. **20**(2), 163–180 (2017)
14. Rodrigue, J.P., Comtois, C.: The environmental impacts of transportation. In: The Geography of Transport Systems. Routledge, New York (2013)

15. Moore, C.C.S., Nogueira, A.R., Kulay, L.: Environmental and energy assessment of the substitution of chemical fertilizers for industrial wastes of ethanol production in sugarcane cultivation in Brazil. Int. J. Life Cycle Assess. **22**(4), 628–643 (2017)
16. Price, R.: The modernization of rural France: communications networks and agricultural market structures in nineteenth-century France, vol. 13. Routledge (2017)
17. Hyland, M.F., Mahmassani, H.S.: Taxonomy of shared autonomous vehicle fleet management problems to inform future transportation mobility. Transp. Res. Record: J. Transp. Res. Board **2653**, 26–34 (2017)
18. Muñuzuri, J., Cuberos, M., Abaurrea, F., Escudero, A.: Improving the design of urban loading zone systems. J. Transp. Geogr. **59**, 1–13 (2017)
19. Roca-Riu, M., Cao, J., Dakic, I., Menendez, M.: Designing dynamic delivery parking spots in urban areas to reduce traffic disruptions. SVT Working Paper (2017)
20. Huang, C.C., Huang, C.C., Lai, P.C.: Handling, packaging, and transportation of Phalaenopsis plants for marine shipment. In: Orchid Biotechnology III, pp. 137–143 (2017)
21. Luo, J., Wyatt, J., van der Weerden, T.J., Thomas, S.M., de Klein, C.A., Li, Y., Ding, W.: Potential hotspot areas of nitrous oxide emissions from grazed pastoral dairy farm systems. In: Advances in Agronomy (2017)
22. Adekomaya, O., Jamiru, T., Sadiku, R., Huan, Z.: Minimizing energy consumption in refrigerated vehicles through alternative external wall. Renew. Sustain. Energy Rev. **67**, 89–93 (2017)
23. Rühl, F., Boltze, M.: Freight transport demand management: influencing the freight transport demand within traffic management. In: Dynamic and Seamless Integration of Production, Logistics and Traffic, pp. 163–184. Springer International Publishing (2017)
24. Corazzin, M., Schermer, M., Park, S.Y.: Tools to retain added value in dairy farms: The South Korea case. J. Asian Rural Stud. **1**(2), 81–96 (2017)
25. Seifu, E., Buys, E.M., Donkin, E.F.: Significance of the lactoperoxidase system in the dairy industry and its potential applications: a review. Trends Food Sci. Technol. **16**(4), 137–154 (2005)
26. Wakeland, W., Cholette, S., Venkat, K.: Food transportation issues and reducing carbon footprint. In: Green Technologies in Food Production and Processing, pp. 211–236. Springer, US (2012)
27. Kumar, R., Prabhakar, R.K.: Opportunities and challenges in Indian dairy industry supply chain: a literature review. Int. J. Logist. Supply Chain Manag. Perspect. **2**(4), 791 (2013)
28. Prakash, S., Prakash, S., Soni, G., Soni, G., Rathore, A.P.S., Rathore, A.P.S., Singh, S.: Risk analysis and mitigation for perishable food supply chain: a case of dairy industry. Benchmark. Int. J. 24(1), 2–23 (2017)
29. Haq, A.N., Boddu, V.: Analysis of enablers for the implementation of leagile supply chain management using an integrated fuzzy QFD approach. J. Intell. Manuf. **28**(1), 1–12 (2017)
30. Mangla, S. K., Kumar, P., Barua, M. K.: Risk analysis in green supply chain using fuzzy AHP approach: a case study. Resour. Conserv. Recy. **104**, 375–390 (2015)

Smart Focused Web Crawler for Hidden Web

Sawroop Kaur and G. Geetha

Abstract Huge amount of useful data is buried under the layers of hidden web that is accessible when submit forms are filled by users. Web crawlers can access this data only by interacting with web-based search forms. Traditional search engines cannot efficiently search and index these deep or hidden web pages. Retrieving data with high accuracy and coverage in hidden web is a challenging task. Focused crawling guarantees that the document that is found has a place with the particular subject. In the proposed architecture, Smart focused web crawler for hidden web is based on XML parsing of web pages, by first finding the hidden web pages and learning their features. Term frequency–inverse document frequency will be used to build classifier in order to find relevant pages, using completely automatic adaptive learning technique. This system will help in increasing the coverage and accuracy of retrieved web pages. For distributed processing, MapReduce framework of Hadoop will be used.

Keywords Hidden web · Focused crawler · MapReduce

1 Introduction

The Internet is one of the wealthiest sources of data with its continuously extending and evolving attributes. The web empowers access to numerous website pages with different types of interactive media, for example, content, video. The number of websites and number of Internet users have been constantly expanding. This makes the field of web crawling an interesting research direction. No matter how refined the users are, and which search engine they are using, they are not able to search whole information on the Internet. There is a part of the Internet that we call hidden

S. Kaur (✉) · G. Geetha
Lovely Professional University, Phagwara, Punjab, India
e-mail: srbal@gmail.com

G. Geetha
e-mail: gitaskumar@yahoo.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_42

web or deep web that consists of huge amount of information. Crawlers of generic search engines are not intended to look progressively into database content. For instance, if client needs to look at literacy rates of nations, it is conceivable just utilizing unique database by powerfully produced Hypertext Markup Language (HTML) pages. Some databases are not free; database content is only accessible to those users who have subscription. For example, subscriptions of research paper from ScienceDirect. It is not like traditional crawlers cannot search deep website at all; they index only part of the websites. General search engines cannot search file formats other than HTML, real-time content, websites with login authorization, and websites that are not for open access [1]. Crawling of web pages also differs according to search engines; sometimes some items one search, engine cannot find, but others can locate it. The most popular question in context of hidden web is why to search hidden web and when it is appropriate to use it. The specialized content is the main reason why researchers should use invisible web. When searcher has to utilize a particular search interface, these query-specific interfaces give a high-level control over search results; moreover, the level of accuracy and recall is high in case of deep or invisible web [2].

Another question that arises is what are the features of hidden web content or what is the reason why generic search engines cannot find this web pages that are related to deep web. Deep web is database-oriented and dynamic content-based; crawlers of general-purpose search engines are not designed to query the databases. Usually the databases are subscription-based and accessible to only who have subscribed them. Though some general-purpose search engines index deep websites but they partially index, those parts that are not indexed become part of hidden web. The Internet consists of information that may be structured, semi-structured, and non-structured; even formats for information are different. It also takes time for new formats to appear in search engines. The websites which have login-based authorization, not for open access, blocked by robot exclusion protocol, and the sites that before sending information require information from users; for example, websites related to job search, movie search, and travel sites all are part of deep web that general search engine crawlers cannot crawl. It is almost unfeasible to calculate the volume of data available on the deep web; total data in digital form has reached 6 ZB [3]. And after few years, volume is believed to be increased more than the estimation. Hidden web can be based on single web pages or database information. Usually the creators of single web pages are individual users. Though these web pages may have important information, non-availability of links produces hindrance for the web crawlers to search these pages [1].

To access structured data, user has to go through query interface [4]. Structured data is scattered over multiple web databases so it is a challenge for the search engines to locate such kind of data. One of the pioneer works that has been done in this area is to address this problem, form-focused crawler, and adaptive crawler finds and explores online databases on a particular subject, respectively [5, 6]. This paper proposed a distributed deep web crawler, named as Smart focused web crawler, for hidden web with the goal of achieving wide coverage and high efficiency of web pages.

2 Literature Review

To crawl deep web, primary step is to find where the sources of deep web reside, and then similar sources should be selected to extract the content located in deep web sources [7]. Low harvest rate and low coverage of the hidden web directories are huge issues [8–11]. Page classifier of best focused crawler guide the search for relevant pages, but even with improvement, harvest rate is low [12]. Enhancements to this crawler use additional classifier and do not follow all links in pages. On basis of feedback from baseline classifier, apprentice prioritizes the links in the frontier considering characteristics of good links [13]. Both form-focused crawler (FFC) and adaptive crawler for hidden web (ACHE) are focused crawlers. The FFC has three classifiers. ACHE has improved FFC with adaptive link learning procedure, and feature selection is automatic [5, 6].

Search engines' inability to assign keywords to matching form fields is the reason they are not able to find deep web sources. Purpose of HTML tags is to tell us how things are supposed to look on screen; it does not signify what source it actually is. The eXtensible Markup Language (XML) query languages understand keywords as names of elements that are equivalent to concepts and values of instance, but are usually ambiguous to the user. XML-based crawler can use HTML style href links, link structure to uncover the union of web pages as a suitable query result [14]. Though knowledge discovery from XML documents lies in effective organization of XML documents to retrieve useful information from them, it helps in easily finding relevant documents instead of searching entire collection of XML documents [15]. The power of XML lies in its ability to present web data in a tree or hierarchical-structured form and to represent content in various formats. Data storage in this format has several advantages such as more meaningful information retrieval and knowledge discovery [16]. On this basis, the proposed architecture has one component called convertor, which will convert the non-XML format into XML format. HTML documents are first transformed into XML-based DOM trees and then will be queried and searched for possible links [17]. One another issue in deep web crawling is ranking. Page rank algorithms do not work well with deep web, because it has nothing to do with content; it is entirely based on link structures. Query-independent technique considers web structure, content, and usage mining for ranking of deep web documents. In this way, it considers all factors related to web mining for page rank calculation, and the downside is its being user-based. User priorities can be changed as the time will pass [18].

While Deep Web Search Engine Ranking Algorithm has taken into account both query-independent and query-dependent factors, under three categories of visually related, distance/time-related and data-related factors, ten quality factors are considered in combination with page rank. Scalability and minimal processing are the two best plus points of this technique [19]. Query-independent and query-dependent factors are considered, but it is a very time-consuming technique [20]. Page rank and term weighting technique-based ranking algorithm has taken into account query-independent factors. Though it returns important pages, the results are lacking in similarity factor [21]. Query-dependent factors with Syntactic and Semantic Structure mining, matching on the basis of conceptual, linguistic, and ontology have yielded good results. But it is prone to scam websites, and same title and name without any important information can be used [22]. The broad review of existing web crawlers, their types, and strategies being used is given in [23]. Ranking in the proposed architecture is based on the weighting; high rank will be given to term with high weight.

3 Architecture

Seed sites are list of initial URLs that are utilized to start crawling process to search related pages and domains. For seed sites, initially a list of URLs will be used from Tel-8 dataset available in web browsable form. Depending on the data format, a transformation and semantic annotation into XML is performed by the converter component. Frontier 1 is a queue which contains the to be downloaded URLs by the crawler. The order of URLs in frontier is politeness and relevance bound. The proposed crawler will engage adaptive site learner to learn the features of deep websites, i.e., whether the website contains one or more searchable forms. Minimum number of features that could be considered for it are URL, anchor text, and text around URL.

It is observed that websites from same domains have some similarities (see Figs. 1 and 2); there is similarity in pattern of travel websites. When user fills and submits the initial search form, there could be next child form for results in specialized manner. Site Classifier helps in achieving more accurate results by categorizing URLs into relevant or irrelevant on the basis of homepage content. The proposed system has to do website and form classification. To filter out irrelevant sites, most common method is using term frequency–inverse document frequency (TF–IDF). An improved version of classifying webpages will be used for classifying webpages but with different sections of webpage, that is URL, keywords, anchor text, body and text around URL [24]. JavaMI toolbox is a good help for finding mutual information between the web pages, in order to put similar web pages into one class. Decision on forms whether it is relevant or not is difficult, as each form can have multiple text input, buttons, bullets, etc., so crawler also needs to classify the non-searchable form from searchable forms according to the following criteria.

- Non-searchable forms are those that are used for conducting survey, sending e-mail, and web account login.

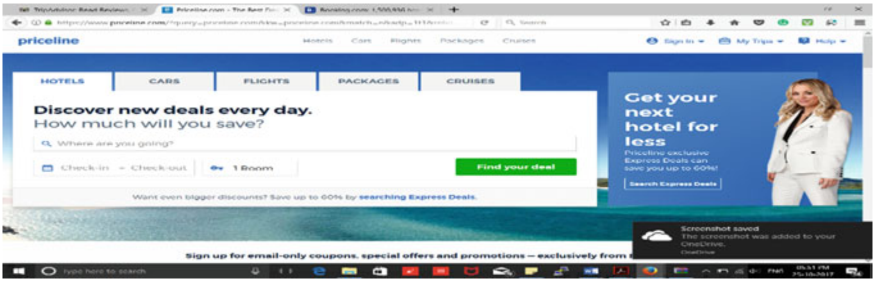


Fig. 1 Illustration of website with search interface

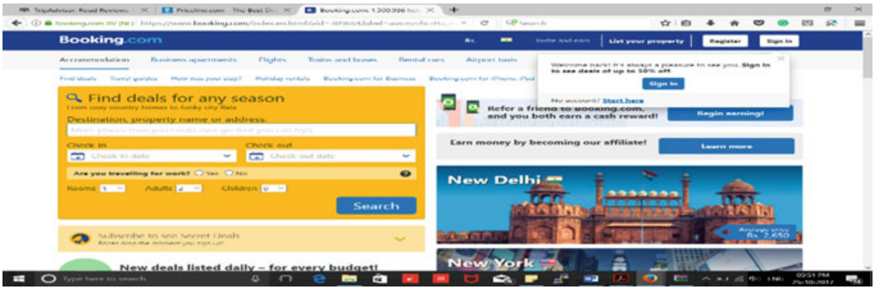


Fig. 2 Illustration of website with search interface similar to Fig. 1

Searchable form should satisfy the following conditions:

- Form has text input file.
- At least one of the set of keywords such as "search," "query," or find [25].
- Query interface with at least two query interfaces for structured data [26].

A crawler will check whether the page belongs to deep website or not, on the basis of the above-mentioned assumption. It parses the webpage, extracts links, and downloads extracted links to find if the links are related or not. A page is said to be relevant if:

- Searchable form is present on page.
- If count of fetched deep websites is higher than threshold [27].

Threshold is user-defined, such as 5, 10, and 50, according to the volume of data and requests of search. For example, we can decide seed sites are 100 or more for experiment. From this number, relevant sites are extracted for site frontier. Score is assigned to unvisited sites that have similarity to the previously discovered deep websites. After giving rank to the site, it needs to be classified as whether it is relevant to topic or not, for a focused crawl by classifier 1. Crawling begins when the URL is classified as related and ignored if irrelevant and next URL is selected from the frontier. Topic is checked for its relevancy from the home page of webpage [18]. For searching inside the web page, breadth-first search will be used, to increase the coverage. Now the search needs to stop somewhere. By default, there is no limit on the depth of crawling. Maximum depth decided is three. Crawler will stop within decided depth, when certain number of pages in each depth is reached. If in certain depth, and after crawling predefined number of pages, crawler is not able to find searchable forms, it moves straight to the subsequent depth. Predefined number is 10; if 10 pages are fetched, but no page consists of searchable form up to depth of three of webpage, crawler will stop and will pick new URL. The Document Object Model (DOM) is ideal for HTML and XML documents parsing in order to form a tree-like structure. Dynamic extraction is easy using DOM as web pages differ in structure. Structure of entire web page can be represented in tree form to find relevant content, and even it is easier to navigate through this model. Balanced link prioritizing is used to prioritize highly relevant links; it is required because in server directories links are unevenly distributed. Even with balance prioritizing problem that occurs is prioritizing the links using relevance can be resulted in biasing toward some directories. Crawler misses the searchable forms from the low rank directories. Directories with few links can be merged to avoid bias.

Ranking is very important phase in deep web, as data is hidden behind the HTML forms and inferring any relationship between them is very difficult. Ranking mechanism prioritizes links for crawlers to quickly find searchable forms. As page rank is not much useful in deep web. To filter searchable forms, domain-independent

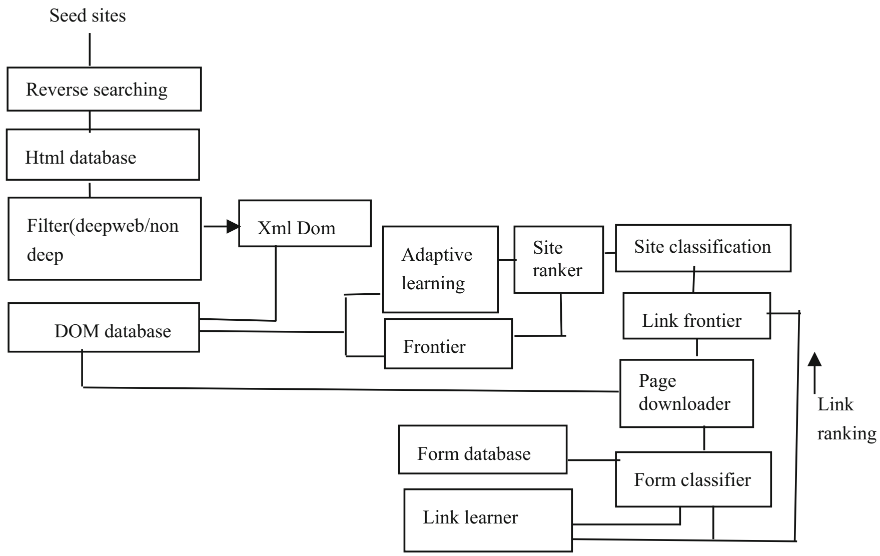


Fig. 3 Smart crawler for hidden web

searchable form classifier will be used [28]. Page downloader directly fetches out center page of the website from the DOM database (Fig. 3).

Links that are extracted in the candidate frontier are ranked by link ranking. Form database contains collection of sites; it collects all data which got input from the form classifier. Architecture consists of three databases to collect starting URLs, XML database, and database to collect searchable forms. Adaptive learner will learn the features of deep websites, searchable forms. Searchable form classifier will use decision tree, and to search domains, Domain searchable form classifier (DSFC) employs support vector machine [29]. To classify whether the page is relevant or not, Naive Bayes classifier will be used. For limiting the scope few kinds of forms like those that require personal information, such as login, username etc. kind of keywords associated, on forums there are feedback forms, comment submission are not under consideration. For distribution, map functions are to be distributed across several machines. The input URLs are split into sets for each map function. Different parallelly working machines will work on input splits [30]. Reducers are distributed using a function for partition. Partition function is used for division of intermediate key space into parts. The number of partitions will depend on the number of URLs in site database. Tel-8 and common crawl datasets will be used. MapReduce function will be called, and the input will split into 64 MB plus copies of this on other clusters. Master will assign each worker one map task or a reduce task. Worker reads the content of input split, parse the pairs of key/value, pass it to map function, and values will be buffered in memory in local disk periodically with locations known to master. After the reducer has all the data, it is sorted and iterated whenever a unique key is encountered. The final output is given by reduce function. MapReduce returns back to initial user code after completion of task. Web crawling at large scale have to face failures, but MapReduce library is designed to tolerate machine failures while processing large amounts of data.

4 Conclusion

We have demonstrated the design of Smart focused web crawler for hidden web, to accomplish wide coverage and high efficiency. Crawler first harvests web pages then does extensive exploration of those sites to find next related web pages. XML helps hierarchical representation of semi-structured and finding relationships between data items. Generic search engine crawlers cannot dig deep into deep web sources due to their inability of assigning keywords to matching form fields. XML support this and has quality of explaining or interpreting meaning of keywords as names of elements that are matching to concepts and instance values. Two learning mechanisms are provided to learn paths of deep websites and forms, respectively, to enhance the efficiency of classifiers. Ranking is based on weights using improved (TF–IDF) technique.

References

1. Scheeren, W.O.: The Hidden Web: A Sourcebook. ABC-CLIO (2012)
2. Sherman, C., Price, G.: The Invisible Web: Uncovering Information Sources Search Engines Can't See. Information Today Inc, Medford, New Jersey (2001)
3. Idc worldwide predictions 2014: Battles for dominance–and survival—on the 3rd platform (2014). http://www.idc.com/research/Predictions14/index.jsp
4. Dragut, E.C., Yu, C., Meng, W.: Meaningful labeling of integrated query interfaces. In: Proceedings of the 32nd International Conference on Very Large Data Bases, pp. 679–690. VLDB Endowment (2006)
5. Barbosa, L., Freire, J.: Searching for hidden-web databases. In: WebDB, pp. 1–6 (2005)
6. Barbosa, L., Freire, J.: An adaptive crawler for locating hidden-web entry points. In: Proceedings of the 16th International Conference on World Wide Web, pp. 441–450. ACM (2007)
7. Olston, C., Najork, M.: Web crawling (foundations and trends®). Inf. Retr. **4**(3), 175–246 (2010)
8. Madhavan, J., Jeffery, S.R., Cohen, S., Dong, X., Ko, D., Yu, C., Halevy, A.: Web-scale data integration: you can only afford to pay as you go. In: CIDR (2007)
9. https://brightplanet.com/2012/06/the-deep-web-surfacing-hidden-value/. Accessed 20 Oct 2017
10. He, B., Patel, M., Zhang, Z., Chang, K.C.C.: Accessing the deep web. Commun. ACM **50**(5), 94–101 (2007)
11. Wright, A.: Searching the deep web. Commun. ACM **51**(10), 14–15 (2008)
12. Chakrabarti, S., Van den Berg, M., Dom, B.: Focused crawling: a new approach to topic-specific web resource discovery. Comput. Netw. **31**(11), 1623–1640 (1999)
13. Chakrabarti, S., Punera, K., Subramanyam, M.: Accelerated focused crawling through online relevance feedback. In: Proceedings of the 11th International Conference on World Wide Web, pp. 148–159. ACM (2002)
14. Graupmann, J., Biwer, M., Zimmer, C., Zimmer, P., Bender, M., Theobald, M., Weikum, G.: COMPASS: a concept-based web search engine for HTML, XML, and deep Web data. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases, vol. 30, pp. 1313–1316. VLDB Endowment (2004)
15. Nayak, R., Senellart, P., Suchanek, F.M., Varde, A.S.: Discovering interesting information with advances in web technology. ACM SIGKDD Explor. Newsl. **14**(2), 63–81 (2013)
16. Suchanek, F.M., Varde, A.S., Nayak, R., Senellart, P.: The hidden web, xml and the semantic web: scientific data management perspectives. In: Proceedings of the 14th International Conference on Extending Database Technology, pp. 534–537. ACM (2011)
17. Tatli, Eİ., Urgun, B.: WIVET—benchmarking coverage qualities of web crawlers. Comput. J. **60**(4), 555–572 (2017)
18. Ahuja, B., Anuradha: SCUM: a hidden web page ranking technique. Int. J. Innov. Res. Adv. Eng. **1**(10) (2014)
19. Wong, B.W.F.: Deep-web search engine ranking algorithms. Doctoral Dissertation, Massachusetts Institute of Technology
20. Balakrishnan, R.: Trust and Profit Sensitive Ranking for the Deep Web and On-line Advertisements. Arizona State University (2012)
21. Batra, N., Kumar, A., Singh, D., Rajotia, R.N.: Content based hidden web ranking algorithm (CHWRA). In: 2014 IEEE International Advance Computing Conference (IACC), pp. 586–589. IEEE (2014)
22. Chahal, P., Singh, M., Kumar, S.: Ranking of web documents using semantic similarity. In: IEEE 2013 International Conference on Information Systems and Computer Networks (ISCON), pp. 145–150. IEEE (2013)
23. Bal, S.K., Geetha, G.: Advances in web crawler. Int. J. Control Theory Appl. **9**(45), 9–30 (2016)

24. Lu, H., Zhan, D., Zhou, L. and He, D.: An improved focused crawler: using web page classification and link priority evaluation. In: Mathematical Problems in Engineering (2016)
25. Madhavan, J., Ko, D., Kot, Ł., Ganapathy, V., Rasmussen, A., Halevy, A.: Google's deep web crawl. Proc. VLDB Endow. **1**(2), 1241–1252 (2008)
26. Wu, W., Doan, A., Yu, C.: Webiq: learning from the web to match deep-web query interfaces. In: Proceedings of the 22nd International Conference on Data Engineering 2006, ICDE'06, pp. 44–44. IEEE (2006)
27. Barbosa, L., Freire, J.: Combining classifiers to identify online databases. In: Proceedings of the 16th International Conference on World Wide Web, pp. 431–440. ACM (2007)
28. Cope, J., Craswell, N., Hawking, D.: Automated discovery of search interfaces on the web. In: Proceedings of the 14th Australasian Database Conference, vol. 17, pp. 181–189. Australian Computer Society Inc. (2003)
29. Zhao, F., Zhou, J., Nie, C., Huang, H., Jin, H.: SmartCrawler: a two-stage crawler for efficiently harvesting deep-web interfaces. IEEE Trans. Serv. Comput. **9**(4), 608–620 (2016)
30. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)

Current Methodologies for Energy Efficient Cloud Data Centers

Sanjay Patel and Krutika Vyas

Abstract Cloud computing is a novel paradigm which is currently defining a new style of system deployment and is realizing the vision of delivering computing as a utility. This is resulting in a heightened cost of ownership, a reduced return on investment, smaller profit margins, decreased reliability and availability of data center resources, and above all it adversely affects the environment through its increasing carbon footprint. While meeting Quality of Service constraints, providers also need to deliver services to users which meet current green criteria. Servers in data centers consume a large amount of energy, making cooling systems necessary. As a result, it is of utmost importance to optimize energy utilization in these datacenter servers. To achieve this we can consume less power of virtualization in cloud computing. Virtualization is a technique which is very useful in the consolidation of virtual machines. In this chapter, a detailed explanation of the current and relevant methodologies relating to energy efficient cloud data centers is presented and the limitations of those methodologies and associated techniques are discussed.

Keywords Cloud computing · Green computing · Green IT
VM consolidation · VM placement · VM selection · Host underload detection
Host overload detection

1 Introduction

Several methods and methodologies have evolved for use in the consolidation of virtual machines (VM) relating to heterogeneous work methods in the cloud computing environment. Its embrace completely different reasonably work, it's

S. Patel (✉) · K. Vyas
Department of Computer Engineering, LDRP-ITR, Gandhinagar, Gujarat, India
e-mail: sanjaypatel54@gmail.com

K. Vyas
e-mail: krutika_ce@ldrp.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_43

having completely different parameters, such as the number of CPUs needed and the buffer size and length of input and output files. To manage heterogeneous work, various methods are employed as described below [1, 2]:

Dynamically modify heterogeneous workload on same physical h/w: In this method, the management of heterogeneous workloads in different virtualized data centers is explored. In order to compare and improve the performance of different heterogeneous workloads, various utility functions are employed.
Cooperative resource provision and management policies: In the same organization, Jing et al. [1] proposed the use of Phoenix cloud to manage heterogeneous workloads. This results in an increased number of completed jobs and is carried out using various simulation kits. This method provides cooperative management and introduces resource provision policies.
MapReduce model to manage heterogeneous workloads: In order to manage heterogeneous workloads, Mell and Grance [2] proposed a new model which directly employs MapReduce. The workload is divided into three main parts, namely I/O bound, CPU-bound without shuffle, and CPU-bound. Simulation results have been obtained by using a triple queue scheduler mechanism the throughput of Hadoop by 30%.
Automatic management of data centers using real-time environments and simulation: To manage the data centers running heterogeneous workloads automatically, Jing et al. [1] introduced a new method whereby the heterogeneous workloads on any server are collocated, and by using resource allocation high-performance goal has been achieved without heterogeneous workloads. In this method, implementation has employed both real-time environments and simulations.
Scheduling methods for high-performance computing (HPC) PC jobs and web-based applications: In order to manage heterogeneous workloads, such as web-based applications and HPC jobs in virtualized data centers, a method was proposed by Mell and Grance [2], which provides maximum benefits to cloud providers. The main aim is to provide energy efficiency to the clients or consumers, Service Level Agreement (SLA) violation penalties, and fault tolerance. With this technique, simulation results show an improved performance and an energy efficiency on the provider side of 15%.

2 Energy-Aware Resource Utilization Requirement

A migration depends mainly on three things: the complexity of the appliance architecture; how loosely coupled the application is; and the amount of effort that is willing to be placed into the migration. Optimization issues within which the input is received online, and within which the output should be designated as the online area unit, are referred to as online issues. Algorithms that are designed for online issues are referred to as online algorithms [3, 4].

2.1 Single VM Migration

In this method for hosting various VMs, a single physical server and host are allocated. The time is split into N time frames and time should be decreased, with the time of each frame being less than 1 s. The cloud provider or resource give to users the physical server consumed energy and the resource provider is responsible for the cost. It is calculated by t1pC1p, where t1p is the time period and C1p is the total power and energy cost (power and energy per given unit time). The single parameters of CPU performance resource usage and the host capacity of various VMs are used. The dynamic workload was provided to each and every VM, and CPU usage varied by VM. The host was either underloaded or overloaded. In order to access CPU performance, all VMs requested by the users, along with the capacity of the CPU demands, will be available. We define some methodologies for CPU demands of the performance of available capacity, between users and resource provider the SLA violation can happen. If a SLA violation is incurred at the provider side, it is represented by Cvtv, where Cv is the cost or penalty of the SLA violation, and tv generally refers to a duration of time SLA violation. In the absence of this generality, the authors define Cp = 1 and Cv = s, where s 2 B+. This is same as defining Cp = 1:0 = s and Cv = 1 [5].

2.2 Dynamic VM Migration

Dynamic migration of generally support the various VMs moments between physical nodes from which VMs created. Sometime VMs do not use resources provided by the server and can be resized logically so as to require a minimum amount of resources from the consolidation of physical nodes, while unused resources off mode or switched off to reduce power consumption and to enable the cloud data center to consume less power also. To provide the high- performance power using the resource utilization with providing the SLAs without focusing the VMs allocation to the process for energy consumption [6, 7].

There are three main issues that require definition when referring to energy efficiency and performance.

First, the server could reduce Quality of Service (QoS) parameters, such as reliability it's cycling the excessive.

Second, it is sometimes necessary to turn off resources in risky dynamic environments as the workload may be variable and consolidation aggressive as point view of random workload. In addition, due to the workload some VMs may not get the required amount of resources from the physical server and as a result the provider may fail to meet the required QoS.

Third, in a virtualized environment SLA, accurate performance parameters must be provided as a matter of course. As per QoS requirements, these policies and mechanisms can be helpful in minimizing energy used and increasing power efficiency without compromising performance requirements [8].

3 Heuristics for Distributed Dynamic VM Consolidation

Dynamic consolidation of VMs will be achieve with support from the study of historical information relating to resource utilization by VMs.

This problem is separated into four parts [9]:

(1) Host is verified as a limited number of process. It ends up being migrated or in either the off or sleep mode;
(2) Host is verified as overladen. To reduce the load, one or more migrations are required or an additional host provides various VMs;
(3) From an overloaded host, the selected VMs must be migrated;
(4) From various VMs from the host select one placement or finding a new one.

4 Host Underload Detection

Using the full host detection formula, all full nodes are first realized. The system then realizes minimum utilization compared with different hosts and provides them to the VMs. If such a system placement is feasible, then once migration is finished the supply the host should revert to off or sleep with the aim of saving energy. If all the VMs from the supply cannot be placed then the host may be kept active [10, 11].

5 Host Overload Detection

To avoid the degradation of performance associated with a SLA violation every VMs figure host sporadically the host overload detection algorithm rule to VMs which is de-consolidate once required.

Several heuristic methods are available for the detection of an overloaded host [12].

A static processor utilization threshold: Using easy processor utilization threshold characteristic the both of overload and not overload various states of a host.
A Markov process model host overload detection: This is a mathematical framework for use in applied mathematical modeling of universal processes.
An adjusting utilization threshold: Median Absolute Deviation: The planned adaptive-threshold algorithmic rule adjusts the value of the processor utilization threshold counting various strength deviation of the processor utilization. The lower value of the higher threshold for utilization purpose. This can be explained by the fact that a greater deviation will increase the probability of processor utilization reaching a maximum and cause a violation of the SLA.
An adjusting utilization threshold: Interquartile Range: In descriptive statistics, the interquartile range (IQR), also known as the midstream or middle fifty, may be a life of applied mathematics dispersion. Unlike the (total) variance, the interquartile variance may be a strong piece of data, for the breakdown purpose for 25%, and thus, commonly most popular entire vary. In the shake, cruciate energy and power distribute, half IQR are similar to the MAD.
Local regression methods: The basic aim for native local regression methodology fitting easy techniques localized subsets knowledge to make up a system curve approximates first data.
Robust native regression: This methodology is generally used in the estimation of successive identifications. If the equalities are not same square measure glad, the physical host is identified as overloaded. This is defined as robust native regression as a result of the algorithm involved.

6 Existing VM Consolidation Strategies

VM consolidation is divided into main four areas as illustrated in figure along with the algorithms and strategies employed. In this section, VM placement strategies are discussed [13, 14] (Fig. 1).

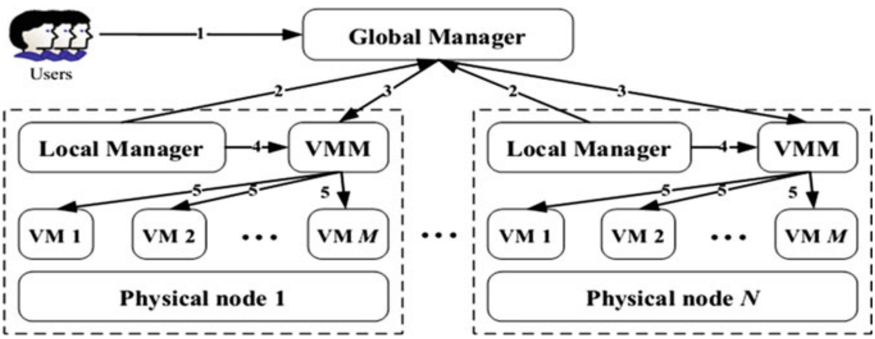


Fig. 1 The system model [15]

The basic aim is to improve the quality of the system, the host underload or overload and VM placement algorithm can be used. This can follow that computer code and also it's a layer of the system is layer native and managers of the systems (Fig. 2). The native managers can reside in each and every module of the virtual machine manager (VMM) layer. The aim of this is to continue monitoring the VMs and to detect host overload and underload. In a scenario of host overload detection, a manager runs and initiates a physical host selection algorithm for the selection of a VM. Another part of the system is the global manager. This resides at the top of the system and its function is to oversee the local managers which are managing the VMs. Whatever decision is taken by the local manager, for optimized VM placement the global manager assigns the commands to the local manager. The main role of the VMM is to perform VM migration and all power modes node [16].

The VM placement refers to as the or mapping as the multidimensional bin packing problem. Regarding mapping, there are several important objectives, and each available item on the tuple should contain the smallest number of bins possible. The physical machines are considered to be bins with a fixed capacity vector.

The main aim of the VM must place into the physical machine representing the physical server capabilities.

The general steps of VM placement algorithms are:

Step 1: All VM requirements by the clients are sorted into decreasing order.

Step 2: Mapping of each VM to PM is performed according to the heuristic definition. Heuristics techniques for VM placement are:

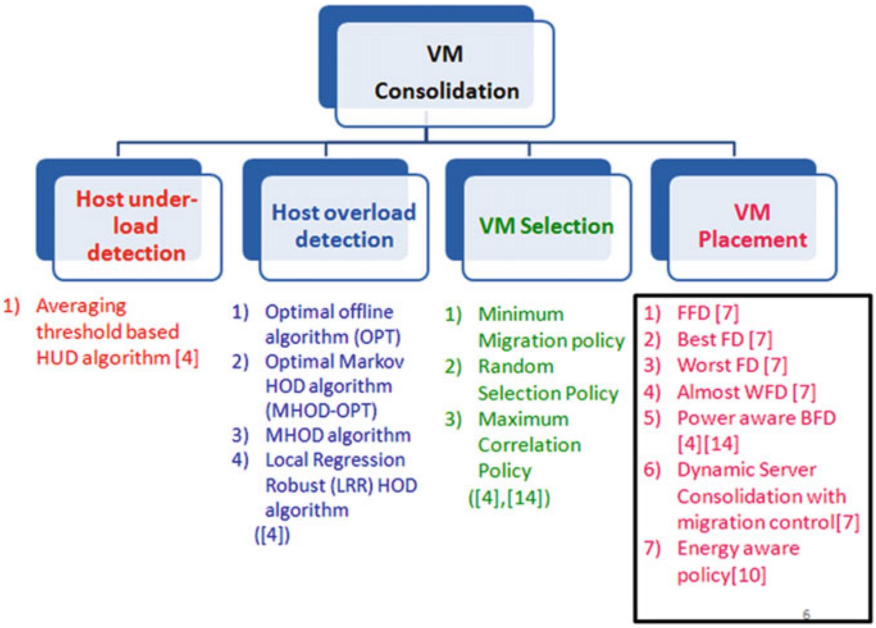


Fig. 2 Subdivision of VM consolidation problems and their related algorithms [17]

(A) **First Fit Decreasing (FFD)**: The VM is mapped to the first PM with available capacity
(B) **Best Fit Decreasing (BFD)**: In this scenario, VMs are mapped to all available PMs left over the least the mapped to the obtained PMs
(C) **Worst Fit Decreasing (WFD)**: In this technique, the biggest left over during VM mapped with PM to the obtained the PMs
(D) **Almost Worst Fit Decreasing (AWFD)**: In this mapping technique, all VMs are mapped to the second best PM for least over the space between all
(E) **Power Aware Best Fit Decreasing (PABFD)**: Each VM should be allocated to a specified host and this reduces the amount of power consumed
(F) **Dynamic consolidation with migration control**: For a steady demand this gives the guarantee of providing stability to the VMs and ensures that VMs with steady requirement do not migrate at any cost.

7 Limitations of Existing Algorithms and Methodologies

Based on a literature survey, existing algorithms and methodologies exhibit numerous limitations and disadvantages. The limitations of existing algorithms and methodologies are detailed below [18].

Energy efficient scheduling of virtual machines in a cloud with deadline constraint: This algorithm is not applicable for I/O intensive or network intensive VMs [7].
Efficient multi-tenant virtual machine allocation in cloud data centers: In this technique, the goal of optimization is obtained by the tedious task. It also integrates the LP-MKP algorithm into various open source cloud computing platforms such as open stack and cloud stack [7].
Real-time tasks oriented energy-aware scheduling in virtualized clouds: The maximum number of CPU cycles is assigned to the available VMs and various tasks must be updated dynamically [8].
Optimized task scheduling and resource allocation in a cloud computing environment using an Improved Differential Evolution Algorithm (IDEA): The major disadvantage of this technique is that the processing time of each and every sub-task depends on a resource. As a result, pre-emption is not allowed in this method [8].
An energy-aware fault tolerant scheduling framework for soft error resilient cloud computing systems: The major disadvantage of this method is that it does not give a guarantee to execute within a deadline and it has compatibility issues when more than two VMs run [8].
Dynamic resource allocation strategies in cloud data centers: This reservation-based resource needs more migration and therefore systems become highly overloaded [9].

Energy-efficient resource allocation and provisioning framework for cloud data centers: Daily dynamic requests received for that must work with improving the workload prediction module. Testing in other architectures is also required [9].
Resource allocation optimization in a data center with energy storage devices: In this method, it is necessary to address an analysis of the power hierarchy in the data center, Complex power models also need to be addressed [10].
Dynamic resource allocation using virtual machines for cloud computing environments: Based on the predicted future resource demands of VMs, the evaluation of resource allocation status can be updated. This method does not provide an efficient prediction for dynamic and efficient requests [10].
A green energy-efficient scheduling algorithm using the DVFS method for cloud datacenters: The system model is very complicated to implement in real-time cloud environments and heterogeneous servers. Whatever servers are chosen for a particular job must satisfy QoS requirements [10].
QoS-based efficient resource allocation in cloud computing: In this method, a proposed algorithm is implemented in a CloudSim toolkit. This is also the case in other cloud simulators and real-time cloud environments [11].

8 Conclusion and Future Directions

It can be concluded that energy efficiency is important in a data center. To enable VM consolidation, virtualization is a very important technique. In this chapter I have considered network topology information only in terms of reducing the energy consumption incurred by network components when the VM has migrated a long way from the source. In future work energy consumption may be further reduced by considering the topology formed by the communicating VMs. These communicating VMs may have physical nodes providing costly data transfer between each other. It can also be concluded that there are numerous limitations and disadvantages present in existing methodologies.

References

1. Jing, S.-Y., Ali, S., She, K., Zhong, Y.: State-of-the-art research study for green cloud computing. J. Supercomput. (Springer) **8** (2013)
2. Mell, P., Grance, T.: The NIST Definition of Cloud Computing. Special Publications (2013). https://doi.org/10.6028/NIST.SP.800-145
3. Feller, E., Rilling, L., Morin, C.: Energy-Aware Ant Colony Based Workload Placement in Clouds, Institut National de Recherche en Informatique et en Automatique. ISSN: 0249-6399, inria-00594992, version 1, 23 May 2011
4. Buyya, R., Beloglazov, A., Abawajy, J.: Energy-efficient management of data center resources for cloud computing: a vision, architectural elements and open challenges. In:

Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) (2010)
5. Beloglazov, A., Buyya, R., Choon Lee, Y., Zomaya, A.: A taxonomy and survey of energy-efficient data centers and cloud computing systems. IEEE (2010)
6. Minas, L., Ellison, B.: Energy efficiency for information technology: how to reduce power consumption in servers and data centers. Intel Press (2009)
7. Barroso, L.A., Holzle, U.: The case for energy-proportional computing. Computer 33–37 (2007)
8. Wei, G., Liu, J., Xu, J., Lu, G., Yu, K., Tian, K.: The on-going evolutions of power management in Xen. Technical Report, Intel Corporation (2009)
9. VMware Inc., vSphere resource management guide (2009)
10. How VMware virtualization right-sizes IT infrastructure to reduce power consumption (2009)
11. Qumranet Inc, KVM: kernel-based virtualization driver. White Paper (2006)
12. Nathuji, R., Schwan, K.: VirtualPower: coordinated power management in virtualized enterprise systems. ACM SIGOPS Oper. Syst. Rev. **41**(6), 265–278 (2007). Institute of Standards and Technology 1-4244-0344-8/06/2006. ACM
13. Raghavendra, R., Ranganathan, P., Talwar, V., Wang, Z., Zhu, X.: No “power” struggles: coordinated multi-level power management for the data center. SIGARCH Comput. Archit. News **36**(1), 48–59 (2008)
14. Stillwell, M., Schanzenbach, D., Vivien, F., Casanova, H.: Resource allocation using virtual clusters. In: Proceedings of the 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid 2009), Shanghai, China, 2009, pp. 260–267 (2009)
15. Verma, A., Ahuja, P., Neogi, A.: pMapper: power and migration cost aware application placement in virtualized systems. In: Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware, pp. 243–264. Springer-Verlag, New York, Inc. (2008)
16. Buyya, R., Beloglazov, A., Abawajy, J.: Energy-efficient management of data center resources for cloud computing: a vision, architectural elements, and open challenges. In: Proceedings of the 2010 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA 2010), Las Vegas, USA, 12–15 July 2010
17. Kim, K.H., Beloglazov, A., Buyya, R.: Power-aware provisioning of cloud resources for real-time services. In: Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science (MGC 2009), Urbana Champaign, Illinois, USA, pp. 1–6 (2009)
18. Beloglazov, A., Abawajyb, J., Buyyaa, R.: Energy aware resource allocation heuristics for efficient management of data centers for Cloud computing. Future Gener. Comput. Syst. **28**, 755–768 (2012)
19. Beloglazov, A., Buyya, R.: Energy efficient resource management in virtualized cloud data centers. In: 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (2010)
20. Kliazovich, D., Bouvry, P., Khan, S.U.: GreenCloud: A Packet-Level Simulator of Energy-Aware Cloud Computing Data Centers. Springer Science Business Media, LLC (2010)
21. Beloglazov, A., Buyya, R.: Energy efficient resource management in virtualized cloud data centers. In: 2011 11th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing (2011)
22. Wang, L., von Laszewski, G., Huang, F., Dayal, J., Frulani, T., Fox, G.: Task scheduling with ANN-based temperature prediction in a data center: a simulation-based study. Springer-Verlag, London Limited (2011)
23. Mao, L., Liu, B., Lin, W.W.: An energy-efficient resource scheduling algorithm for cloud computing based on resource equivalence optimization. Int. J. Grid High Perform. Comput. **8** (2), 43–57 (2016)

An Enhanced Scheme for PHR on Cloud Servers Using CP-ABE

Nishant Doshi, Mayur Oza and Nikita Gorasia

Abstract In today's digital world, cloud computing gives the illimitable possibilities. Personal Health Record (PHR) enables patients to store, share, and access personal health data in centralized way such that it can be accessible from anywhere and anytime. Indeed, coalescing of PHR with cloud gives incipient horizons for medical fields to be digitalized in centralized storage but it comes with major concern i.e. security. There are many researchers are work in securing PHR which stored in cloud utilizing gave approaches but it's not enough to secure it. So there is desideratum for incipient technology as Attribute Based Encryption that secure PHR with providing many functionalities such as revocation of utilizer, delegation of other utilizer access, accountability, probing over encrypted files, multi-authority and many more.

Keywords CP-ABE · Cloud data security · PHR · Attribute based encryption

1 Introduction

Personal Health Records (PHR) is electronic data which contain patient's electronic health records. PHR sanctions utilizer to store, share and retrieve medical data with friends, family and medicos. PHR is store in centralized way so PHR can be simply accessible from anywhere and anytime. But health data is sensitive, so disclosure of PHR can put patient in hazard. Albeit, access of health data in the professional medical domain is tightly controlled by subsisting regulations, such as the U.S. Health Indemnification Portability and Accountability Act (HIPAA).

The aim of research to find current trends to secure PHR on cloud utilizing different techniques and finding best technique that provides most applicability in

N. Doshi (✉)
Pandit Deendayal Petroleum University, Gandhinagar, India
e-mail: Doshinikki2004@gmail.com

M. Oza · N. Gorasia
Marwadi Education Foundation, Rajkot, India

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_44

authentic world. Here these work shows that Attribute Based Encryption (ABE) is good technique to secure PHR with providing many functionalities. We wanted to enhance one of the scheme that gives accountability with probing capabilities over encrypted data blocks.

In current days, working in immensely colossal company engender sizably voluminous amount of data so infrastructure is not adequate for companies to store it, that's why companies move towards to store data at cloud. However, cloud computing additionally comes with some security quandaries that can be solved afore use for that companies. Out-sourcing of data may be obviated from confidentiality, integrity and privacy of client's information that must be bulwarked by some mechanism. As an evolution of cloud computing its accessible as software-as-a-service or platform-as-a-service gives facilely access from anywhere and anytime by just simply connecting to server.

Therefore, amalgamation of PHR with cloud gets incipient horizons for medical records to store and retrieve in centralized way that can be facilely manage with diminutive effort of patients, medicos and care givers. There is always a jeopardy to store a data on cloud and storing data of patient's health is genuinely critical to store at untrusted server because any misuse or alteration of data can cause crucial damage on patient's health or his reputation. Thus the desideratum of securing PHR data at cloud.

In older days the health records are store in medical journals/notes and manage by hospitals and medical stores, but the management of hardcopy is tedious to store, share and search for some records. However, the appearance of the digitalization of this medical records are converted into digital copies which is kenned as Electronic Health Records (EHRs). Its similarly manage by hospitals and facile to probe the records, but patient has no control on it. And it additionally increases the cost to stores more data, needs more data centers and webservers. There is additionally problem with medical records which are not facile to apportion because it is managed by Hospitals and not under control of patient. Data storage are not able to elastically stretchable according to comes with incipient data. However, data is stored on cloud servers gives that accommodation well with Database-as-a-service. PHR are engendered, manage and control by Patient itself. PHR is electronic data which contain patient's electronic health records. PHR sanctions patients to store, retrieve and apportion medical data with friends, family and medicos. PHR is store in centralized way so it can be openly accessible from anywhere and anytime. But health data is sensitive, so inopportune disclosure of PHR can put patient in vulnerably susceptible condition. The detailed description of PHR is given in [12, 13].

Attribute Based Encryption is an elongated work to Identity Based Encryption (IBE) where identity field of utilizer holds descriptive attributes rather than string as in IBE. In ABE utilizer has identity as w attributes data is encrypted utilizing some w' attributes. So when utilizer want to decrypt this data then the attributes and w need some threshold level d (predefined) of homogeneous attributes than only he can decrypt that data.

In ABE there are two variants predicated on placing attributes and access structure (attribute policy) i.e. (i) CP-ABE: In CP-ABE scheme the access structure is residing on cipher text and culminate utilizer has a list of attribute to verify and match it with access structure for prosperous decryption (ii) KP-ABE: KP-ABE is complementary scheme of CP-ABE where encryption is done with set of attributes and on decryption of data needs access structure.

2 Literature Survey

In 2009, Benaloh et al. [3] suggested that security and privacy is major issue for storing data at third party. For that they suggested to do encryption utilizing symmetric key afore storing data to cloud servers. They propose scheme for different caliber of key for different part of data. But this scheme was not fortify to fine-grained access or multi-owner of file. However, Damiani et al. [7], Atallah et al. [2], Wang et al. [18] and Boldyreva et al. [5] gives solution for securing outsourcing data utilizing Identity Based Encryption and Hierarchical Identity-Based Encryption (HIBE) techniques. In HIBE the outsourcing data is encrypted utilizing Identity Based Encryption but key is hierarchical so utilizer only can access data if they have at some threshold level of key.

On other hand, Attribute Based Encryption is introducing by Sahai et al. [17] and more work on ABE to provide more functionalities in like Ciphertext Policy Based ABE and Key Policy Based ABE in Bethencourt et al. [4], Chase et al. [6], Waters [8]. After the first integration of ABE with PHRs by Ibraimi et al. [11] in 2009 which gives securing outsourcing of data by encryption with ABE. They provide flexible delegation and facile revocation of attribute access for utilizer. Afterwards many researchers working on it gives incipient functionalities to utilizer and more security than other schemes Li et al. [13, 14], Akinyele et al. [1], Xhafa et al. [19], Dubovitskaya et al. [9], Fabian et al. [10]. Brief survey in these field of Securing PHR by CPABE provided by Oza et al. [16] with analysis of all schemes and their quandaries.

3 Preliminaries

3.1 Bilinear Pairing

Consider two cyclic multiplicative groups of prime order p are G and $G_T \cdot e$ called bilinear paining if and only if $e\colon G \times G \rightarrow G_T$ has following properties:

1. Bilinearity: $e(g^a, g^b) = e(g,g)^{ab}$ for all $a, b \in {}_R Z_p$.
2. Non-degeneracy: there such $g_1, g_2 \in G$ exist such that $e(g_1, g_2) \neq 1$.
3. Computable: It is feasible to compute $e(g_1, g_2)$ for all $g_1, g_2 \in {}_R G$.

3.2 Access Structure

L represents list of attributes as $L = [L_1, L_2, \ldots, L_n]$ and Accesss structure $W = [W_1, W_2, \ldots, W_n]$. Here Access structure is said to be accepted ($L \models W$) if $L_i = W_i$ else it's not accepted.

4 Xhafa's Scheme

In Xhafa et al. [19] analysis part of scheme they claim that the scheme is secure under Decisional bilinear, Diffie-Hellman Decisional-linear Assumption. In their scheme they consider timing for encryption as $(4n + 4)T_{exp} + T_{expt}$ and for decryption step timing is as $((\Sigma_{k \in [N]}) \cdot ((4n_k + 4\rho + 1)T_{pair} + 3T_{exp}$ where T_{exp} for timing of exponential in G, T_{expt} for timing of exponential in G_T, T_{pair} for timing of pairing, n for total number of attribute values, n_k for number of attributes managed by AA_k.

Analysis of Xhafa's scheme in tabular form presented in Table 1. These table shows that the time involution and execution time is more due to astronomically immense number of operations. However we proposed incipient scheme that required minuscule number of operation that reduce execution time that is presented in next section.

5 Proposed Scheme

The proposed scheme consists of seven algorithms as follows. The block diagram between them is given below in Fig. 1.

Setup: This algorithm run CA and it optate bilinear group G of prime order p with generator as g. Now CA optate two random β, γ. It calculates h and Y and output of

Table 1 Analysis of Xhafa's scheme

	Setup AttKeyGen encryption decryption				Trace
Select random	N	2N	1	–	s
Hashing	–	5N	2N	–	2 * s * N
Exponent	N	13N	8N	–	8 * s * N
Pairing	N	–	–	8 (N * n_k) + 1	–
Multiplication	–	6N	1	2 (N * n_k) + 3	s
Division	–	N	–	–	–
Addition	–	N	–	–	–
Concatenate	–	–	18N	–	18 * s * N

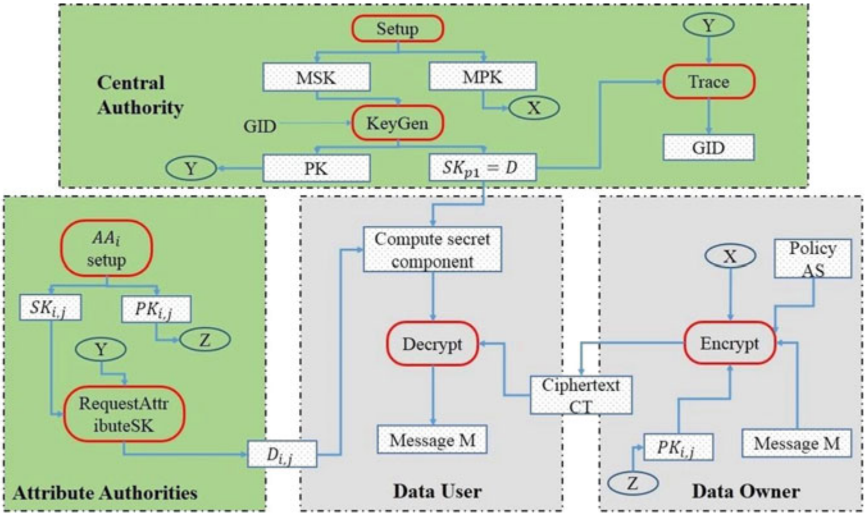


Fig. 1 Framework for securing PHR by CP-ABE

this step master public key as parameter MPK and master secret key MSK as follows:

$$\text{MPK} = \{G, g, h = g,\ Y = e(g,\ g)^{\gamma}\}\ \text{MSK} = (\beta,\ \gamma)$$

AA

$\boldsymbol{AA_i}$**Setup** → In this algorithm run by every AA to generate PK and SK for every attribute i.

$$SK_i = \text{GID}_{AA_i} * \boldsymbol{\alpha}_{i,j}\ PK_i = g^{\text{GID}_{AA_i} * \boldsymbol{\alpha}_{i,j}}$$

KeyGen(MSK, GID) → This algorithm run by CA. It takes input as MSK and Global Identity of user GID. Output of this algorithm is part of secret key and full private key for user with GID. Additionally, in this step we take one table that contains GID and public key pair for user identification in tracing of misbehaving user.

$$D = \frac{y + r * GID_{user}}{\beta},\quad PK_i = g^{r * GID_{user}}$$

RequestAttributeSK(PK, SK$_i$) → This algorithm run by every AAs corresponding to attributes of utilizer. It takes input as public key of utilizer and secret key of Attribute ascendancy that is corresponding to attribute. It outputs the component of secret key that is utilizable for engendering users secret key that use for decrypting data.

$$D_{i,j} = (PK)^{GID_{AA_i} * \alpha_{i,j}}$$

Encrypt (M, W) → This algorithm run by client (data owner) that want to secure data. It takes Message M and Access Structure W as input. Output of this algorithm is ciphertext CT. This is similar to that in [18].

$$C_1 = M \cdot Y^s$$

$$C_2 = \left(\prod_{t \in W} PK_{i,j} \right)^s$$

$$C_3 = h^s = g^{\beta s}$$

$$C_4 = g^s$$

Decrypt (CT, PK_u): This algorithm run by cloud user that want to access file. It takes input as ciphertext CT and public key PK and secret key SK of cloud user. This is somewhat similar functioning to that in [18].

$$\frac{C_1 \cdot e(PK, C_2 \cdot C_4)}{e(C_3, D) \cdot e(C_4, \prod_{t \in AS} D_{i,j})}$$

Trace^D(PK_u) → GID: This algorithm run by CA. It takes input as public key PK from misbehaving device and gives identity GID of that user.

5.1 Analysis

The analysis of the proposed scheme is given in Table 2.

Table 2 Analysis of proposed scheme

Select random	Multiplication	Exponential	Pairing	Division	Addition
Setup 2	–	1	1	–	–
AA_iSetup 1	N	–	–	–	–
KeyGen 1	2	2	–	1	1
RASK	–	M	–	–	–
Encryption	p + 1	p + 3	–	–	–
Decryption	M + 3	–	3	–	–

6 Conclusion and Future Work

As we shown in this work that current techniques for storing data on cloud is not enough for security and efficiency. So here we provide a secure and efficient scheme for securing health data. We additionally provide comparison between Xhafa's scheme and our approach for number of operations and required timing. We additionally implemented proposed scheme by utilizing rudimental toolkit of CP-ABE that provide efficient and secure encryption of data. We withal explicated the securing sample health dataset as UCI Diabetic dataset [15] by our proposed scheme.

Here we endeavoring to cover all aspect of securing health data but still programming part can be amending if anyone is fascinated to secure it by some incipient or elongated scheme. In this paper we take sample data of Diabetic from UCI dataset but this work can be very utilizable in authentic dataset of any Hospital or health record collectors. As growing of ABE variation there is incipient possibilities to coalesce more that variation for securing data.

References

1. Akinyele, J.A., Pagano, M.W., Green, M.D., Lehmann, C.U., Peterson, Z.N., Rubin, A.D.: Securing electronic medical records using attribute-based encryption on mobile devices. In: Proceedings of the 1st ACM workshop on Security and Privacy in Smartphones and Mobile Devices, pp. 75–86. ACM (2011)
2. Atallah, M.J., Blanton, M., Fazio, N., Frikken, K.B.: Dynamic and efficient key management for access hierarchies. ACM Trans. Inf. Syst. Secur. (TISSEC) **12**(3), 18 (2009)
3. Benaloh, J., Chase, M., Horvitz, E., Lauter, K.: Patient controlled encryption: ensuring privacy of electronic medical records. In: Proceedings of the 2009 ACM Workshop on Cloud Computing Security, pp. 103–114. ACM (2009)
4. Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-policy attribute-based encryption. In: 2007 IEEE Symposium on Security and Privacy, SP'07, pp. 321–334. IEEE (2007)
5. Boldyreva, A., Goyal, V., Kumar, V.: Identity-based encryption with efficient revocation. In: Proceedings of the 15th ACM Conference on Computer and Communications Security, pp. 417–426. ACM (2008)
6. Chase, M., Chow, S.S.: Improving privacy and security in multi-authority attribute based encryption. In: Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 121–130. ACM (2009)
7. Damiani, E., di Vimercati, S., Foresti, S., Jajodia, S., Paraboschi, S., Samarati, P.: Key management for multi-user encrypted databases. In: Proceedings of the 2005 ACM Workshop on Storage Security and Survivability, pp. 74–83. ACM (2005)
8. Doshi, N., Jinwala, D.: Constant ciphertext length in multi-authority ciphertext policy attribute based encryption. In: 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011), Allahabad, pp. 451–456 (2011)
9. Dubovitskaya, A., Urovi, V., Vasirani, M., Aberer, K., Schumacher, M.I.: A cloud based e-health architecture for privacy preserving data integration. In: ICT Systems Security and Privacy Protection, pp. 585–598. Springer (2015)
10. Fabian, B., Ermakova, T., Junghanns, P.: Collaborative and secure sharing of healthcare data in multi-clouds. Inf. Syst. **48**, 132–150 (2015)

11. Ibraimi, L., Petkovic, M., Nikova, S., Hartel, P., Jonker, W.: Ciphertext-policy attribute based threshold decryption with flexible delegation and revocation of user attributes (2009)
12. Kaelber, D.C., Jha, A.K., Johnston, D., Middleton, B., Bates, D.W.: A research agenda for personal health records (PHRs). J. Am. Med. Inform. Assoc. **15**(6), 729–736 (2008)
13. Li, M., Yu, S., Ren, K., Lou, W.: Securing personal health records in cloud computing: patient-centric and fine-grained data access control in multi-owner settings. In: Security and Privacy in Communication Networks, pp. 89–106. Springer (2010)
14. Li, M., Yu, S., Zheng, Y., Ren, K., Lou, W.: Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. IEEE Trans. Parallel Distrib. Syst. **24**(1), 131–143 (2013)
15. Lichman, M.: UCI machine learning repository (2013). http://archive.ics.uci.edu/ml
16. Oza, M., Gorasiya, N., Doshi, N.: Securing Personal Health Records in the Cloud by Using Attribute Based Encryption. A Review. Grin Publisher, Verlag, Germany (2016)
17. Sahai, A., Waters, B.: Fuzzy identity-based encryption. In: Advances in Cryptology–EUROCRYPT 2005, pp. 457–473. Springer (2005)
18. Wang, W., Li, Z., Owens, R., Bhargava, B.: Secure and efficient access to outsourced data. In: Proceedings of the 2009 ACM Workshop on Cloud Computing Security, pp. 55–66. ACM (2009)
19. Xhafa, F., Feng, J., Zhang, Y., Chen, X., Li, J.: Privacy-aware attribute-based PHR sharing with user accountability in cloud computing. J. Supercomput. **71**(5), 1607–1619 (2015)

Decision Tree Algorithms for Prediction of Heart Disease

Srabanti Maji and Srishti Arora

Abstract In the present scenario, maximum causes of death are heart disease. Many researches are taking place to detect all types of heart diseases at very early stage. Scientists are using various computational techniques to predict and prevent heart diseases. Using data mining techniques, the number of tests that are required for the detection of heart disease reduces. In this paper, hybridization technique is proposed in which decision tree and artificial neural network classifiers are hybridized for better performance of prediction of heart disease. This is done using WEKA. To validate the performance of the proposed algorithm, tenfold validation test is performed on the dataset of heart disease patients which is taken from UCI repository. The accuracy, sensitivity, and specificity of the individual classifier and hybrid technique are analyzed.

Keywords Data mining · Decision tree · Heart disease

1 Introduction

Artificial intelligence is a way of making a computer, a robot, or software think intelligently. Machine intelligent unit makes our system smart without explicit programming. The two main types of machine learning techniques used to classify dataset include supervised learning and unsupervised learning. In the supervised learning process class label is known, while in the unsupervised learning process they are unknown. Clustering comes under the category of unsupervised learning method, whereas classification and regression are predictive methods [1]. The different types of techniques are shown in Fig. 1.

S. Maji (✉) · S. Arora
Dehradun Institute of Technology, Dehradun, Uttrakhand, India
e-mail: srabantiindia@gmail.com

S. Arora
e-mail: srishtiarora35@ymail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_45

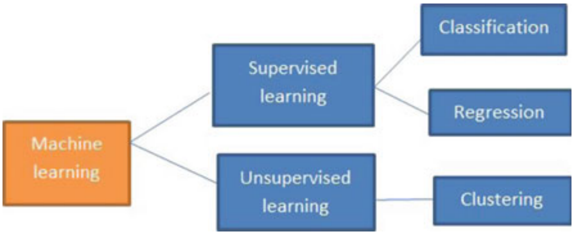


Fig. 1 Categories of machine learning techniques

Data mining [2] is utilized for exploring, analyzing, and extracting data of different fields and discovers meaningful information from it; i.e., raw data is converted into useful information. This extracted information can be useful for different aspects like healthcare industry, financial estimation, weather forecasting, study of organic composite and medical treatment [3, 4]. It uses available datasets to perform its applications. Machine learning and data mining together can be used to develop spam mail detectors, self-driven cars, speech recognition, face recognition, recommendation systems, etc.

Various diseases can be diagnosed like heart diseases, cancer, thyroid, diabetes. Heart disease prediction system can assist medical professionals in predicting heart disease based on the clinical data of patients [5]. The main risk factors for heart disease are [6]:

- High LDL (bad cholesterol).
- Smoking.
- Low HDL (good cholesterol).
- Uncontrolled hypertension.
- Uncontrolled diabetes.
- Corpulence.
- Physical inactivity.
- Uncontrolled stress.

Therefore, by utilizing various attributes of heart disease attributes, our proposed model can predict and easily diagnose the patients with heart disease [7].

2 Related Work

Different authors have represented their research by exploring various techniques which include classification, association mining, clustering, decision tress in different health fields. Various chronic diseases like asthma, blood pressure, diabetes cannot be cured easily, but high risk of diseases can be controlled with accurate and timely update data of the patients.

The diagnosis of diseases including blood pressure and diabetes with the help of neural networks was introduced [8]. Various experiments were performed on the collected dataset of different diseases. The technique is trained by using training dataset and was tested by using 13 attributes which include blood pressure, age. The supervised network was applied to diagnose heart diseases. When the new dataset is inserted, it compares the result and finally produces a list of diseases of the patient.

In 2013 [9] presented a paper in which they implemented Naïve Bayes and decision tree algorithms for prediction of heart disease. Their main purpose was to diagnose and analyze heart disease with less number of attributes.

In 2014 [10], M. A. Nishara Banu and B. Gomathy, proposed an article in which the data is first preprocessed and stored in the database. Then, this processed data is clustered using clustering algorithms which includes K-means clustering technique. The patterns observed are then classified into different classes using the C4.5 algorithm. The performance of decision tree algorithm is measured on the basis of information entropy. The result demonstrates that the designed system is useful for prediction of heart attack successfully.

In [11], the authors have created a hybrid model using data mining and artificial neural network (ANN) techniques in which a multilayer perceptron neural network algorithm is used. This system helps in detection of disease at early stages. The experimental result shows that using ANN heart disease is predicted with nearly 100% accuracy and in less time. Thus, an ANN technique overpowers HDPS and is found to be the best.

Classification [12] methods are most efficient for predicting the heart disease when the dataset contains the large number of attributes. In this paper, the authors use information gain split as it gives more number of attributes than the others. Thus, ANN and Naïve Bayes give accuracy than the other techniques. The information attribute filtering is best for predicting heart disease as it reduces the number of clusters.

In [13], the authors introduced a model for heart disease prediction with data mining algorithms which include Naïve Bayes classifier and decision tree. The validation tests are performed to predict the accuracy of the classifier. The analysis shows that the decision tree performs better than Naïve Bayes classifier for the given dataset.

To predict and diagnose the cardiovascular disease [14], the authors have compared RIPPER classifier, decision tree, artificial neural networks (ANNs), and support vector machine (SVM). After experimentation, it was concluded that out of all four classification techniques, SVM turns out to be the best technique for heart disease prediction with highest accuracy.

3 Preliminaries

3.1 Artificial Neural Network

ANN or artificial neural network is a machine learning technique. It is similar to neural structure of brain. These networks consist of large number of nodes which are used to perform computational and mathematical operations. It is not used for predictive analysis as it overfits the model or relation. There are mainly three layers of ANN, namely input layer, hidden layer, and output layer. The algorithm is as follows:

1. The data is to be processed and is given to the input layer.
2. The values are transferred to the hidden layer using some random weight value which is further transferred to output layer.
3. For every input value, the output is compared with the desired output.
4. If the output is correct, no alteration is done in the weight value. Otherwise, modify the weight value according to the desired result.
5. The information is processed, and output layer gives the output.

3.2 C4.5 Algorithm

Information gain splitting attribute is bias to the attributes with lots of values. Therefore, C4.5 is used instead of information gain to reduce the influence of biasing. It is used for implementing classification rules in the form of decision trees for a given set of data. In WEKA, J48 works as C4.5. In this algorithm, the entire data needs to be fit in the memory. It is more useful as it handles missing attributes. Also, tree post-pruning can be performed using this algorithm.

3.3 Dataset

In our research, the experiment is conducted using WEKA tool. This tool is the combination of machine learning and data mining techniques. The information of UCI repository is introduced in the database. There are total 13 attributes and 270 instances in this dataset. In order to perform analysis in WEKA tool, the datasets must be in the Attribute-Relation File Format (ARFF). This tool lets us pre-process and filters data based on different useful attributes. The best splitting attribute is selected. Finally, results are compared between different algorithms on the basis of accuracy, sensitivity, relative errors, etc., and thus, the conclusion is inferred from it. This trained model can be used for prediction of heart disease and can definitely

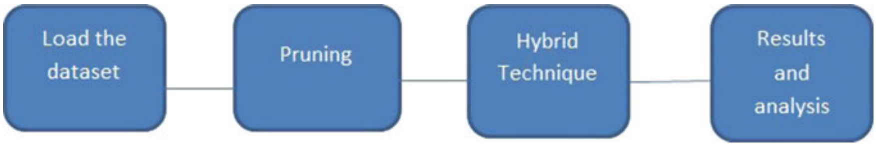


Fig. 2 Proposed model

help to diagnose patient's heart disease well in advance. This reduces the number of tests. Thus, disease can be cured at the right time and lives of thousands of people can be saved. The proposed model is shown in Fig. 2.

4 Methodology

In today's era, millions of people are suffering from some sort of heart disease every year. Heart disease has become one of the biggest killers for both men and women around the world. Medical diagnosis plays a very important role to diagnose any type of heart disease. To reduce the cost of clinical tests, proper and trained computer systems should be build which are capable of taking effective decisions. Data mining implies the use of software techniques for finding patterns and consistency in recorded set of data.

Many researchers are developing different hybrid techniques by combining different data mining techniques to help doctors and other medical experts in the heart disease prediction. Artificial neural network, Naive Bayes, decision trees, support vector machine, etc., are some of the machine learning techniques used for detection of different diseases. The different open-source frameworks which are available for performing these techniques include R, Python, and Apache Mahout.

In our proposed system, decision tree technique, i.e., C4.5 algorithm is combined with ANN and named as hybrid DT to produce the desired result. The dataset is loaded, and pruning is done to remove noise. Then, proposed technique is applied on the clean dataset. The accuracy, sensitivity, and specificity of individual algorithm and the hybrid DT are determined. The result is validated using tenfold validation test. Thus, it concludes which algorithm works the best and is faster in detecting the symptoms of heart disease for a given dataset.

5 Application of Data Mining

Data mining has been used to find large amount of uncover patterns from stored data among researchers, as it provides several benefits to healthcare industry. The objectives cum applications of data mining together with machine learning are discussed as follows [15]:

- Managing hospital resources: Machine learning lets us build a model which manages all hospital resources. It helps in detection of chronic diseases, and based on this prediction, effective treatment can be given to the patients on timely basis. Reports play a vital role for utilizing and managing hospital resources effectively. Various healthcare centers across the world provide different guidelines for cost and treatment of diseases to improve health and quality of life of people.
- Building relationships with customers: Mining techniques are used by the researchers in the medical field to understand customer's behavior and their needs depending on the basis of various patterns. This helps in building better relationships with the customers. Customers also feel satisfied, and it creates a huge impact in rising up of the healthcare industry.
- Improvised treatment techniques: With the advancement in artificial intelligence, systems are made intelligent by applying machine learning techniques. The systems become intelligent by learning from the past experiences. Thus, patients can easily compare which technique is the best. It helps to identify the side effects of various treatments and helps to develop smarter techniques for the treatment.
- Patient's support and care: A huge amount of data is collected and stored in the computer to improve quality and standards of living of the patients. Different models are constructed by applying different techniques on the basis of a large amount of data. This is used to make improvised decisions which promote success to healthcare industry. Thus, satisfaction level of patients is achieved.

6 Result

By applying different data miming learning techniques and conducting experiments on the given dataset, we conclude that the hybrid model has the highest accuracy. It enhances the performance of heart disease prediction model. The comparison of accuracies, sensitivity, and specificity of classification techniques is shown in Table 1.

The graphical representation between ANN, C4.5, and hybrid DT is depicted in Fig. 3.

Table 1 Comparison of different techniques

Classification technique	Accuracy (%)	Sensitivity (%)	Specificity (%)
ANN	77.40	77.40	21.7
C4.5	76.66	76.7	24
Hybrid-DT	78.14	78	22.9

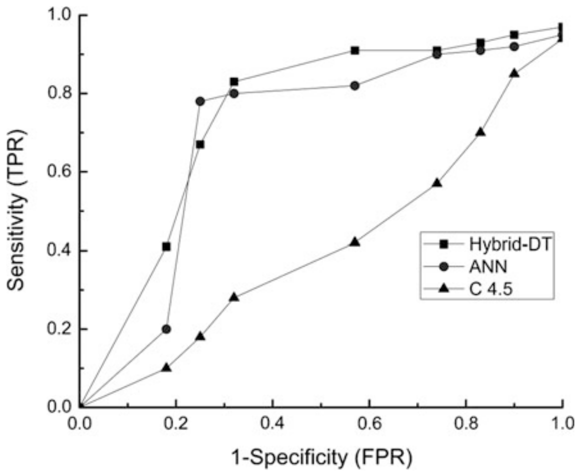


Fig. 3 Performance comparison of classification techniques

7 Conclusion

Statistical models are not useful for the prediction of diseases as these models are incapable of storing categorical values and large number of missing values. Machine learning technique (MLT) is used. It is used in combination with data mining techniques to classify different kinds of diseases based on patient's records. Accuracy is important for data mining in medical industry. Different algorithms can be applied for detecting different types of diseases. This in turn makes the system intelligent. More combinational models are constructed to predict the heart disease which can help doctors in prediction of different types of heart diseases at an early stage.

It is analyzed that hybrid DT works out the best for the given dataset with respect to the individual algorithms. Thus, it promotes better results for prediction of heart disease.

8 Future Work

In future, a Web-based application or a tool can be made using decision tree techniques for people living in rural areas to easily detect the disease. Also, the algorithms implemented can be improvised further by combing different techniques and making a hybrid system to improve the performance and efficiency of the system. Also, the algorithms can be tried on other datasets to make an expert model to enhance the results. This way different disease can be diagnosed by implementing different algorithms. Different algorithms like association mining can also be applied.

References

1. Chandna, D.: Diagnosis of heart disease using data mining algorithm. Int. J. Comput. Sci. Inf. Technol. **5**(2), 1678–1680 (2014)
2. Nikhar, S., Karandikar, A.M.: Prediction of heart disease using machine learning algorithms. IJAEMS **1–2**(6) (2016)
3. Shetty, A., Naik,C.: Different data mining approaches for predicting heart disease. Int. J. Innov. Res. Sci. Eng. Technol. **5**(Special Issue 9) (2016)
4. Tomar, D., Agarwal, S.: A survey on data mining approaches for healthcare. **5**(5) (2013)
5. Dangare, C.S., Apte, S.S.: Improved study of heart disease prediction system using data mining classification techniques. Int. J. Comput. Appl. **47**(10), 44–48 (2012)
6. Yanwei, X., Wang, J., Zhao, Z., Gao, Y.: Combination data mining models with new medical data to predict outcome of coronary heart disease. In: Proceedings of the International Conference on Convergence Information Technology, pp. 868–872 (2007)
7. Patel, J., Upadhyay, T., Patel, S.: Heart disease prediction on using machine learning and data mining techniques. IJCSC (2016)
8. Guru, N., Dahiya, A., Rajpal, N.: Decision support system for heart disease diagnosis using neural network. Delhi Bus. Rev. **8**(I) (2007)
9. Patel, S.B., Yadav, P.K., Shukla, D.P.: Predict the diagnosis of heart disease patients using classification mining techniques. IOSR J. Agric. Vet. Sci. (IOSR-JAVS) **4**(2) (2013)
10. Nishara Banu, M.A., Gomathy, B.: Disease forecasting system using data mining methods. In: International Conference on Intelligent Computing Applications (2014)
11. Sharma, N., Sharma, P.: A review-heart disease prediction using data mining technique. IJRASET **2**(IV) (2014)
12. Ratnam, D., Hima Bindu, P., Mallik Sai, V., Rama Devi, S.P., Raghavendra Rao, P.: Computer-based clinical decision support system for prediction of heart diseases using Naïve Bayes algorithm. Int. J. Comput. Sci. Inf. Technol. (2015)
13. Nikhar, S., Karandikar, A.M.: Prediction of heart disease using machine learning algorithms. IJAEMS **2**(6) (2016)
14. Kumari, M., Godara, S.: Comparative study of data mining classification methods in cardiovascular disease prediction (2017)
15. Shouman, M., Turner, T., Stocker, R.: Using data mining techniques in heart disease diagnosis and treatment, pp. 173–177 (2012)

Green Communication for Wireless Cooperative Networks: A Survey

Sonika Pahuja and Poonam Jindal

Abstract For next-generation wireless networks, radio frequency (RF) energy transfer and harvesting emerged as a viable solution to the energy-constrained nodes. In this paper, comprehensive literature review on recent research progress in wireless-powered cooperative network (WPCN) with RF energy harvesting capability is presented. First, overview of green communication with distinct classification and protocols is given. Further, comparison of half-duplex (HD) and full-duplex (FD) relaying methods is presented. Finally, open research challenges are envisioned.

Keywords Energy harvesting (EH) · Relaying protocols · Throughput
Wireless-powered cooperative network (WPCN)

1 Introduction

There is a tremendous increase in wireless communication devices since last decade [1]. In 2012, different sectors of information and communication technologies (ICTs) consumed about 4.7% of world's total electrical energy [2]. Therefore, saving energy is quite essential for enhancing the lifetime of nodes in wireless networks. Energy-constrained nodes in cooperative communication require recharging or replenishment after certain time interval. To overcome this limitation, wireless-powered cooperative communication network (WPCN) has emerged as a promising solution. The concept of simultaneous wireless information and power transfer (SWIPT) enables the system to harvest energy from radio frequencies

S. Pahuja (✉)
Department of Electronics and Communication, HMRITM, Hamidpur, New Delhi, India
e-mail: sonika.pahuja@gmail.com

P. Jindal
Department of Electronics & Communication, National Institute of Technology Kurukshetra, Kurukshetra, Haryana, India
e-mail: poonamjindal81@nitkkr.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_46

(RFs) broadcasted by energy transmitters to charge the batteries. This is quite desirable where the nodes are not reachable easily and recharging/replacing batteries incurs economical overheads. Hence, relay node is capable to harvest energy from RF signals. This is first investigated in [3] where energy harvested amplifies and forward (AF) relay is analyzed. Among different energy harvesting (EH) protocols, time switching-based relaying (TSR) and power splitting-based relaying (PSR) are analyzed for delay-limited and delay-tolerant transmission mode. In WPCN, relay node acts as wireless-powered node and it divides the received signal into two parts in time, power, or spatial domain. Among which, one of the part is used for energy harvesting and another part of received signal is used for information processing.

In the literature, there exist two types of energy harvesting relays/jammers in WPCN: half duplex (HD) and full duplex (FD). In HD or two-phase communication, the total time allotted to relay is divided between power transfer (PT) and information transmission (IT) time. Therefore, time allocation between PT and IT must be carefully designed to improve the throughput.

This paper presents the diverse protocols of green communication for cooperative networks and further investigates the usage and comparison of half-duplex and full-duplex WPCN considering single or multiple relays with single antenna and multiple antenna.

The rest of the paper is outlined as follows: Section 2 presents the concepts and distinct protocols of green communication. Section 3 explains half-duplex (HD) and full-duplex relaying used for EHN and comparison based on protocol, methodology used along with system model consideration. Section 4 focuses on future directions and open research issues, and conclusion is drawn in Sect. 5.

2 Green Communication

In cooperative communication, relay node acts as a helper node for relaying the confidential information between main nodes or jamming the information to the eavesdropper. Therefore, relay node either behaves as helper node or friendly jammer (FJ). However, relay node is energy constrained and replacing or recharging the batteries periodically may be hazardous or inconvenient sometimes (especially for the nodes located at remote area or sensors embedded within human body). As compared to conventional energy sources like solar, thermal, wind, radio frequency (RF) energy is upcoming solution for such cases. RF signal carries information and energy at the same time, thus providing energy to relay nodes.

There are various methods of green communication used for wireless networks. Figure 1 classifies distinctly the EH protocols and EH jammers used for cooperative network. As relay nodes in WPCN either act as helper node or behave as jammer for confidential information transmission and reception. The system architecture, techniques, circuit implementation, and protocols used in radio frequency energy harvesting networks (RF-EHN) are well demonstrated [4], also focusing on other

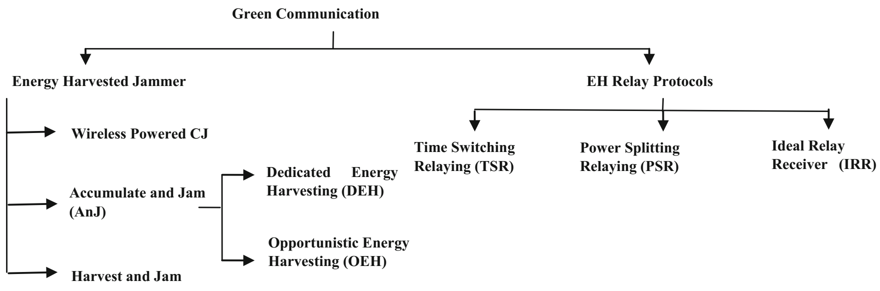


Fig. 1 Classification of green communication methods

design issues. For RF-EHN, three receiver architecture designs are discussed: time switching receiver (TSR), power splitting receiver (PSR), and integrated relay receiver (IRR). For evaluation of RFEH circuits, matching circuits are used to investigate the performance [5] with different circuit elements.

(A) EH jammers:

As depicted in Fig. 1, there are different approaches/methods of EH jammers used for cooperative networks. Full-duplex (FD) multiple antennas FJ without embedded power supply but deployed with energy harvesting unit and rechargeable energy storage are discussed in [6] which proved to be better performer over conventional HD counterpart. Two modes dedicated energy harvesting (DEH) and opportunistic energy harvesting (OEH) are taken into account. Source sends dedicated energy-bearing signal, and jammer performs energy harvesting in DEH mode, while in OEH information signal is sent by source to destination and simultaneously wireless-powered jammer transmits jamming signal using harvested energy to confound eavesdropper. Liu et al. [7] proposed two-phase communication protocol for wireless-powered cooperative jammer where energy is harvested by signal transmitted by source. High SNR and antenna regime are considered for increasing throughput with fixed rate transmission. FD information source capable of self-energy recycling, powered by energy source for transmitting confidential information, is given in [8]. Optimal solution is achieved by maximum achievable secrecy rate (ASR) at information receiver and minimum required energy harvesting at energy receiver. SRM is achieved by jointly designing energy and information beamforming for secrecy rate enhancement of proposed model.

(B) EH relay protocols:

EH-AF and EH-DF protocol are proposed [9] where relay nodes harvest energy from RF signals of source by power splitting receiver (PSR). For maximizing ASR, best relay is selected at destination. Exact and asymptotic SOP is used as performance metric where EH-DF protocol outperforms EH-AF protocol. Salem et al. [10] studied EH-based AF multi-antenna, HD relay network with EH relaying protocols: TSR, PSR, and IRR in presence of passive eavesdropper. ESC for all EH relaying protocol is derived, and PSR protocol performs better as compared to TSR is concluded. Hoang et al. [11] considered N intermediate nodes behaving as energy harvesting nodes with TSR protocol equipped with energy harvesting circuitries. Two schemes are verified: one by selecting best relay for forwarding information with jammer random and another with best jammer selection for sending jamming signals while relay is random. It is observed that system performance degrades if the information processing time is less as compared to time required for energy harvesting.

Energy-constrained relay-assisted communication between source and destination using energy harvesting relaying protocols: TSR and PSR are adopted in [3]. To analyze the throughput of both the protocols, delay-limited and delay-tolerant transmission modes are used taking outage probability and ergodic capacity as performance parameter. In terms of throughput at relatively low SNR and high transmission rate it is shown that TSR outperforms PSR protocol.

In WPCN, many factors affect the secret transmission of information, viz. distance between (1) source and eavesdropper node (d_{se}), (2) relay node and eavesdropper node (d_{re}), (3) source and relay node (d_{sr}), (4) relay and destination node (d_{sd}), path loss coefficient, time splitting, and power splitting ratios. For single hop cooperative network with TS and PS EH protocol using AF and DF is analyzed in [12] Fig. 2 depicts the variation of distance between relay and eavesdropper node (d_{re}) with respect to the secrecy rate of WPCN using DF and AF full-duplex relay (FDR) without EH. Figures 3 and 4 represent the usage of two EH protocols: TSR and PSR, respectively. It is clear that by increasing d_{re} secrecy rate of the cooperative network improves.

In the literature, there exist two ways of designing receiver: the ideal receiver design where information and power transfer are done simultaneously at relay node and another practical receiver design where information decoding and energy harvesting are done separately. This paper focuses and compares the both HD and FD relaying methodologies used in RF-EHN. Green communication is applicable

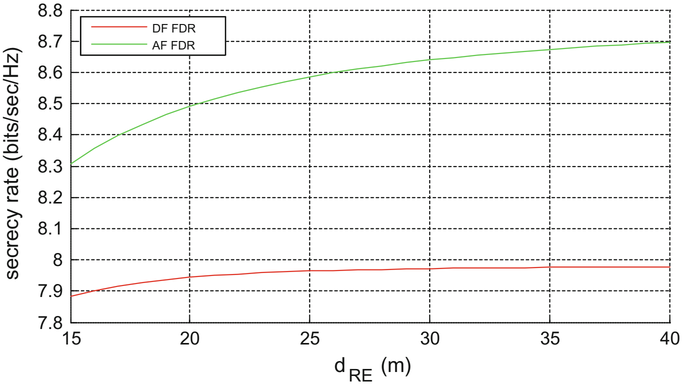


Fig. 2 Effect of d_{RE} on secrecy rate for conventional system without EH (d_{SR} = 10 m, d_{RD} = 15 m) [12]

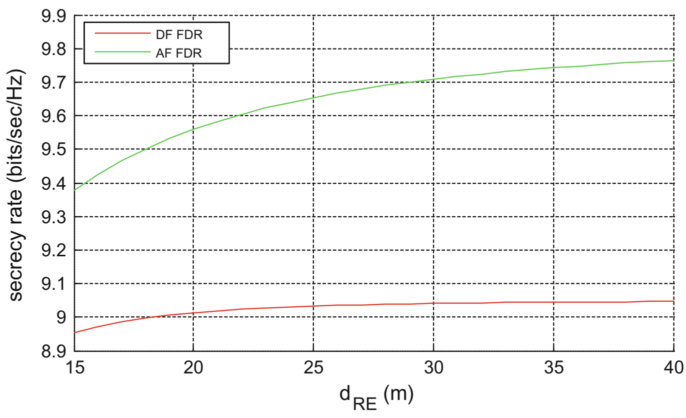


Fig. 3 Effect of d_{RE} on secrecy rate for system with TS EH (d_{SR} = 10 m, d_{RD} = 15 m, d_{BS} = 7 m, and d_{BR} = 7 m) [12]

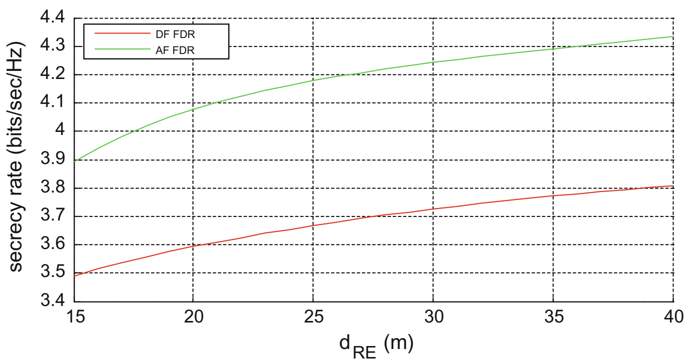


Fig. 4 Effect of d_{RE} with secrecy rate for system with PS EH (d_{SR} = 10 m, d_{RD} = 15 m, d_{BS} = 7 m, and d_{BR} = 7 m) [12]

nowadays for many areas in wireless networks, viz. unmanned aerial vehicles, cognitive radio, 5G networks [13–15].

3 Half-Duplex (HD)/Full-Duplex (FD) Relaying

In the literature, mainly HD communication network exists where relay node can either receive data or transmit data at a time but FD doubles the spectral efficiency as transmission and reception over same frequency band are enabled simultaneously. Due to simplified system design of HD relaying, this architecture is mainly used in traditional wireless systems but resulting in significant spectral loss. Therefore, FD relaying received attention of the researchers.

Comparison of FD relaying methods and protocols used is depicted in Table 1. The system model considered in all the methods considered single source node and single destination node. The number of antennas used at source, destination, and relay nodes is also represented.

FD jammer provided with energy harvesting unit and rechargeable energy storage is proposed in [6]. Here, relay node behaves as friendly jammer (FJ) and helps in transmission of secret information. Two modes are proposed dedicated energy harvesting (DEH) and opportunistic energy harvesting (OEH). Liu et al. [16] investigated power splitting for FD relay network to optimize minimum outage probability. It is shown that proposed partial channel state information (CSI) and full CSI power splitting scheme outperform fixed power splitting scheme. Optimization of both information and energy beamforming vectors for FD self-energy recycling WPCN is proposed in [8]. To maximize secrecy rate of the system, partial Lagrange dual method with two-stage optimization is used. Self-energy recycling is proposed by [17], and without the need of time switching and power splitting factor, how uninterrupted information transfer occurs. Zhong et al. [18] provided analytical expression of throughput for three communication modes: instantaneous mode, delay transfer, and delay-constrained mode using dual hop AF, DF relay.

Methods used in HD relaying are explained in Table 2. All methods used in this scheme use single source and destination node in WPCN. Best relay selection scheme for maximizing achievable secrecy rate (ASR) using EH-DF and EH-AF protocols is proposed in [9]. Golden selection search method is used to optimize the power splitting ratio of both methods. Other parameters affecting the performance of system are number of relay nodes, distance between relay and eavesdropper, noise at RF baseband unit, and energy conversion efficiency.

Two-phase communication protocol for optimizing throughput of wireless-powered FJ with fixed rate transmission is well investigated in [7]. Two cases are analyzed for increasing throughput: single antenna jammer and multiple antenna jammers. Energy consumption and energy balanced cases are also considered for IT and PT blocks.

Multiple intermediate nodes are considered [11], among which relay and jammer selection is done using proposed algorithm. Ergodic secrecy capacity (ESC) for all three EH protocols is derived in [10]. Various other parameters are also investigated which effect system performance like EH time, relay location, artificial noise (AN) power, power splitting ratio, number of antennas at relay node, and EH efficiency. It is shown through simulation that in terms of secrecy capacity the PSR protocol is better than TSR. Chen et al. [19] propose harvest-then-cooperate protocol where source and relay nodes are not provided with embedded power supply; rather, they depend on signal broadcasted by hybrid access point (AP) in wireless cooperative network scenario. Zhang and Chen [20] proposed three distinct schemes for throughput maximization, viz. source wireless power transfer (S-WPT), destination WPT (D-WPT), and joint source destination WPT scheme (SD-WPT).

Table 1 Comparison of FD single/multiple relay in green communication

Relay working as	Method used	Parameter	Knowledge of CSI (main nodes)	Antennas at source node	Antennas at destination node	Antennas at relay node	Eavesdropper
Friendly jammer [6]	DEH and OEH energy accumulation and consumption procedure	SOP, probability of positive secrecy capacity	Imperfect	One	One	N	Passive, single antenna
Information source (IS) [8]	Self-energy recycling with secure beamforming	Secrecy rate maximization (SRM)	Perfect	M	One	N + 1	Potential, single antenna
EH relay [16]	Power splitting scheme	Outage probability	Perfect and partial	One	One	Two, transmit and receive	N.A.
Wireless-powered AF relay [17]	Self-energy recycling	Throughput	Perfect	M	One	N + 1	N.A.
EH relay AF, DF dual hop [18]	TSR in instantaneous transmission, delay-tolerant and delay-constrained modes	Throughput	Perfect	One	One	Two, transmit and receive	N.A.

Where *AF* amplify and forward, *CSI* channel state information, *DEH* dedicated energy harvesting, *DF* decode and forward, *N.A.* not applicable, *OEH* opportunistic energy harvesting, *SOP* secrecy outage probability

Table 2 Comparison of HD single/multiple EH relay in green communication

Relay working as	Method used	Parameter	CSI	Antennas at source	Antennas at destination node	Relay nodes and antennas	Eavesdropper
Friendly jammer (FJ) [7]	PT—power transfer and IT—information transfer	Throughput	Perfect	One	One	Single or Nj antenna	One passive, single antenna
EH relay [9]	EH-AF, EH-DF PSR protocols with best relay selection	Exact and asymptotic SOP ASR	Perfect	One	One	M trusted terminals, single	One, single active
EH-AF relay network [10]	TSR, PSR, IRR	ESC	Perfect	One	One	N antenna	One passive, single antenna
Relay and FJ [11]	DF protocol, TSR bR-rJ and rR-bJ method	Exact and asymptotic SOP	–	One	One	M + 1 nodes	Passive
AF and selection combining technique [19]	Harvest-then-cooperate (HTC)	Average throughput, time allocation, relay number, relay position	Perfect	One	N.A	One, single	N.A.
DF relay [20]	S-WPT, D-WPT, SD-WPT	Outage probability	——	One	One	Single	N.A.

Where *ASR* achievable secrecy rate, *bR-rJ* best relay and random jammer, *ESC* ergodic secrecy capacity, *N.A.* not applicable, *rR-bJ* random relay and best jammer

For multi-user environment, applicability of WPT is essential requisite for current era of exponential growth of wireless devices. Although, few work is done by researchers in WPCN for multi-user scenario.

4 Future Directions

In WPCN, open research issues are: implementation of RF-EHN in multi-user environment where scheduling schemes can be used. Multi-relay nodes can also be used where efficient relay and jammer selection algorithms improve the system performance. Moreover, transmit power for RF device is typically low. Thus, employing multiple antennas at relay node helps in improving transmission efficiency but on the other hand resulting in increased power consumption. Therefore, methods for optimizing trade-off between transmission efficiency and power consumption may be proposed. Optimal relay selection can be done based on high energy conversion efficiency. In RF-powered devices, power consumption is a critical issue; hence, redesigning the conventional schemes and protocols to optimize power consumption along with maintaining secrecy constraints of the cooperative network still requires efforts. Implementation of green communication in dynamic environment where energy harvesting rate is varying needs to be explored.

5 Conclusion

In green communication, using RF energy harvesting system, throughput improves significantly using FD relaying, as compared to HD relaying architecture. For optimal time split and power split, system throughput substantially enhances using FD relay. Moreover, potential gain is realized using channel statistics only without the knowledge of instantaneous CSI. Therefore, FD relaying is a promising solution for implementation of future wireless cooperative networks.

References

1. More than 50 billion connected devices, Ericsson, white paper (2011). http://www.akos-rs.si/files/Telekomunikacije/Digitalna_agenda/Internetni_protokol_Ipv6/More-than-50-billion-connected-devices.pdf
2. Gelenbe, E., Caseau, Y.: The impact of information technology on energy consumption and carbon emissions, pp. 1–15 (2015). https://doi.org/10.1145/2755977
3. Nasir, A.A., Zhou, X., Durrani, S., Kennedy, R.A.: Relaying protocols for wireless energy harvesting and information processing. IEEE Trans. Wirel. Commun. **12**(7), 3622–3636 (2013)

4. Lu, X., Wang, P., Niyato, D., Kim, D.I., Han, Z.: Wireless networks with RF energy harvesting: a contemporary survey. IEEE Commun. Surv. Tutor. **17**(2), 757–789 (2015)
5. Aggarwal, et al.: Exact performance evaluation of RF energy harvesting with different circuit elements. IETE Tech. Rev. 1–9 (2017)
6. Bi, Ying, Chen, He: Accumulate and jam: towards secure communication via a wireless-powered full-duplex jammer. IEEE J. Sel. Top. Signal Process. **10**(8), 1538–1550 (2016)
7. Liu, W., Zhou, X., Durrani, S., Popovski, P.: Secure communication with a wireless-powered friendly jammer. IEEE Trans. Wirel. Commun. **15**(1), 401–415 (2016)
8. Wu, W., Wang, B., Deng, Z., Zhang, H.: Secure beamforming for full-duplex wireless powered communication systems with self-energy recycling. IEEE Wirel. Commun. Lett. **6** (2), 146–149 (2017)
9. Son, P.N., Kong, H.Y.: Co-operative communication with energy harvesting relays under physical layer security. IET Commun. J. **9**(7), 2131–2139 (2015)
10. Salem, A., Hamdi, K.A., Rabie, K.M.: Physical layer security with RF energy harvesting in AF multi-antenna relaying networks. IEEE Trans. Commun. **64**(7), 3025–3038 (2016)
11. Hoang, T.M., Duong, T.Q., Vo, N.S., Kundu, C.: Physical layer security in cooperative energy harvesting networks with a friendly jammer. IEEE Wirel. Commun. Lett. **6**(2), 174–177 (2017)
12. Jindal, P., Sinha, R.: Physical layer security with energy harvesting in single hop wireless relaying system. In: Kim, K., Joukov, N. (eds.) Information Science and Applications 2017, ICISA 2017. Lecture Notes in Electrical Engineering, vol. 424. Springer, Singapore (2017)
13. Gao, H., Ejaz, W., Jo, M.: Cooperative wireless energy harvesting and spectrum sharing in 5G networks. IEEE Access (Special section on Green Communications and Networking for 5G Wireless) **4**, 3647–3658 (2016)
14. Li, K., Ni, W., Wang, X., Liu, R.P., Kanhere, S.S., Jha, S.: Energy-efficient cooperative relaying for unmanned aerial vehicles. IEEE Trans. Mobile Comput. **15**(6), 1377–1386 (2016)
15. Verma, D.K., Chang, R.Y., Chien, F.-T.: Energy-assisted decode-and-forward for energy harvesting cooperative cognitive networks. IEEE Trans. Cogn. Commun. Netw. (2017). https://doi.org/10.1109/tccn.2017.2709759
16. Liu, H., Kim, K.J., Kwak, K.S.: Power splitting for full-duplex relay with wireless information and power transfer (2015). arXiv:1504.04697v1 [cs.IT]
17. Zeng, Y., Zhang, R.: Full-duplex wireless-powered relay with self-energy recycling. IEEE Wirel. Commun. Lett. **4**(2), 201–204 (2015)
18. Zhong, Caijun, Suraweera, Himal A., Zheng, Gan, Krikidis, Ioannis, Zhang, Zhaoyang: Wireless information and power transfer with full duplex relaying. IEEE Trans. Commun. **62** (10), 3447–3461 (2014)
19. Chen, H., Li, Y., Rebelatto, J.L., Uchôa-Filho, B.F., Vucetic, B.: Harvest-then-cooperate: wireless-powered cooperative communications. IEEE Trans. Signal Process. **63**(7), 1700–1711 (2015)
20. Zhang, C., Chen, Y.: Wireless power transfer strategies for cooperative relay system to maximize information throughput. IEEE Access **5**, 2573–2582 (2017)
21. Morsi, Rania, Michalopoulos, Diomidis S., Schober, Robert: Multiuser scheduling schemes for simultaneous wireless information and power transfer over fading channels. IEEE Trans. Wirel. Commun. **14**(4), 1967–1982 (2015)

Self-training with Neighborhood Information for the Classification of Remote Sensing Images

Prem Shankar Singh Aydav and Sonajharia Minz

Abstract Acquiring labeled data is very tough and time-taking process due to various constraints present in the area of remote sensing. Supervised classification of remote sensing images requires huge amount of labeled data for the better performance. Many semi-supervised techniques have been developed and explored which require very few labeled data to train the classifiers. Self-training is a popular semi-supervised technique which trains any supervised classifiers in various iterations with the help few labeled and large pool of unlabeled samples. Traditional self-training method is not suitable for the remote sensing images because it does not utilize spatial properties of the image. In this paper, an enhanced version of traditional self-training method has been proposed which utilize spatial properties for confident sample selection. The experimental results show that the proposed method has been performed better than traditional self-training method.

Keywords Remote sensing · Semi-supervised · Self-training
Support vector machine

1 Introduction

Remote sensing is a science and art to capture information about any object or area through a device which is not in contact with that object or area. Remote sensing images are one of the very important spatial data sets captured by remote sensors of different types. It has many applications in many areas like agriculture, land use, change detection, glacier monitoring, urban growth study, forest monitoring, and environmental study. Data mining and machine learning techniques play very important role to capture interesting information from the images. Supervised,

P. S. S. Aydav (✉)
DTE, Lucknow, UP, India
e-mail: premit2007@gmail.com

S. Minz
JNU, New Delhi, India

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_47

unsupervised, and semi-supervised techniques are main categories which are widely used in analysis of remote sensing data. Supervised techniques require good amount of labeled information to classify the remote sensing images while unsupervised techniques do not need labeled data [1–3]. In the area of remote sensing, it is very hard to collect labeled information due to many constraints. In the absence of the labeled information, unsupervised techniques may be used to extract information from remote sensing images but these techniques do not provide true distribution of classes present in the data sets. On the other hand, supervised classification techniques require good amount of labeled information to train the classifiers. It is very hard to train supervised classifiers with limited amount of data. To overcome this problem, many semi-supervised techniques are explored in the area of remote sensing which require limited amount of labeled data and lot of unlabeled data to develop a classifier model [4, 5].

In the literature, many semi-supervised techniques have been defined which can be divided into three broad categories [6]. Graph-based semi-supervised techniques are first category in which a graph is created with labeled and unlabeled samples, and it tries to minimize error by assigning some label to neighboring nodes. The main assumption in graph-based semi-supervised method is that nodes connected with highest weight should share same class label [7].

In second category of semi-supervised techniques, the objective functions of the traditional supervised classifiers are modified which also take care of unlabeled samples. These techniques mainly add a regularization term based on the cluster assumption. The support vector machine (SVM) is modified in many ways to avail the benefit of the unlabeled samples. The objective of this semi-supervised SVM is to move decision boundary in low density region while keeping labeled samples correctly classified. In remote sensing area, semi-supervised SVM has been explored in many ways [8–11]. The neural network objective function is also modified to integrate unlabeled samples information to learn the weight of neural network. Semi-supervised neural network is used in many ways to classify the remote sensing images [12]. Generative models are also used in semi-supervised techniques in which it is assumed that the label and unlabeled samples comes from same distribution. The main disadvantages of second category algorithm are that while manipulating objective functions most of them become nonlinear optimization problems. In third category, traditional classifier is trained in iterative manners without changing their original objective functions. The category is known as self-labeled semi-supervised techniques. The main advantages of these algorithms are that it does not modify the original classifiers. Self-training, co-training, and committee-based semi-supervised techniques are the three leading techniques of self-labeled semi-supervised techniques [13]. Self-labeled semi-supervised techniques are also widely used for the classification of remote sensing images. Self-training is a very basic and effective self-labeled semi-supervised technique which trains a single classifier in iterative manner. In this paper, self-training technique is modified for the remote sensing images which take help from spatial neighborhood pixels to add labeled samples for the iterative training process.

Rest of the paper is organized as follows: Section 2 describes the self-training process, Sect. 3 describes the proposed methodology, Sect. 4 describes experimental results and analysis, and the last section presents the conclusion of the work.

2 Self-training

Self-training is a very effective semi-supervised approach which is widely used in many areas. It trains a supervised classifier with the help of very few labeled and pool of large unlabeled samples. It is a wrapper technique in which any classifier may integrate. The algorithm starts with training of a supervised classifier with few labeled samples, and the trained classifier is used to classify the unlabeled samples. The most confident samples from the unlabeled samples are added in labeled data samples, and this process continues in many iterations. The whole procedure is summarized in Algorithm 1. Let U is the unlabeled samples, L is the given few *labeled* samples, X is vector which contains d features, and f is the classifier which has to be learnt. Let probability of a sample X to belong a particular class c is denoted by $p(X^C)$. The proposed method starts with training of f with initial set of labeled pixels set L. The trained classifier f is used to classify the unlabeled samples U, and few confident samples having highest probability are added to the labeled pool L. The whole procedure is summarized in Algorithm 1.

Algorithm-1 ***Self- Learning***

Input: A supervised learner f, labeled data pool L, unlabeled data set U, Maximum iteration number *itr*,

Output *A supervised Learner f*

Method:

1. Train classifier f with initial labeled data set L.

Repeat

2. Apply learner f on data set U and get probability p_{ij} of each samples
3. Acquire pixels set (t_i) from each class having highest probability
4. Add sample set $(T = \bigcup_{i=1}^{w} t_i)$ to label pixel set $(L=LUT)$
5. Remove added samles T from U $(U = U - T)$
6. Train again classifier f with newly labelled sample set L

Until maximum iteration *itr*

7. Return trained classifier f

End

3 Proposed Self-training Approach

The main properties of the remote sensing images are high correlation among neighboring pixels. It is very high probability that the neighboring pixels in the images share the same class. One major problem with traditional self-training approach is that if misclassified unlabeled samples are selected as highly confident samples, then it increases noise in labeled data pool. The added misclassified samples may degrade the accuracy of the classifier. To decrease this effect, the spatial correlation properties of the image data set are utilized in the proposed model. The proposed technique measures the confidence of a pixel with integration of their neighborhood pixels. The technique does not consider only the confidence of the single pixel but also the neighborhood's information to decide the confidence of a pixel to belong in a particular class. The proposed technique removes the noisy pixels and selects some informative as well as confident samples using spatial properties of the image data sets.

Let I is the image data set, L is the limited labeled pixels set, x_{ij} is the pixel in spatial domain, and f is the classifier which has to be learn. Let probability of a pixel x_{ij} to belong a particular class c is denoted by $p(x_{ij}^{c})$. The proposed method starts with training of f with initial set of labeled pixels set L. The trained classifier f is used to calculate probability $p(x_{ij}^{c}))$ of each pixel in the image data set I. Then, in each iteration the confidence of each pixel is calculated as weighted sum neighborhood pixels probability:

$$conf\left(x_{ij}^{c}\right) = \sum_{k \in N_{ij}} w_k p\left(x_{ij}^{c}\right) \tag{1}$$

The label of each pixel is determined through maximum weighted neighborhood probability:

$$p'(x_{ij}^{c}) = \sum_{k \in N_{ij}} w_k p(x_{ij}^{c}) \tag{2}$$

$$\text{Label}\left(x_{ij}\right) = \arg(\max(p'(x_{ij}^{c})) \tag{3}$$

where $p'(x_{ij}^{c})$ is neighborhood probability of x_{ij} pixel to the cth class and N_{ij} is the neighborhoods of x_{ij} pixel. The set of confident samples selected by Eq. 1 is added to labeled data pool which is used for next iteration training. The whole process is summarized in Algorithm 2. According to spatial resolution and properties of the images, weight w_k of the pixel x_{ij} with their neighbors may be chosen in different ways. In this paper, Gaussian and average weights are considered.

Algorithm 2- *Proposed Self-Training with Neighbourhood Information*

Input: A supervised learner *f*, labeled data pool *L*, Image data set *I*, unlabeled data set *U⊆(I-L)*, Maximum iteration number *itr,*

Output: A supervised Learner f

1. Train classifier *f* with initial labeled pixels set *L*

Repeat

2. Apply trained learner *f* on image data set *I* and get probability p_{ij} of each pixels
3. Apply equation 1 to get neighbourhood probability of each pixel x_{ij}
4. Get label using equation 3 and confidence with equation 1 of samples belongs to *U*
5. Acquire pixels set (t_i) from each class having highest confidence
6. Add pixel set ($T = \bigcup_{i=1}^{w} t_i$) to label pixel set (*L=LUT*)
7. Remove added pixels T from U ($U = U - T$)
8. Train again classifier *f* with newly labeled pixel set *L*

Until maximum iteration *itr*

9. Return trained classifier *f*

End

4 Experimental Result and Analysis

To show effectiveness of the proposed method, two benchmark remote sensing images have been used in the experiments. The data sets are hyper-spectral remote sensing images known as Indian Pines and Salinas's data sets. The average weight and Gaussian weight are considered in the spatial neighborhood window for the proposed method. The probabilistic SVM is used as supervised classifier [14]. The traditional self-training method with support vector machine is abbreviated as SSVM, the proposed method with average weight is abbreviated as SSVMAVG, and the proposed method with Gaussian weight is abbreviated as SSVMGAUS. Each experiment on both data sets is performed till tenth iterations.

4.1 Salinas Data

Salinas hyper-spectral remote sensing data sets collected over the area of Salinas Valley, California. The data set contains 200 spectral bands with spatial resolution of 3.7-m pixels (Fig. 1). The experiments are performed with 16 classes of the data sets. In the experiments, ten random set of the labeled data have been randomly chosen. The size of the spatial window has been taken as 5×5. The regularization

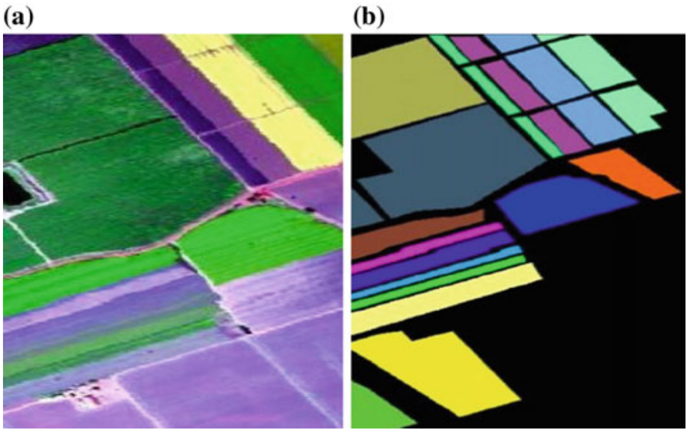


Fig. 1 **a** Salinas image, **b** ground truth

parameter and Gaussians kernel parameter of the SVM classifier have been acquired with cross-validation method. The experiments have been performed till ten iterations, and the most confident pixel have been added in the label pool from each of the respective class in every iteration. The average error graph has been shown in Fig. 2. The error graph of SSVM shows that there is reduction in error during initial iterations, and after that the error has increased slightly during the different iterations. The graph of the proposed methods SSVMAVG and SSVMGAUS shows that the error has been reduced in progress of the iterations. The comparative study of the graph indicates that the proposed method has performed very better than the

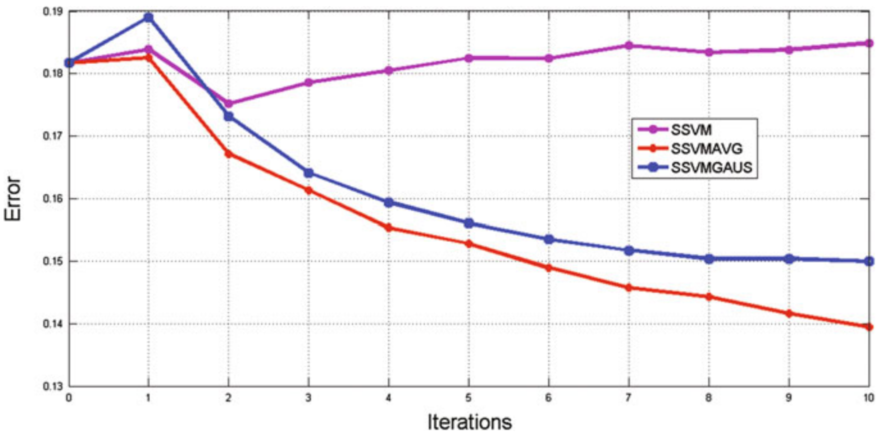


Fig. 2 Salinas error graph

Table 1 Salinas kappa coefficient and accuracy

Data sets	SVM		SSVM		SSVMGAUS		SSVMAVG	
	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy
1	0.7979	0.8173	0.7803	0.8019	0.8370	0.8532	**0.8437**	**0.8595**
2	0.8032	0.8236	0.7999	0.8197	0.8351	0.8514	**0.8355**	0.**8517**
3	0.8020	0.8208	0.8015	0.8204	0.8091	0.8279	**0.8306**	**0.8471**
4	0.7923	0.8135	0.8038	0.8230	0.8233	0.8406	**0.8340**	**0.8504**
5	0.7896	0.814	0.7866	0.8084	0.7967	0.8171	**0.8275**	**0.8451**
6	0.8096	0.8287	0.8316	0.8480	**0.8555**	**0.8697**	0.8518	0.8665
7	0.8028	0.8225	0.7640	0.7867	0.8473	0.8624	**0.8653**	**0.8786**
8	0.7872	0.8099	0.7950	0.8152	0.8509	0.8658	**0.8690**	**0.8822**
9	0.7933	0.8126	0.7937	0.8134	0.8589	0.8729	**0.8586**	**0.8725**
10	0.8040	0.8227	0.7947	0.8146	0.8217	0.8392	0.8356	**0.8518**
Mean	0.7982	0.8183	0.7951	0.8151	0.8335	0.8500	0.8452	**0.8605**

traditional self-training approach. In proposed methods, the error plot of SSVMAVG is always below than SSVMGAUS but both have reduction in error during iterations. Table 2 contains mean accuracy and mean kappa coefficient achieved in tenth iterations. The analysis of the results is shown in Table 1 and indicates that the proposed methods SSVMGAUS and SSVMAVG have achieved better accuracy than the SSVM. The proposed method SSVMAVG accuracy and kappa coefficient are **0.8725** and **0.8586**, respectively which is highest than all other methods.

4.2 Indian Pines Data

This hyper-spectral remote sensing image was collected in northwestern Indiana over the Indian Pines site (Fig. 3). The data set contains 200 spectral bands. The original data set contains 16 classes but the experiments have been performed with 12 major classes of the data sets. The image and its ground truth are shown in Fig. 3a, b, respectively. In the experiments, ten random set of different labeled samples are selected for the experiments. In each random set, six label pixels from each class have been selected.

The parameters of Gaussian kernel and the regularization parameter are determined with cross-validation method. In proposed method, the window size has been taken as 5 × 5. In each iteration, the most confident sample is added in the labeled data pool for every technique.

The graph shown in Fig. 4 indicates that the very less error has been reduced by the traditional method SSVM during iterations. The purposed method SSVMGAUS

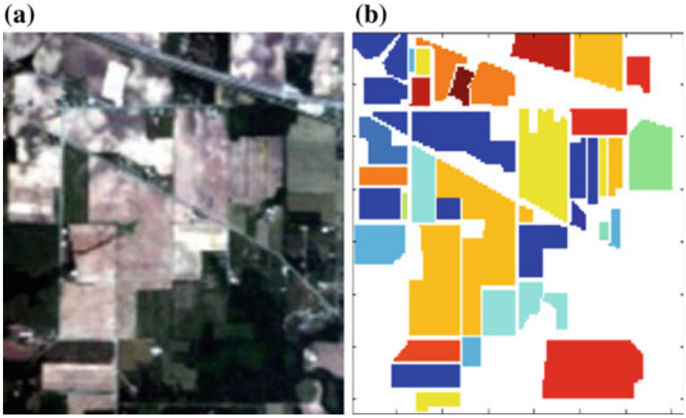


Fig. 3 a Indian Pines image, b ground truth of Indian Pines

has reduced error in first few iterations, and then it has been saturated for further iterations. The proposed method SSVMAVG has performed very better than other two methods. The method has reduced error in almost all iterations with very high rate. The data shown in Table 2 also confirm that the kappa accuracy of SSVMAVG and SSVMGAUS is better than SSVM. The highest accuracy and kappa coefficient is achieved by the purposed method SSVMAVG which are 0.5414 and 0.4844, respectively.

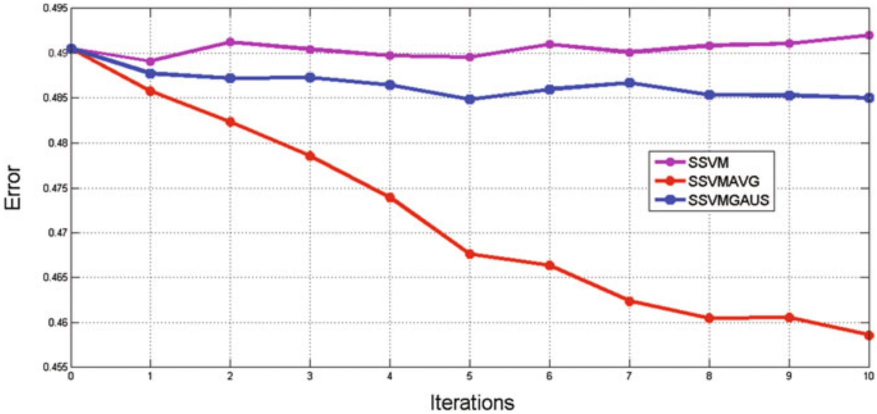


Fig. 4 Error graph of Indian Pines

Table 2 Indian Pines kappa coefficient and accuracy

Data sets	SVM		SSVM		SSVMGAUS		SSVMAVG	
	Kappa	OA	Kappa	OA	Kappa	OA	Kappa OA	
1	0.4884	0.5419	0.5115	0.5652	0.5044	0.5600	**0.5297**	**0.5844**
2	0.4337	0.4894	0.4373	0.4928	0.4495	0.5051	**0.4766**	**0.5280**
3	0.4657	0.5238	0.4403	0.4994	0.4577	0.5172	**0.5206**	**0.5745**
4	0.4054	0.4655	0.4082	0.4694	0.4150	0.4765	**0.4436**	**0.5012**
5	0.4493	0.5114	0.4565	0.5160	0.4430	0.5071	**0.4753**	**0.5322**
6	0.4530	0.5118	0.4451	0.5049	0.4616	0.5214	**0.4727**	**0.5317**
7	0.4398	0.5004	0.4232	0.4853	0.4261	0.4882	**0.4665**	**0.5279**
8	0.4184	0.4856	0.4178	0.4852	0.4191	0.4873	**0.4578**	**0.5205**
9	0.4661	0.5249	0.4615	0.5217	0.4803	0.5388	**0.4984**	**0.5535**
10	0.4823	0.5400	0.4782	0.5402	0.4902	0.5483	**0.5027**	**0.5603**
Mean	0.4502	0.5095	0.4480	0.5080	0.4547	0.5150	**0.4844**	**0.5414**

5 Conclusion

In this paper, an improved self-training method has been proposed for the semi-supervised classification of the remote sensing images. The method assumes the image data set as pool of unlabeled samples with their spatial relationship. The proposed method takes help of neighboring pixels to decide the confident samples which are added into labeled data pool. The proposed method has capability to handle noisy pixels, and it also selects the informative samples. The experimental results show that the proposed method has performed better than traditional self-training method.

References

1. Gong, B., Im, J., Mountrakis, G.: An artificial immune network approach to multi-sensor land use/land cover classification. Remote Sens. Environ. **115**(2), 600–614 (2011)
2. Pal, M., Mather, P.M.: An assessment of the effectiveness of decision tree methods for land cover classification. Remote Sens. Environ. **86**(4), 554–565 (2003)
3. Aydav, P.S.S., Minz, S.: Generalized fuzzy c-means with spatial information for clustering of remote sensing images. In: 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC), pp. 1–5. IEEE (2014)
4. Aydav, P.S.S., Minz, S.: Semi-supervised learning for the classification of remote sensing images: a literature review. Adv. Comput. Sci. Inf. Technol. 4, 10–15
5. Persello, C., Bruzzone, L.: Active and semisupervised learning for the classification of remote sensing images. IEEE Trans. Geosci. Remote Sens. **52**(11), 6937–6956 (2014)
6. Zhu, X.: Semi-supervised learning literature survey. Comput. Sci. (University of Wisconsin-Madison) **2**(3), 4 (2006)
7. Subramanya, A., Talukdar, P.P.: Graph-based semi-supervised learning. Synth. Lect. Artif. Intell. Mach. Learn. **8**(4), 1–125 (2014)

8. Chapelle, O., Sindhwani, V., Keerthi, S.S.: Optimization techniques for semi-supervised support vector machines. J. Mach. Learn. Res. 9(Feb), 203–233 (2008)
9. Chi, M., Bruzzone, L.: Semisupervised classification of hyperspectral images by SVMs optimized in the primal. IEEE Trans. Geosci. Remote Sens. **45**(6), 1870–1880 (2007)
10. Chi, M., Feng, R., Bruzzone, L.: Classification of hyperspectral remote-sensing data with primal SVM for small-sized training dataset problem. Adv. Space Res. **41**(11), 1793–1799 (2008)
11. Gómez-Chova, L., Camps-Valls, G., Munoz-Mari, J., Calpe, J.: Semisupervised image classification with Laplacian support vector machines. IEEE Geosci. Remote Sens. Lett. **5**(3), 336–340 (2008)
12. Ratle, F., Camps-Valls, G., Weston, J.: Semisupervised neural networks for efficient hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. **48**(5), 2271–2282 (2010)
13. Triguero, I., García, S., Herrera, F.: Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study. Knowl. Inf. Syst. **42**(2), 245–284 (2015)
14. Wu, T.F., Lin, C.J., Weng, R.C.: Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. 5(Aug), 975–1005 (2004)

Instance Selection Using Multi-objective CHC Evolutionary Algorithm

Seema Rathee, Saroj Ratnoo and Jyoti Ahuja

Abstract Data reduction has always been an important field of research to enhance the performance of data mining algorithms. Instance selection, a data reduction technique, relates to selecting a subset of informative and non-redundant examples from data. This paper deals with the problem of instance selection in a multi-objective perspective and, hence, proposes a multi-objective cross-generational elitist selection, heterogeneous recombination, and cataclysmic mutation (CHC) for discovering a set of Pareto-optimal solutions. The suggested MOCHC algorithm integrates the concept of non-dominating sorting with CHC. The algorithm has been employed to eight datasets available from UCI machine learning repository. The MOCHC has been successful in finding a range of multiple optimal solutions instead of yielding a single solution. These solutions provide a user with several choices of reduced datasets. Further, the solutions may be combined into a single instance subset by exploiting the promising characteristics across the potentially good solutions based on some user-defined criteria.

Keywords Multi-objective optimization · CHC algorithm · Instance selection KNN

S. Rathee (✉) · S. Ratnoo
Guru Jambheshwar University of Science and Technology, Hisar, India
e-mail: seema27rathee@gmail.com

S. Ratnoo
e-mail: ratnoo.saroj@gmail.com

J. Ahuja
Government Post Graduate College for Women, Rohtak, India
e-mail: kwatra.jyoti@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_48

1 Introduction

In real-life situations, data is not clean enough to be processed straight away and most often it is accompanied by noise, inconsistencies, and redundancies. Data reduction is one of the important data cleaning tasks to make data mining processes effective. Data reduction process involves the selection of more informative and relevant features (feature selection) or instances (instance selection) or both (dual selection) [1–3]. This study exclusively addresses the problem of instance selection (IS) in context to K-nearest neighbor (KNN) classifier.

KNN is one of the most popular and widely used classification algorithms. But it is not without limitations. The performance of KNN algorithm is susceptible to the value of the parameter K, and the sheer size of data makes KNN unsuitable for many problems due to the scalability issues. Therefore, we need to scale down the training data by removing the redundant instances. Instance selection based on prototype selection (PS) is a method for scaling up KNN classifier to deal with large datasets.

As instance selection in databases can be visualized as a search problem, it can be worked out using evolutionary meta-heuristics. Cano et al. [1] conclude that, out of the several variations of genetic algorithms, cross-generational elitist selection, heterogeneous recombination, and cataclysmic mutation (CHC) perform best for instance selection. It is a non-conventional genetic algorithm which puts together an elitist selection, an efficient recombination operator, and a restart process on reaching immature convergence.

Garcia et al. [2–8] have proposed many methods to deal with the scalability issue but deviations in accuracy and reduction rates are still there. Both factors are in disagreement because maximizing data reduction causes a decrease in accuracy and vice versa. Therefore, prototype selection should be considered as a multi-objective optimization problem.

The CHC like any other GA converges to a single best solution and does not deal with the problem of competing objectives. This paper suggests a multi-objective CHC (MOCHC) to address the issue of instance selection for the task of classification. The suggested algorithm finds a Pareto-optimal set of solutions (subsets of instances) instead of a single solution, and the user can have several choices of reduced instance subsets according to their preference for higher accuracy or higher reduction rate. We have used KNN classification algorithm for the task of classification. Here, we have been successful to have produced multiple prototypes based on multi-objective criteria, i.e., accuracy and reduction rate. To the best of our knowledge, hybridization of MOGA with CHC is the first attempt toward instance selection.

Rest of the paper is structured as follows. Section 2 presents the related literature in context to the proposed work. Section 3 describes the essential background for multi-objective optimization. Section 4 explains the proposed algorithm. Section 5 gives experiment design and results. The last section concludes the paper.

2 Literature Review

Prototype selection is a fertile field of research and encloses a vast literature. Most commonly used PS algorithms are based on NN rules and aim at improving the effectiveness of KNN classifier. Some of them are condensed nearest neighbor (CNN), edited nearest neighbor (ENN), reduced edited NN (RENN), reduced NN (RNN), model class selection system (MCS), Shrink, IB2, and IB3. There are some other ordered removal-based methods also, i.e., Drop1, Drop2, Drop3 [9].

Authors have also successfully applied evolutionary algorithms to resolve the problem of instance selection [1, 10–13]. A survey of evolutionary algorithms in instance selection [14] confirms the superiority of GAs as these provide maximum classification performance by efficient exploration of solution space to give a single best solution. The CHC has shown very promising results in the domain of instance selection [1, 10, 11]. Details of the peculiarities of the CHC algorithm are given in [1].

Subsequently, considering instance selection as a multi-objective optimization problem, some authors have also worked on multi-objective counterparts of GAs for instance selection [12, 15]. MOGAs have been recognized in the last decade as excellent methods to explore and find an approximation to the Pareto-optimal front for data reduction problems. It would be interesting to investigate the performance of multi-objective version of CHC as it (CHC) is being considered as the most powerful GA for instance selection.

3 Multi-objective Optimization

When, there are conflicting objectives, no single solution is adequate that maximizes simultaneously all of the objectives. Thus, the solutions that provide a quality trade-off among the objectives are to be investigated. Pareto optimality offers a method to settle the tradeoffs among the Pareto-optimal solutions. It is to be noted that solution **A** dominates solution **B** (denoted by $\mathbf{A} > \mathbf{B}$) if and only if **A** is not inferior to **B** in any of the objectives, and there must be at least one objective with respect to **A** is superior. If a solution is not dominated by any other solution in the solution space, it is known to be Pareto optimal. The Pareto-optimal set consists of all the Pareto-optimal solutions. The curve passing through the Pareto-optimal solutions is called the Pareto front [16].

Multi-objective evolutionary algorithms have gained popularity because these work with a population of diverse solutions that converge to a set of Pareto-optimal solutions. A large number of EAs have been proposed for solving problems that involve multiple and conflicting objectives. These include non-dominated sorting genetic algorithm II (NSGA-II) [17], strength Pareto evolutionary algorithm 2 (SPEA2), Pareto archived evolution strategy (PAES). Out of these, NSGA-II is the frequently applied multi-objective genetic algorithms. In this paper, we have hybridized CHC and the concept of Pareto-based non-domination sorting from NSGA-II to address the problem of instance selection. The suggested algorithm tries to simultaneously optimize data reduction rate and accuracy of KNN classifier.

4 Multi-objective CHC Algorithm for Instance Selection: MOCHC

This section describes the design of MOCHC for instance subset selection. The whole MOCHC procedure combines the non-dominated sorting and replacement strategy of NSGA-II (a well-known MOGA) [17] and CHC genetic algorithm as shown in Fig. 1.

The non-dominated sorting and replacement strategy of NSGA-II provide the multi-objective environment for the two distinct objectives, i.e., accuracy and reduction rate. The CHC algorithm integrates a conservative selection mechanism and a highly disruptive crossover operator that maintains diversity and conserves the better fit individuals in the population. The main challenge of the MOCHC is to defend elitism as well as diversity with respect to the conflicting objectives (accuracy and reduction rate) in the GA population.

In the proposed design, the most significant modification concerns the evaluation of chromosomes and replacement strategy. The chromosomes are evaluated on the basis of two objectives and sorted using non-dominated sort. Then, CHC runs its self-controlled phases with the committed objectives by combining elitism and HUX (a different crossover operator) method to enhance the effectiveness of the algorithm.

The old GA population is swapped using replacement strategy of NSGA-II. Accordingly, two successive (parent and child) populations are merged and the non-dominated solutions falling on Pareto fronts of different levels are selected to form new population for the next generation. If the new population size is bigger than the pre-specified population size, then chromosomes from the last added front are selected one-by-one based on the crowding distance. With the progressive evolution of the population, it starts converging and the number of individuals satisfying the incest condition rises. In other words, the GA population is stalled, and then a restart process is initiated by applying bit-flip mutation operator with very high probability. The restart process keeps a percentage of the highly ranked individuals intact and applies mutation to the remaining. The chromosome representation, fitness function, and genetic operator employed are described below.

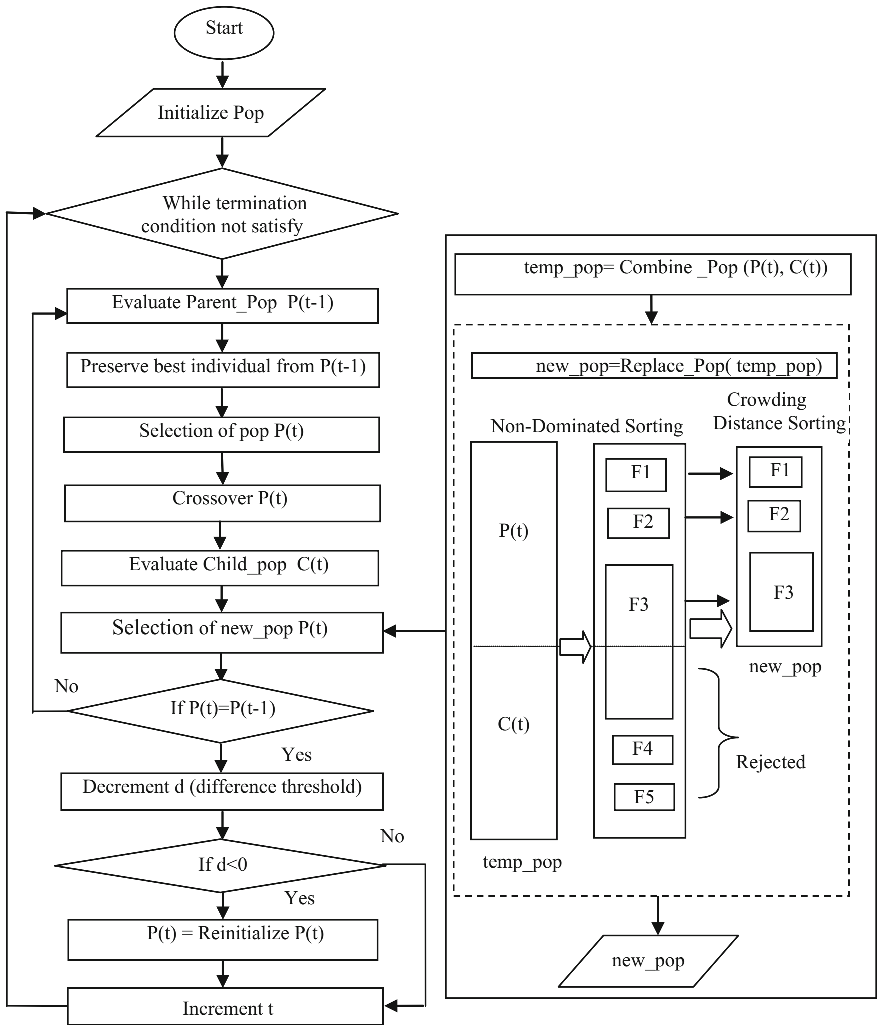


Fig. 1 Flowchart of MOCHC

4.1 Chromosome Representation

The initialization of individuals in the GA pop is made randomly. Each chromosome represents an instance contained in a dataset. A chromosome corresponds to a sequence of the digits, 0’s and 1’s. The bit value 0 signifies the absence of an instance, while the bit value 1 indicates the presence of an instance.

4.2 Fitness Function

A multi-objective approach considers each objective as a separate goal. Ideally, we would like to simultaneously maximize accuracy as well as reduction rate. In reality, it is not possible to find such an ideal solution since the objectives are in disagreement with each other. In multi-objective perspective, both the objectives are dealt independent of the other. The fitness function considers the two objectives, f1 being the predictive accuracy of classifier and f2 is the percentage data reduction as given below.

$$\mathrm{f1} = \mathrm{Accuracy}$$

$$\mathrm{f2} = 100 \times (|\mathrm{DS}| - |\mathrm{RS}|)/|\mathrm{DS}|$$

where DS and RS are the cardinalities of the original and the reduced datasets.

4.3 Genetic Operators

The CHC uses selection with elitist strategy, HUX crossover operator, and simple bit-flip mutation. The HUX crossover swaps half of the bits which differ in the parent individuals. The bit positions to be swapped are randomly selected. CHC genetic algorithm does not apply mutation operator during recombination phase like other genetic algorithms. It is used with a high probability (0.35 usually) only when CHC gets stuck in sub-optimal regions of search space.

5 Experimental Design and Result Analysis

The proposed MOCHC has been implemented in MATLAB environment. The other experimental details are given below.

5.1 Description of the Datasets

The proposed MOCHC is applied to eight datasets available from UCI machine learning repository. These contain diverse number of attributes and instances. Table 1 summarizes all the datasets.

Table 1 Description of the datasets

Sr. no.	Datasets	#Instances	#Features	#Class
1.	Yeast	1484	8	10
2.	Segment	2310	19	7
3.	Hypothyroid	3772	29	4
4.	Page-blocks	5473	10	5
5.	Optdigits	5620	64	10
6.	Mushroom	5644	22	2
7.	Letter	20,000	16	26
8.	Shuttle	58,000	9	7

5.2 Genetic Parameters

The parameter setting is an important issue for any evolutionary algorithm, and it influences the performance of stochastic algorithms. The parameters used for MOCHC are: the chromosome length as per the number of instances in a dataset, population size = 50, number of generations = 100, mutation threshold = 0.35, difference threshold = one-fourth of length of chromosome, preserved population (no. of elite kids) = 10%, convergence limit = 10 generations.

5.3 Results

Due to lack of space, it is not possible to present all the non-dominated solutions for all the datasets. Hence, we have depicted the non-dominated solutions for the *Segment* and *Page-blocks* datasets only. The resulting Pareto-optimal solutions for these datasets are given in Table 2.

The solutions reflect a trade-off between the two objectives, accuracy, and reduction rate. Figure 2 shows the non-dominated solutions and their respective Pareto fronts in a graphical manner. A user can select any instance subset from the Pareto-optimal solution as per certain requirements or preferences. Yet a single

Table 2 Pareto solutions in case of segment and page-blocks dataset

Sr. no	Segment dataset		Page-blocks dataset	
	% Accuracy (f1)	% RR (f2)	% Accuracy (f1)	% RR (f2)
1.	62.5000	52.5974	90.7063	47.1267
2.	63.5593	51.9481	91.6716	45.1865
3.	64.5455	45.4545	90.6475	48.1468
4.	74.7664	44.3723	92.3704	44.1259
5.	71.75	45.1321	94.0369	44.1169

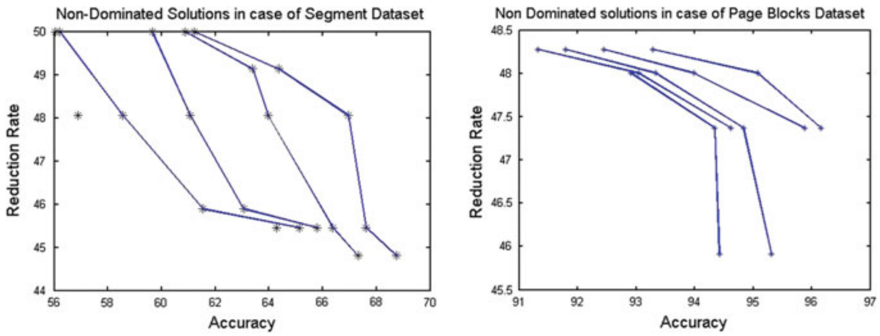


Fig. 2 Pareto solutions for *segment* and *page-blocks* datasets

instance subset can also be constructed by combining common instances falling in many of the Pareto-optimal solutions.

After all, the solutions are optimized for one or the other objectives. Now the question is that from which optimal solutions the common instances may be combined to construct a single subset of instances? This may be decided by defining a threshold for accuracy and reduction rate. One may choose to combine the common instances from the solutions that have accuracy and reduction rates above the given thresholds. We have taken these thresholds to be 60% and 0.45, respectively. The accuracy and reduction rates obtained after the application of MOCHC on various datasets are briefed in Table 3. The percent accuracy obtained by running KNN on the complete datasets is given in the first column of Table 3.

The proposed algorithm has achieved more than 50% data reduction rate across all the datasets. A comparison of accuracy obtained by KNN on full datasets with the accuracy obtained by applying KNN on the reduced datasets reflects no significant decrease in the accuracy in the later case. There is only one dataset (segment) where accuracy decreases from around 89% to around 80%. The differences

Table 3 Results of datasets for MOCHC algorithm

Datasets	KNN on whole dataset	KNN on reduced datasets by MOCHC	
	%Accuracy	% Accuracy	% RR
Yeast	97.2973	96.1988	53.3784
Segment	88.9610	80.4971	54.5022
Hypothyroid	92.4138	92.4528	55.0398
Page-blocks	95.5726	94.8220	54.7715
Optdigits	98.7989	97.5950	54.0214
Mushroom	95.7890	94.9846	54.0071
Letter	99.3375	98.0628	55.0300
Shuttle	99.6802	99.6356	54.5749

for the rest of the datasets are insignificant. The results are encouraging and interesting since despite a reduction rate of more than 50% there is no compromise on the accuracy.

6 Conclusion

Selecting non-redundant and relevant instances is essential for enhancing the performance of classification algorithms. In fact, redundant and irrelevant instances cause confusion for learning algorithms which may lead to poor predictive accuracy and comprehensibility. This paper has proposed a multi-objective CHC (MOCHC) to address the problem of data reduction, specially, instance selection in a multi-objective framework. The proposed approach has successfully discovered a range of Pareto-optimal solutions as an alternative of a single optimal solution. The main contribution of this paper is to combine the concept of non-dominating sorting and replacement of NSGA-II, the most frequently used multi-objective GA, with CHC, a genetic algorithm that has given the best results so far for the instance subset selection problem. We have also suggested a strategy to combine the instances from various Pareto-optimal solutions to give a single subset of instances. The results are encouraging across almost all the datasets with high accuracy and significant reduction rates.

In future, the results need to be confirmed on more datasets and we intend to apply MOCHC for simultaneous reduction of instances and features for the problem of dual selection.

References

1. Cano, J.R., Herrera, F., Lozano, M.: Using evolutionary algorithms as instance selection for data reduction in KDD: an experimental study. IEEE Trans. Evol. Comput. **7**(6), 561–575 (2003)
2. de Haro-García, A., García-Pedrajas, N.: A divide-and-conquer recursive approach for scaling up instance selection algorithms. Data Min. Knowl. Disc. **18**(3), 392–418 (2008)
3. García-Pedrajas, N., de Haro-García, A., Pérez-Rodríguez, J.: A scalable approach to simultaneous evolutionary instance and feature selection. Inf. Sci. **228**, 150–174 (2013)
4. Derrac, J., Triguero, I., Garcia, S., Herrera, F.: Integrating instance selection, instance weighting, and feature weighting for nearest neighbor classifiers by co-evolutionary algorithms. IEEE Trans. Syst. Man Cybern. B Cybern. **42**(5), 1383–1397 (2012)
5. Garcia-Pedrajas, N.: Constructing ensembles of classifiers by means of weighted instance selection. IEEE Trans. Neural Netw. **20**(2), 258–277 (2009)
6. García-Osorio, C., de Haro-García, A., García-Pedrajas, N.: Democratic instance selection: a linear complexity instance selection algorithm based on classifier ensemble concepts. Artif. Intell. Rev. **174**(5–6), 410–441 (2010)
7. García, S., Luengo, J., Herrera, F.: Instance selection. In: Data Preprocessing in Data Mining, pp. 195–243. Springer International Publishing, Switzerland (2015)

8. Derrac, J., García, S.: Herrera, F.: Stratified prototype selection based on a steady-state memetic algorithm: a study of scalability. Memet. Comput. **2**(3), 183–199 (2010)
9. Olvera-López, J.A., Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Kittler, J.: A review of instance selection methods. Artif. Intell. Rev. **34**(2), 133–143 (2010)
10. Cano, J.R., Herrera, F., Lozano, M.: Strategies for scaling up evolutionary instance reduction algorithms for data mining. In: Ghosh, D.A., Jain, P.L.C. (eds.) Evolutionary Computation in Data Mining, vol. 163, pp. 21–39. Springer, Berlin, Heidelberg (2005)
11. Cano, J.R., Herrera, F., Lozano, M.: Stratification for scaling up evolutionary prototype selection. Pattern Recogn. Lett. **26**(7), 953–963 (2005)
12. Rosales-Perez, A., Garcia, S., Gonzalez, J.A., Coello, C.A.C., Herrera, F.: An evolutionary multi-objective model and instance selection for support vector machines with Pareto-based ensembles. IEEE Trans. Evol. Comput. **99**, 1–1 (2017)
13. Cano, J.R., Herrera, F., Lozano, M.: On the combination of evolutionary algorithms and stratified strategies for training set selection in data mining. Appl. Soft Comput. **6**(3), 323–332 (2006)
14. Derrac, J., García., S., Herrera, F.: A survey on evolutionary instance selection and generation. Int. J. Appl. Metaheuristic Comput. **1**(1), 60–92 (2010)
15. Li, J., Wang, Y.: A nearest prototype selection algorithm using multi-objective optimization and partition. In: 2013 Ninth International Conference on Computational Intelligence and Security, pp. 264–268 (2013)
16. Kim, M., Hiroyasu, T., Miki, M., Watanabe, S.: SPEA2+: improving the performance of the strength Pareto evolutionary algorithm 2. In: Edmund, X.Y., et al. (eds.) Parallel Problem Solving from Nature—PPSN VIII, vol. 3242, pp. 742–751. Springer, Berlin, Heidelberg (2004)
17. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

Inclusive Device to Connect People with the Surrounding World

Vishal B. Pattanashetty, Nalini C. Iyer, H. L. Viswanath and Shweta Kore

Abstract The recent innovation in IoT is controlling of devices around us at tip of the finger. The existing IoT devices are used in controlling and monitoring of industrial equipment's, home automation, health care, etc. The advancement in this field of IoT is that the device has to be compatible and secure enough to control and monitor different parameters using a single device. The existing systems are such that one application can control one single application only. This project aims to design a product that integrates different systems into a single device and that can be controlled. This device is capable of configuring and controlling any end device required. This product configures the end device based on the required application. The device is portable, and it works on plug-and-play mechanism; the device has to be plugged to the end device that has to be monitored and controlled, and for instance, it can be used to control the speed of the fan, lights in homes. This product can also be used in industries to monitor the parameter like temperature. The system makes use of a software application to control and configure the end device.

Keywords IoT · WSN · Connected cars · IoT project · Zigbee

V. B. Pattanashetty · N. C. Iyer
B.V. Bhoomaraddi College of Engineering & Technology, Hubballi, Karnataka, India
e-mail: vishalbps@gmail.com

N. C. Iyer
e-mail: nalinic@bvb.edu

H. L. Viswanath
Dayanand Sagar Engineering College, Bengaluru, Karnataka, India
e-mail: hl_viswa@hotmail.com

S. Kore (✉)
J.S.S College, Dharwad, Karnataka, India
e-mail: shwethavishalp@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_49

1 Introduction

Internet of things is a socio-technical phenomena which allows physical entities, vehicles, or additional things rooted with microelectronics or electronics, real-world sensors and link connectivity that allows this object to gather and interchange the facts. It is predictable that by 2020 tens of billions of things will be deployed worldwide. IoT devours stridden out of its roots besides its on the edge of changing the contemporary stagnant Internet to a totally established imminent Internet [1]. Today Internet has become one of the important parts of our lives. In the present day, everything is getting smarter like the devices around us. The evolution in the IoT is visualizing the world where the several objects can sense, communicate, and share information. IoT is trivial since an object that can embody itself digitally becomes improved than the object itself. The data is then allied to the surrounding objects and record data. The present devices are such that one application can control any single parameter like a fan or a light, and there is no heterogeneity. Hence, there is a requirement for the device which can alone control and monitor multiple parameters. We have analyzed different applications and developed a system which controls the different parameters using a single application, and also, the data sensed by the system is stored in the database. This acts like a hardware middleware, which connects the device with the surrounding world. Consequently, there is a much necessity to apprehend in what way the current IoT-based middleware structures work and address the diverse requirement and overriding the existing issues and gaps [2, 3]. This can be achieved by developing a mobile application and a webpage. Web page acts as a database which stores the sensed data sent from the device, and the controlling of the end device can be done based on the data sent to the webpage.

2 Motivation

IoT systems were developed to provide knowledge to an amateur user. IoT devices get compactable to different environments, so it is able to address many heterogeneous devices. The existing devices are such that one single application can control only a single device. The product "Inclusive Device" is a portable device and connects the device with different parameters and also configures the end device so that the device can be compatible with the web page. Our proposed solution aims to overcome the challenges of IoT.

3 Problem Definition

To design a low cost, reliable, user-friendly, portable universal plug-and-play device to connect and coordinate people with various real-time objects through wireless sensor networks and Internet of things.

4 Application

Some of the applications of this project are

- A single device configures and controls the multiple parameters by the mobile application.
- The sensed data of temperature, pressure, or any other parameter is stored in the database through a Wi-Fi shield.
- The device is portable and of plug-and-play mechanism, and it has to be connected to the end parameter that has to be controlled.
- The whole system can be controlled by a mobile application.
- The sensed data by the system is stored in database which can be used further.

5 One Hypothesis, Many Visions

The project requires mainly a mobile application to control the parameter from the remote place, a suitable wireless technology, and a best-suited microcontroller. The survey was made on the best-suited methods required by the project.

5.1 Wireless Technology

The wireless technology had overtaken all the previous technologies which were working with the help of wires. The versatility of technologies has caused to diversify many other fields. There are many technologies which work on wireless, and they work to compete in the race of technological world. In this project, the wireless technology is needed to send the data that is sensed to the database. So out of many technologies, we made comparison between Bluetooth module and the Wi-Fi shield ESP8266. As Bluetooth does not suffice the need of IoT, Wi-Fi module is used to transfer the data to the database. Hence, we come to the conclusion that ESP8266 is best suited for the transmission of the data for this project.

5.2 Controlling/Configuring

Important part of IoT-based concept is controlling the different parameters from the remote places. The controlling can be done by any of the software in the PC or it can be controlled by a mobile application. As nowadays almost everyone is well verse in using the smartphones, the best approach to this is controlling by a web page. The mobile application configures the end device via Arduino so that it can be controlled.

5.3 Core/Processor

As we know that every system needs the controller for the different processing functions and taking decisions based on the data obtained by the sensor. In this project, the microcontroller acts as a mediator between the web page and the hardware. The controller which is choose has to be compatible to flexible programming and easy to use, Arduino board is one such type of controller which is low cost, high speed, and reliable. It makes use of serial communication that is it can be connected to computer with simple USB cable. The selection of Arduino is because it uses ATmega microcontroller and provides the number of libraries to make the programming easier. It uses the Harvard architecture wherein two separate memories, flash and SRAM memory, that is program-code and program-data, respectively. Program-code is stored in flash memory, and program-data is stored in SRAM memory. The Harvard architecture is best suited in embedded system. It makes the complicated process easier.

6 Methodology of Inclusive Device

The three main units required for building up the system are end device, middleware, and control/configure. To achieve the Wi-Fi connectivity for Arduino, ESP8266 module is chosen suitable. The block diagram gives a description of how the system is designed and the data flow.

6.1 End Devices

End devices are the devices which are configured and controlled from anywhere via a mobile application or a Web server. For instance, under home automation lights, machines, electrical outlets, heating and cooling structures are bowed up to a manageable network. End devices are configured using mobile application as well as the Arduino microcontroller. End devices consist of actuators and relays. The project emphasis on two applications, controlling the ON and OFF operation of fan at home, depending on the temperature of data which is sensed through LM35 temperature sensor as well as controlling the interior lights of home. The user has an option to configure the end device to either control fan or control the interior lights of home or both of them.

6.2 Middleware

This section is similar the middleware of IOT, and it binds the mobile application to the hardware which is an end device in this case. Middleware consists of an Arduino, Wi-Fi shield. The Wi-Fi shield provides the Wi-Fi connectivity to the Arduino so that it is capable of receiving the instructions sent by the mobile application to control the end device. Specific Arduino pins are made high/low depending on the requirement of the end device via mobile application, the data which is send to the respective Arduino pin with the help of Wi-Fi shield. LM35 is connected to pin 2 of Arduino which acts as input and the fan (motor) is connected to pin 5 of Arduino which acts as an output. The pin which enables the relay to control the fan is currently made LOW, when the need is not the control the fan but some other end application, the pin has to be made HIGH (configure the end device) to control the fan.

6.3 Control/Configure

The main requirement in the IoT concept is the user-friendly mobile application which is used to control the end applications from the remote places. In order to sense the temperature data and keep a record of it, the database is created using one free hosting site. Arduino code is written such that temperature data is sensed for every 1 s, and the data is dumped to the database by linking to the __ site. The mobile application is designed such that configuration of the end devices as well as controlling of interior home lights and fan can be done easily.

Figure 1 shows different sub-blocks among the three different main unit blocks and how different units of the project are linked. The temperature is sensed by a sensor (LM35). The sensed data is sent to the web page through Wi-Fi shield (ESP8266) and stored in the database to store the record of temperature the database is created from 1 free hosting Web site. Steps to create the database are shown below.

Register and Login to the Web site, to create a new database, go to hosting, and create a new account. Go to index page of free hosting site and select the .php my admin. Click on enter “phpmyadmin,” upload the html code in file manager. Click on the process php file, and page appears on the screen whenever the sensed data in

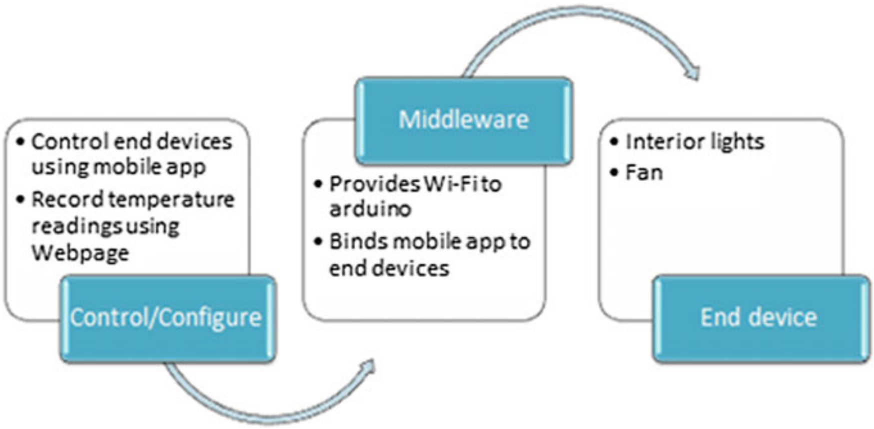


Fig. 1 Three main blocks of the system

linked with the database. The data sensed is obtained and stored on the web page through the Wi-Fi shield.

The ESP8266 is linked with Arduino as explained above. In order to connect the Wi-Fi to Arduino, configuration of ESP8266 has to be done. The serial monitor window of Arduino is used to check the connection of ESP. The Wi-Fi shield has to be configured to set up the communication, and the steps to configure ESP8266 are

The AT commands are sent through the serial monitor to ESP8266. Then the firmware version is done, AT+GMR, and then press OK and get the operation mode. The above steps imply that there is stand-alone with access point mode. Then list all the available Wi-Fi access points. Connect to the Wi-Fi network and IP address is checked. The mobile application is created using android studio. The application is linked with Ethernet.

A web page is created where the temperature data that is sensed using LM35 is stored in the database via a Wi-Fi module. Based on the temperature data obtained, the fan is controlled that is the fan is turned ON or OFF.

7 Result

The address of our webpage is www.inclusivedevice.tk that we use for hosting purpose. Many number of ID can be crated for getting Login. As we Login, the main dashboard will appear which contains temperature data in the form Celsius by using the temperature sensor LM35, configurable pins (D1, D2, and D4) or output pins and current status of all outputs. The temperature input is given to one of the

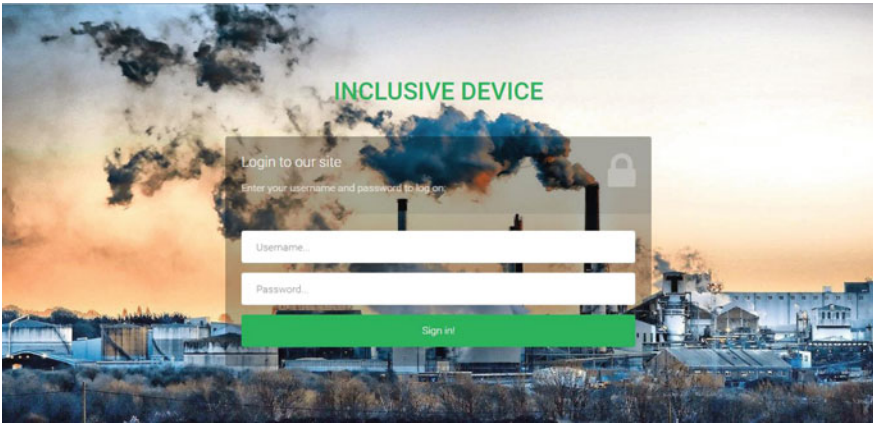


Fig. 2 Screenshot of front view of webpage

digital pins of Arduino. The temperature data gets stored in the database for every 3 s, and whenever the temperature increases beyond the threshold point, the user can turn ON fan automatically from the webpage. Using other three output pins, we can connect n number of end devices such as medical, industrial, home automation, which will be controlled using the web page. The response time from web page to the end devices will take less than 1 s (Figs. 2, 3, 4, and 5).

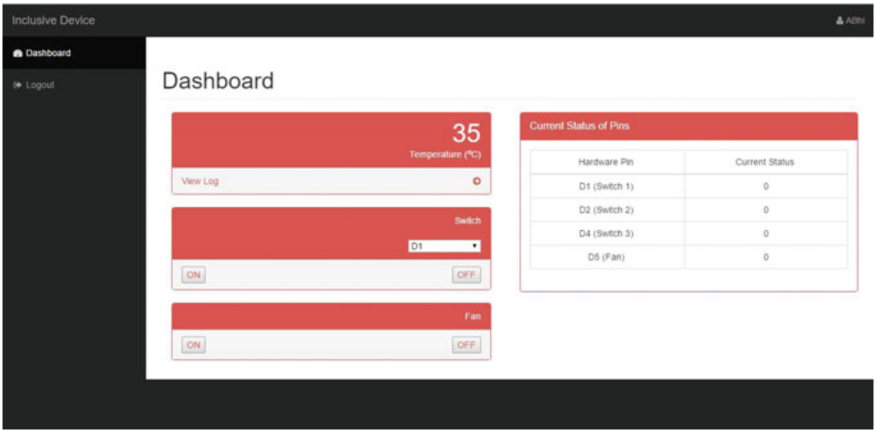


Fig. 3 Dash board view of webpage

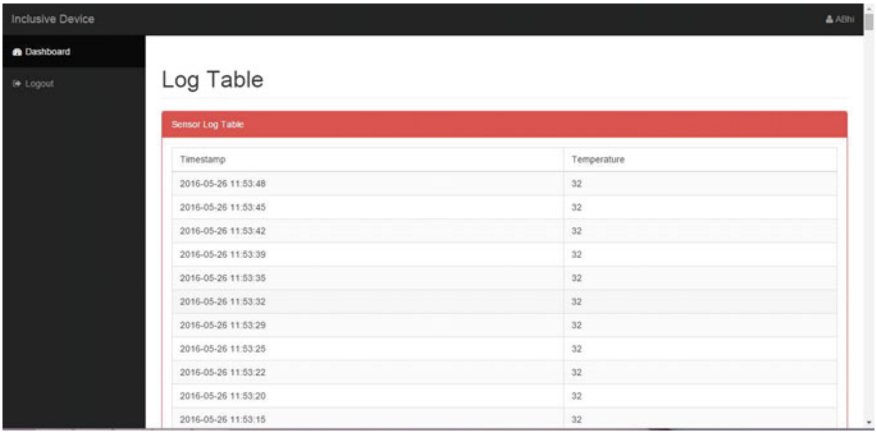


Fig. 4 Database of Temperature sensor

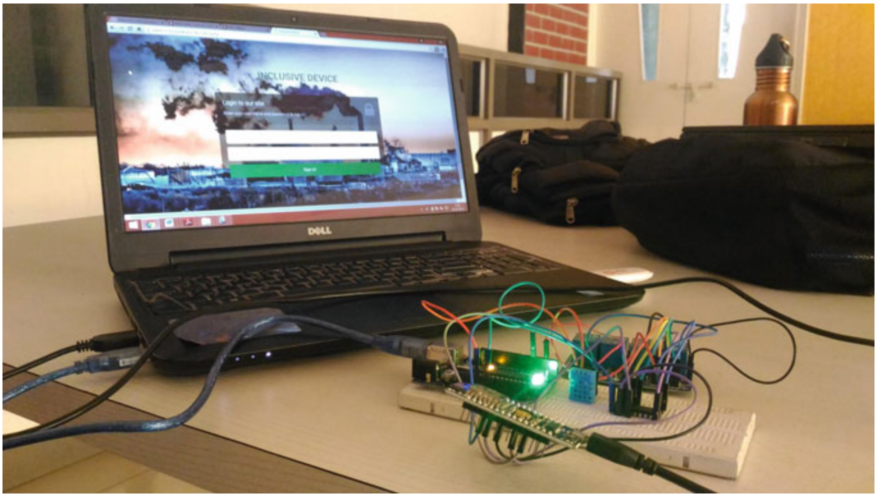


Fig. 5 Overall simulation

8 Conclusion

IoT which is the fast-growing technology is a key for the development of product design and manufacture. There are numerous opportunities for improvement in the field of Internet of things. Identity and privacy are the critical challenges which the IoT has to overcome. The project emphasize on developing an inclusive device to connect people with the surrounding world to configure end devices and control only when there is a requirement, and this method overcomes the chances of someone hacking and trying to control our devices. The project will be a leading edge of innovation among the wider Internet of things.

The Internet of things (IoT) is a fast growing conception that defines a future where everyday physical things will be linked to the Internet and be capable to categorize themselves to other devices. Further, the device can be designed such that it configures all the devices and almost everything will be interconnected such as cows to plants, kids to toys, doors to security, cars to insurance agency, and limbs to healthcare system. Everything can be interred linked.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**, 393–422 (2002)
2. Liu, D.-L. Y. F., Liang, Y. -D.: A survey of the internet of things. In: The 2010 International Conference on Electronic-Business Intelligence (ICEBI) (2010)
3. Bandyopadhyay, S., Sengupta, M., Maiti, S., Dutta, S.: A survey of middleware for internet of things. In: CoNeCo 2011, Ankara, Turkey, 26–28 June 2011
4. Cooper, J., James, A.: Challenges for database management in the internet of things. In: IETE Technical Review, vol. 26, no. 5, pp. 320–329 (2009)
5. Henricksen, K., Robinson, R.: A survey of middleware for sensor networks: state-of-the-art and future directions. In: International Work-shop on Middleware for Sensor Networks, pp. 60–65. Melbourne, Australia, Nov 2006
6. Caceres, R., Friday, A.: Ubicomp systems at 20: progress, opportunities, and challenges. IEEE Pervasive Comput. **11**, 14–21 (2012)
7. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**, 2787–2805 (2010)
8. Belissent, J.: Getting Clever About Smart Cities: New Opportunities Require New Business Models. Forrester Research (2010)

Profiling-Based Effective Resource Utilization in Cloud Environment Using Divide and Conquer Method

Namra Bhadreshkumar Shah, Neel Dhananjaybhai Shah, Jitendra Bhatia and Harshal Trivedi

Abstract "The Cloud" in IT perception is simply a storage which is used to store data persistently through the resources that are available virtually. Computing on the cloud (cloud computing) means storing and accessing data over the Internet (the Cloud). The cloud computing is also regarded as a type of Internet-based computing. Effective resource utilization under cloud framework is one of the prime issues. Load balancer's goal is to provide resources to serve incoming requests as well as fair resource allocation. Its job is to distribute the workload across the available computing resources. Auto-scaling basically provides flexibility with the number of available servers. Most of the load balancing algorithms suffer from starvation due to lack of utilization statistics. In this work, we incorporated profiling of resources in order to utilize them efficiently. Experimental results show that proposed algorithm outperforms traditional load balancing algorithm by 10–20%.

Keywords Cloud computing · Load balancing · Auto-scaling Resource utilization · Divide and conquer

---

N. B. Shah (✉) · N. D. Shah · J. Bhatia
Department of Computer Engineering, Vishwakarma Government Engineering College, Nr. Visat Three Roads, Sabarmati-Koba Highway, Chandkheda, Ahmedabad 382424, Gujarat, India
e-mail: namrashah5@gmail.com

N. D. Shah
e-mail: neel1040@gmail.com

J. Bhatia
e-mail: jitendrabbhatia@gmail.com

H. Trivedi
SoftVan, IIM Road, Ahmedabad 390015, India
e-mail: Harshal@softvan.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_50

1 Introduction

Cloud uses virtualization and automation to provide users its services at a very low cost and at infinite scale [1]. A cloud is a data center or multiple data centers made up of compute and storage resources connected through a network [2].

Cloud computing is the delivery of variegated services over the Internet which here is considered as "the Cloud." These services comprise of databases, servers, software, networking, storage, and many more [3]. These services are delivered on pay-per-use terms. That means a user needs to pay only for those services that one uses and not the other ones. These services are very cost effective. The user does not have to worry about the charges due to the fact that they are very cheap as they are served over the Internet "on demand" rather than a necessity in real world [4]. A cloud network is a combination of multiple computers working together in such a way that they appear to be just one enormous computer resource to a casual observer. It eliminates the need of upfront infrastructure cost. Cloud computing providers offer their services through various models which are called as cloud computing models. There are three types of models, they are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) [5] (Fig. 1).

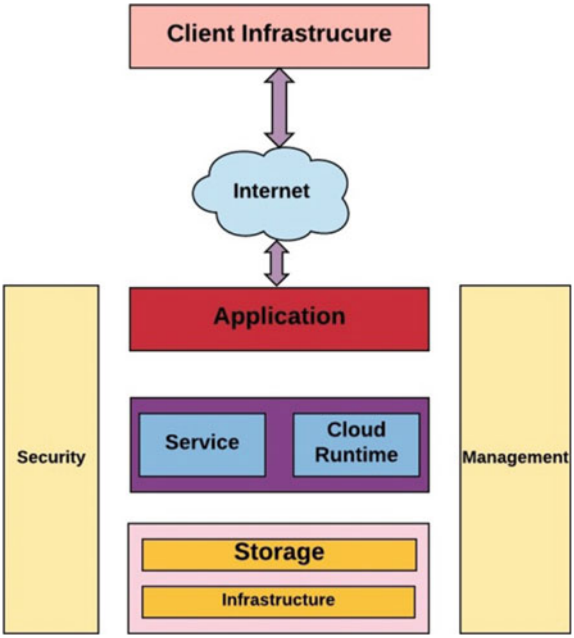


Fig. 1 Cloud computing architecture [6]

Load balancing has always been a challenging task in cloud computing. It plays a very important role in cloud computing. Most of the load balancing algorithms suffer from a common problem of starvation, whereas some of these algorithms which overcome the starvation problem fail to use the servers efficiently. The proposed solution is a step in the right direction in solving the problem of load balancing along with the proposed auto-scaling services which will enhance the Quality of Service (QoS) provided through Internet [7, 8].

Auto-scaling basically provides flexibility with the number of available servers dynamically and also statically in some cases which in turn to no surprise help to enhance the performance of an application drastically [9].

There are many challenges of load balancing but only the important ones are mentioned here. The proposed model gives a solution of these challenges in the later section. Hence, it is important to get somewhat prior knowledge of these parameters.

1.1 Response Time

The amount of time taken by the load balancing algorithm to respond to a request is defined as response time. This parameter should be reduced to enhance performance

1.2 Scalability

It is the ability to scale the servers up or down in case the number of users increases or decreases. The Quality of Service (QoS) should not be compromised.

1.3 Performance

It is to determine how efficiently the system is performing. It must be enhanced in order to provide better quality.

1.4 Fault Tolerance

Despite a server failure, the ability of a system to carry out all requests and to balance the load uniformly is called fault tolerance in this particular context.

Many latest research papers have proposed dynamic load balancing algorithms. Some of them are reviewed in the literature review section along with some of the modern static algorithms. A comparison is made between these dynamic algorithms, and a table is generated based on the outcome of these algorithms to the challenges of load balancing discussed earlier. Some loopholes are shed light to in research gap section. Thereafter, a model is proposed to overcome these challenges and also some new features are included in proposed model section. The proposed model works on the "divide and conquer" approach which will be evaluated in later sections. By using this approach, the cost of the whole process will get reduced significantly.

2 Related Work

There are two kinds of load balancing algorithms in cloud computing, namely static and dynamic. The dynamic algorithms are very well suited for the load balancer due to the fact that these kind of algorithms keep into account the prior set of data, for instance, the performance or the processing power, run-time parameters are also kept in mind before assigning the requests and hence can assign request to the correct server/virtual machine. In contrast, the static algorithms keep track of only the former information and not the run-time parameters. Hence, the resources may get mismanaged and load may not get distributed perfectly.

2.1 Dynamic Compare and Balance Algorithm (DCABA)

This load balancing algorithm decreases the number of servers to support a concept named green computing [10]. The load of the virtual machine that is removed is distributed among all the other servers that are available at that time. It is important to note that a server is only removed when the load on that virtual machine is less than the lower threshold. On multiplying the weight limits of the host machine, we can evaluate upper and lower threshold values. In contrast, if the load on the server is higher than the upper threshold value then its load is distributed among all the other virtual machines in the network. The challenges faced are the selection of migration policies and standards and also the expropriate threshold values. The other limitation is that this algorithm does not scale up the number of nodes or number of servers. It just scales down or adjusts the load between the available virtual machines.

2.2 Dynamic and Integrated Resource Scheduling (DAIRS)

This algorithm calculates the CPU load, network bandwidth, and memory of a server for evaluating the load on that particular machine [11]. This algorithm maintains four different types of queues. These are the waiting queue, request queue, optimizing queue, and delete queue. The waiting queue maintains all the incoming requests that are not allocated any serving machine immediately but are actually waiting for one to be allocated. The request queue contains all the requests that are new. The requests that need to be reallocated are maintained under optimizing queue. And finally, the delete queue contains all the requests whose end time is not known yet. Requests are served from these queue based on the priority level. The waiting queue is having the highest priority, then comes the request queue, then the optimal queue, and finally the delete queue.

2.3 Ant Colony Optimization (ACO)

In ACO, first a head node is selected. The selection of head node depends upon the number of neighbor it has. This is because more the neighbors more the traversal of cloud network. If the head node fails, another head node is selected [12]. This traversal of cloud network is done to keep track of underload and overloaded node. A table is maintained called the pheromone table which contains information about a node's resource utilization. Forward movement in forward direction is made to either underloaded or overloaded node. In backward movement, if the ants encounter an overloaded node they traverse back to the last underloaded node encountered and check if it is still an underloaded node if so all the traffic is routed to this node. This movement helps in maintaining the pheromone table which keeps track of node's various fields. The best path is guaranteed but time to converge it is uncertain. While using system with many nodes, traversal time of ants is sometimes larger than task's execution time. In that case this algorithm renders useless.

Table 1 Dynamic algorithms pros and cons

Algorithm	Advantage(s)	Limitation(s)
DCABA	Green computing is buttressed	Quality of Service is not up to the mark Network bandwidth is consumed in a large proportion
DAIRS	Takes into account the CPU utilization, memory, and network bandwidth for evaluating load on the server	Determining threshold value is a big problem
HBFA	Less VMs are migrated Energy consumption is reduced	Quality of Service is not good
ACO	Easy to understand	Time consumed is very high

2.4 Honey Bee Foraging Algorithm (HBFA)

In HBFA, first an underloaded machine is found out by random scouting. Then all the tasks will be redirected to this machine until a certain threshold value is reached. Once the machine reaches its threshold value, it is called overloaded machine. Then another underloaded machine is searched by random scouting, and tasks are redirected to that new machine. This process is repeated until all tasks are executed [13] (Table 1).

2.5 Round Robin Algorithm (RRA)

In basic Round Robin algorithm, the load balancer aligns all the servers and then it assigns the request to any random server for the first time and then following requests are allocated to next servers in-line. If it encounters the last server, then the load balancer assigns next request to the first server thus maintaining the loop-like flow of request allocation [14]. The load balancer allocates the requests in a circular order to the servers. Some servers get extremely busy and some servers enjoy idleness. It is next to impossible to predict the running time of any request, and hence, this algorithm is not preferred nowadays. Weighted Round Robin algorithm tried to overcome this flaw but failed miserably due to the same reasons. Weighted Round Robin algorithm allocates weights to every node. Requests are assigned to a particular node based on the weight allocated to that specific node.

2.6 Max–Min Algorithm (MaMiA)

The load balancer firstly sorts all the incoming requests depending on the time that will be consumed by the requests. The request of the most time-consuming task will be executed prior than the request of comparatively smaller tasks. This process is performed repeatedly until no more requests are remaining. Then, the task is removed from the list and a new task is executed. The only limitation is that the time to be consumed by every request should be known in advance which is in some cases too much to ask for [15].

2.7 Min–Min Algorithm (MiMiA)

First step in the algorithm is to sort all the incoming requests based on the completion time of each request in the ascending order in the form of a queue. Then, the request with the least completion time will be served first and will be deleted from the queue. Subsequent requests will get executed as their number arrives in the queue [16]. This is a static method, and hence, it is important to know the request parameters well in advance before it can get served. This is the only drawback in this method.

2.8 Opportunistic Load Balancing Algorithm (OLBA)

This static load balancing algorithm aims at keeping every virtual machine busy all the time. Hence, it assigns a task to a node in a completely random fashion. It deals

Table 2 Comparison of various algorithms

Algorithms/ Parameters	Type	Response time	Resource utilization	Scalability	Overhead	Performance	Fault tolerance
DCABA	Dynamic	×	×	√	√	√	×
DAIRS	Dynamic	√	×	×	√	√	×
ACO	Dynamic	×	√	√	×	√	×
HBFA	Dynamic	×	×	√	×	√	×
RRA	Static	√	√	×	√	√	×
MaMiA	Static	√	√	×	√	√	×
MiMiA	Static	√	√	×	√	√	×
OLBA	Static	×	√	×	×	√	×

with the incoming requests very quickly [17]. The current execution task of a virtual machine is not calculated by this algorithm, and hence, the task will process in a slow manner. This is the only major drawback of this algorithm (Table 2).

3 Research Gap

In cloud, one of the prime issues is to handle resource efficiency and utilization along with Quality of Service (QoS). Efficient resource balancing is always a challenging task in cloud computing. There are many existing algorithms for the same but almost all the existing generic algorithms suffer from a common problem of starvation. This problem gets even worse when the workload is bursty in nature and the number of requests demanding for resources increases. Most of these algorithms also suffer from overloading. They reduce the efficiency of server which in turn affects the performance of the service. Hence, the service quality is compromised. To the best of our knowledge, so many research works surprisingly suffer when traffic is increased suddenly or an unusual event descends. Hence, as a result some tools collapse without an indication of abortion which hinders Quality of Service.

4 Proposed Model

4.1 Architecture

The approach used in the algorithm is based on the "divide and conquer" method where basically requests are divided into different queues. Also, on the infrastructure part, nodes are also divided into two parts based on various parameters which will be analyzed further (Fig. 2).

4.2 Data Center Algorithm

The algorithm to be used here runs on a divide and conquer methodology. Firstly, to begin with, an array of nodes is sorted in ascending order based on their "usage factor." The term usage factor will be calculated based on three parameters, namely

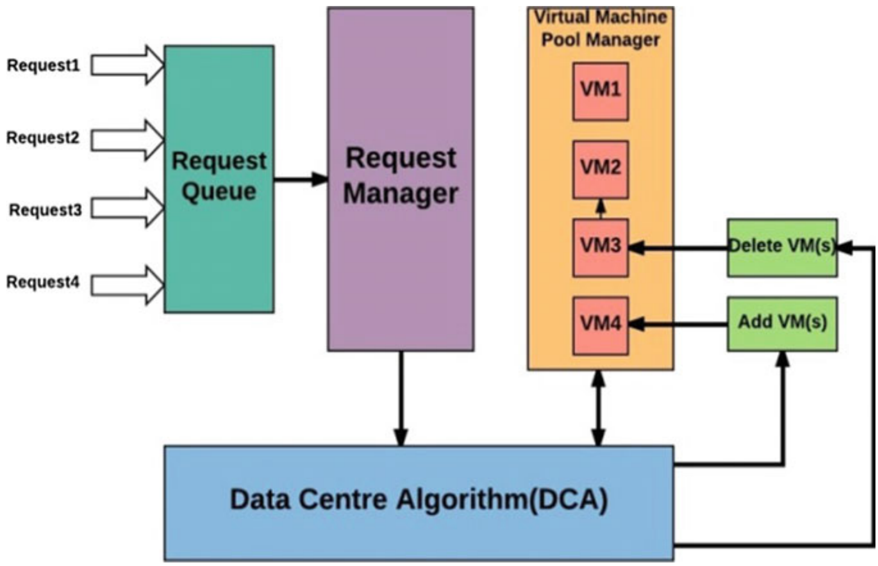


Fig. 2 Proposed architecture

CPU utilization, response time, and success rate of the node. A lookup table will be maintained to keep a track of all these parameters for each node. Here, for each node, CPU utilization and response time should be low and success rate should be high. Mathematically, usage factor for each node will be calculated as

$$\begin{aligned} usage\ factor = {} & 0.5 * cpuUtilization + 0.3 * responseTime(in\ Percentage) \\ & + 0.2 * successRate \end{aligned}$$

Here, 50% of CPU utilization at any instance of time is considered along with 30% of node's response time and 20% of its success rate that indicates successful execution of a process on that VM which is calculated from log heuristics.

First of all DCA algorithm will partition the virtual machine queue into two queues named as Zone A and Zone B based on their usage factor as mentioned above. Zone A indicates the list of underutilized resources, while Zone B contains the list of overutilized resources.

Now, firstly request will be mapped to Zone A followed by the resources listed in Zone B. Also, algorithm will check for aggregated CPU utilization, which if increased the above threshold, then auto-scaling will be done.

Pseudo code

Procedure DataCentreAlgo

Input: *cpuUtilization, responseTime, successRate,CPU_THRESH*

Output: *optimal resource utilization, Auto Scale: YES/NO*

sum← 0

set n= number of VMs // available VMs

For i=1 to n vms

If ($cpuUtilization[ith\ vm] >$ CPU_THRESH) Then

Migrate the few request on under loaded VM(s)

End if

$sum = sum + cpuUtilization[i]$

End for

if ($\frac{sum}{n} > CPU\_THRESH$) then

Call Scale_up() // adds a new VM

else

Call Scale_down() // Shutdown an existing VM

Distribute load on all available VM(S)

Calculate *usage factor*

$usage\ factor = 0.5 * cpuUtilization\ +\ 0.3 * responseTime\ (in\ Percentage) +\ 0.2 * successRate$

Sort all VM's based on *usage factor*

Assign Vms to Both the Zone

Zone_A [] ← First Half of Vms

Zone_B [] ← Second Half of Vms

Serve the Incoming Request(s) in Round Robin Fashion

End Procedure

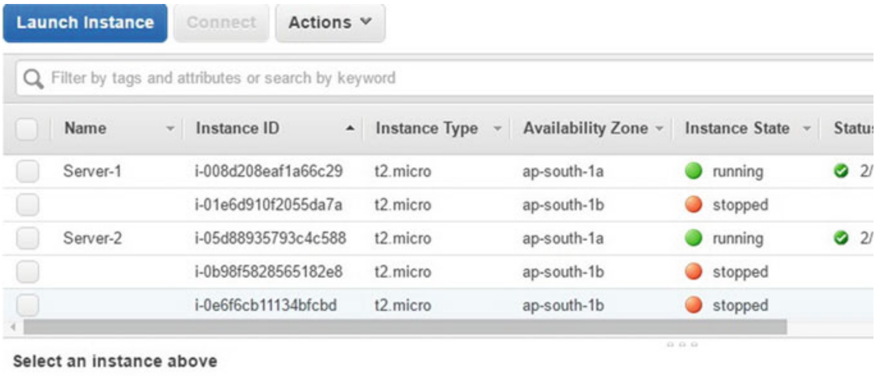


Fig. 3 AWS EC2 dashboard

5 Performance Evaluation

While performing the following experiment with the algorithm applied being mentioned above, it was found that by using the proposed algorithm based on divide and rule strategy, the performance is improved drastically and the load is evenly balanced. Moreover, the requests are properly served and the Quality of Service (QoS) was improved. Also having multiple servers and being that the load was distributed correctly between them benefited to an enormous extent.

Two AWS EC2 instances were created and kept running named Server-1 and Server-2 which are shown in Fig. 3. A Website was created for test purpose and the request made was served by one of the two servers based on the request parameters mentioned in the algorithm. The load was balanced using the data center algorithm (DCA) among these two servers for providing quick and efficient response. The Website is shown (Table 3).

Result shows that DCA algorithm outperforms over conventional load balancing algorithm in terms of an efficient resource utilization. Graph clearly shows that overall resource utilization is more with compare to traditional algorithm (Figs. 4, 5, and 6).

Table 3 Comparison of various algorithms

Parameter	Value
Number of AWS EC2 instances	5
Number of AWS EC2 running instances	2
Number of requests made	1000–5000 http requests
Instance type	Micro

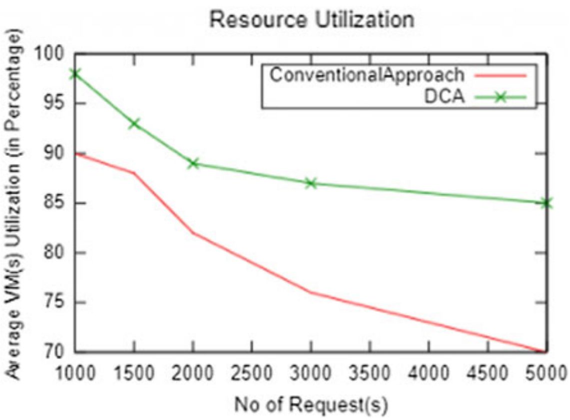


Fig. 4 Analysis graph

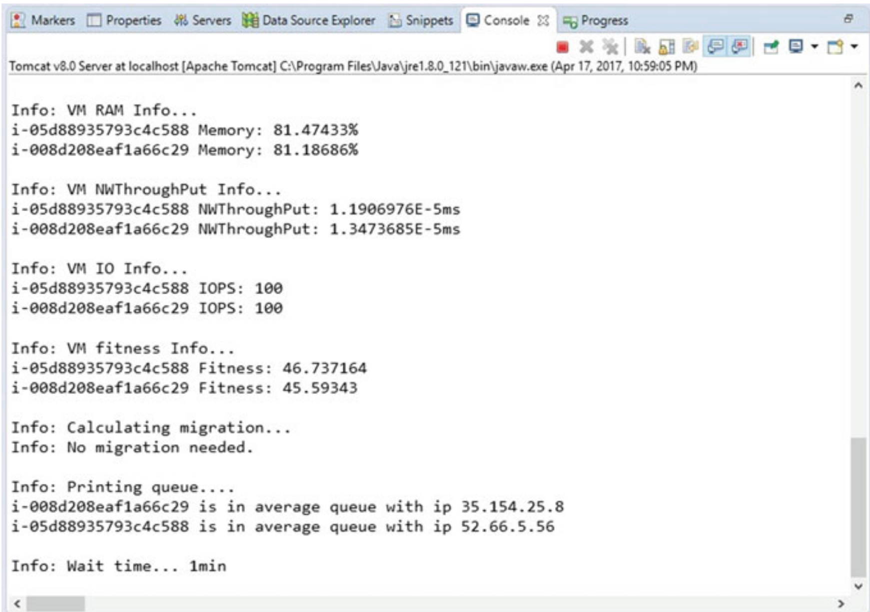


Fig. 5 Execution Snap-1

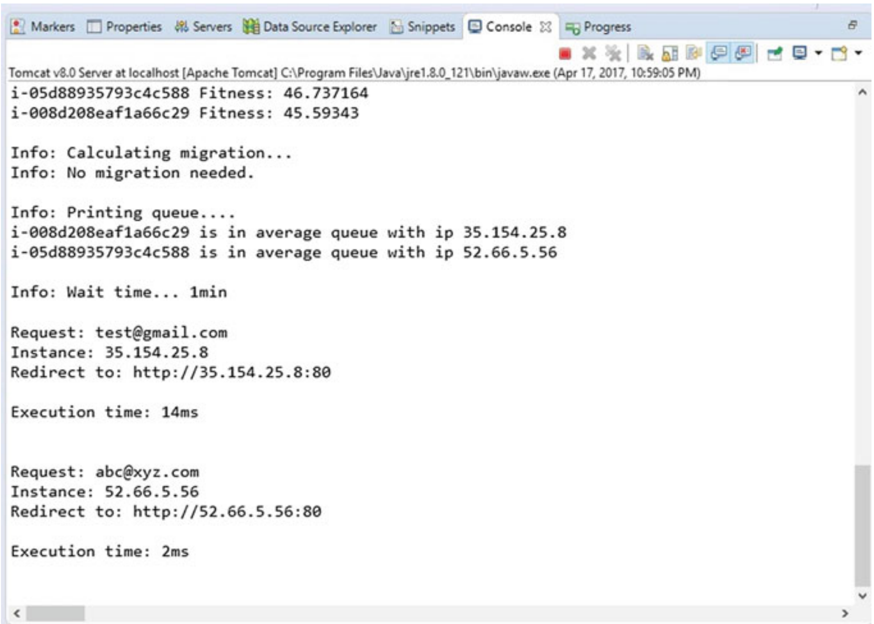


Fig. 6 Execution Snap-2

6 Conclusion

In this work, we surveyed various load balancing techniques for cloud computing. Incorporating the Profiling as a Service in existing infrastructure, we gain an in-depth understanding of network traffic in the cloud. By implementing the proposed Profiling model, one can monitor the usage of resources in data centers. Results show that DCA algorithm outperforms over traditional algorithm in overall resource utilization by 10–20%.

References

1. Reese, G.: Cloud Application Architectures: Building Applications and Infrastructure in the Cloud. O'Reilly Media
2. Mell, P., Grance, T.: The NIST definition of cloud computing. In: National Institute of Standards and Technology, vol. 53, pp. 1–50 NIST, Gaithersburg (2011)
3. Moharana, S.S., Ramesh, R.D., Powar, D.: Analysis of load balancers in cloud computing. Int. J. Comput. Sci. Eng. (IASET) 2, 101–108 (2013)
4. Priority Scheduling Operating System notes. http://www.personal.kent.edu/~rmuhamma/OpSystems/Myos/prioritySchedule.htm. Accessed 13 Oct 2017

5. Beloglazoy, A., Buyya, R.: Energy efficient resource management in virtualized cloud data centers. In: Proceedings of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 826–831. IEEE Computer Society, Washington, DC (2010)
6. Wang, L., Tao, J., Kunze, M.: Scientific cloud computing: early definition and experience. In: The 10th IEEE International Conference Computing and Communications. IEEE Xplore, Dalian, China (2008)
7. Ajith Singh, N., Hemalatha, M.: An approach on semi distributed load balancing algorithm for cloud computing systems. Int. J. Comput. Appl. (IOSR-JCE) **56** (2012)
8. Anand, C., Kapadia, A.: Load balancing algorithm for azure virtualization with specialized VM. Int. J. Innov. Res. Comput. Commun. Eng. (IJIET) **2** (2013)
9. Foster, I., Zhao, Y., Raicu, I., Lu, S.: Cloud computing and grid computing 360-degree compared. In: Proceedings of the Grid Computing Environments Workshop, pp. 99–106. IJSTR (2008)
10. Sahu, Y., Pateriya, R.K., Gupta, R.K.: Cloud server optimization with load balancing and green computing techniques using dynamic compare and balance algorithm. In: Computational Intelligence and Communication Networks (CICN), pp. 527–531 (2013)
11. Tian, W., Zhao, Y., Zhong, Y., Xu, M., Jing, C.: A dynamic and integrated load-balancing scheduling algorithm for cloud datacenters. In: 2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), pp. 311–315. IEEE (2011)
12. Ghafari, S.M., Fazeli, M., Patooghy, A., Rikhtechi, L.: BeeMMT: a load balancing method for power consumption management in cloud computing. In: 2013 Sixth International Conference on Contemporary Computing (IC3), pp. 76–80. IJARIIE (2013)
13. Buyya, R., et al.: Cloud Computing: Principles and Paradigms. Wiley India Edition
14. Shah, M.M.D., Kariyani, M.A.A., Agrawal, M.D.L.: Allocation of virtual machines in cloud computing using load balancing algorithm. Int. J. Comput. Sci. Inf. Technol. (IJCSITS) **3**, 2249–9555 (2013)
15. Parsa, S., Entezari-Maleki, R.: RASA: a new grid task scheduling algorithm. Int. J. Digit. Content Technol. Appl. (IJAIEM) **3**, 91–99 (2009)
16. Chen, H., Wang, F., Helian, N., Akanmu, G.: User-priority guided Min-Min scheduling algorithm for load balancing in cloud computing. In: National Conference on Parallel Computing Technologies (PARCOMPTECH), pp. 1–8. IEEE (2013)
17. Int. J. Comput. Sci. Issues (IJCSI) **9**(1) (2012)

Algorithmic Approach for Error-Correcting Capability and Decoding of Linear Codes Arising from Algebraic Geometry

Arunkumar Patil and Nitin Darkunde

Abstract Error detection and correction of message across noisy channel is one of the important tasks. We tackle the question of finding Gröbner bases of ideals of linear codes [1] arising from algebraic varieties over finite fields. Computation of Gröbner bases [2] of linear codes is a theme of interest to scientists in computational sciences due to its use in decoding and error corrections [1, 3]. Error correction and decoding is one of the main concerns of network and information security. In this paper, generator matrices and Gröbner bases of linear codes associated with Schubert varieties (in some cases) [4] over a finite field with two elements [1] have been computed. We not only found out their error-correcting capability, but also performed decoding of binary Schubert codes [4]. Most of the computations were carried out using open source software SAGE.

Keywords Noisy channel · Binary field · Syndrome decoding · Gröbner Basis Schubert code

1 Introduction

The prime destination of this paper is to compute Gröbner bases of linear codes linked with some special Schubert varieties over binary fields and to study the decoding of Schubert codes. In recent years, Gröbner bases [2] have become a topic of interests for many researchers in mathematics, computer science and allied areas. The techniques of finding Gröbner bases are widely used in computational algebraic

A. Patil
Shri Guru Gobind Singhji Institute of Engineering and Technology, Nanded 431606, India
e-mail: arun.iitb@gmail.com

N. Darkunde (✉)
Swami Ramanand Teerth Marathwada University, Nanded 431606, India
e-mail: darkundenitin@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_51

geometry for ideal membership problem, ideal description problem and implicitization problem, etc. Gröbner bases are used in algebraic coding theory to study structure of linear codes. Borges-Quintana et al. [5] proved that each linear code can be depicted by a binomial ideal. In same paper, authors have determined Gröbner bases of these ideals over $GF(2)$.

The linear codes constructed from higher dimensional algebraic varieties have seen to be of great interest to the people working in algebraic geometry and coding theory. Linear codes from Hermitian varieties, Grassmann varieties, Schubert varieties, determinantal varieties have been studied quite extensively. Grassmann codes have been introduced by Ryan [6, 7] and studied extensively by Nogin [8], Ghorpade and Lachaud [4], Tsfasman and Vlăduţ [9], Ghorpade et al. [10], Hansen et al. [11].

Here we find the Gröbner basis of binary linear code $\mathscr{C}_\alpha(\ell, m)$ associated with Schubert variety [4], denoted by Ω_α. Further, we discuss the decoding of these codes based on decoding technique which has been employed and elaborated in [5].

2 Preliminaries

2.1 *Coding Theory*

Let $GF(q)$ be a finite field with q elements [1, 3], where $q = p^k$, for some prime p and $k \in \mathbb{Z}_+$. By $(GF(q))^n$, we mean a cartesian product of $GF(q)$ with itself n number of times, which is a vector space of dimension n over $GF(q)$. For $0 \neq x \in (GF(q))^n$, we denote its *norm* by $\|x\|$ and define it as the number of entries which are nonzero in an n-tuple $x = (x_1, x_2, \dots, x_n)$. A $k-$dimensional vector subspace of $(GF(q))^n$ over $GF(q)$ is called as $[n, k]_q$-*linear code* [1]. For a linear code $\mathscr{C}$, its (minimum) distance [1] is denoted by $d = d(\mathscr{C})$ and defined as $\min\{\|x\| : x \neq 0, x \in \mathscr{C}\}$. In fact these values are referred as parameters of corresponding code. For more details, one can refer [1, 3]. An $[n, k]_q$-linear code is said to be as *non-degenerate linear code*, if it is not lying in a co-ordinate hyperplane [9] of $(GF(q))^n$. A *generator matrix* [1] for a code $\mathscr{C}$ is denoted by matrix G whose row vectors form a basis set for $\mathscr{C}$, whereas a *parity check matrix* [1] H for code $\mathscr{C}$ is a matrix whose rows form a basis for dual code $\mathscr{C}^\perp$. Also, $v \in \mathscr{C} \iff vH^T = 0$ and $v \in \mathscr{C}^\perp \iff vG^T = 0$. A linear code of distance d is *u-error-detecting* [1] $\iff d \geq u + 1$, whereas a code $\mathscr{C}$ is *v-error-correcting* [1, 12] $\iff d \geq 2v + 1$, where $u, v \in \mathbb{Z}_+$. Hence, $t = \lfloor (d-1)/2 \rfloor$, is the error-correcting capability [12] of a code. Therefore, one always strive for a linear code to have a distance as large as possible.

2.2 Gröbner Bases

Let K be a field [1] and $K[X] := K[x_1, x_2, \ldots, x_n]$ be a polynomial ring [2] in n variables $x_1, x_2, \ldots, x_n$. If $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{Z}^n_{\geq 0}$, where $\mathbb{Z}_{\geq 0} = \mathbb{Z}_+ \cup \{0\}$, then we refer the monomial in $K[X]$ by $X^\alpha := x_1^{\alpha_1} x_2^{\alpha_2}, \ldots, x_n^{\alpha_n}$, and the sum $\alpha_1 + \alpha_2 + \cdots + \alpha_n$ is called as *total degree* [2] of monomial.

A *monomial order* [2] $\prec$ on $K[X]$ is a relation $\prec$ on the set of monomials X^α, such that $\prec$ is a total ordering [2] on $\mathbb{Z}^n_{\geq 0}$ and each non-empty subset of $\mathbb{Z}^n_{\geq 0}$ has a least member [2] w.r.t. $\prec$. One can define ample number of orderings on $K[X]$. The broadly used monomial orders are *lex* order, *revlex* order, *graded reverse lex* order. The degree reverse lexicographic ordering, which is degree compatible ordering, is of immense importance to us.

For a fixed monomial order, and $0 \neq f \in K[X]$, let $\mathcal{MD}(f)$ denote the multi-degree of f which is defined as $max\{\alpha \in \mathbb{Z}^n_{\geq 0} : a_\alpha \neq 0\}$. Also, one can define, the leading coefficient [2] of f, leading monomial [2] of f, leading term [2] of f and total degree [2] of f depicted by $\mathcal{LC}(f)$, $\mathcal{LM}(f)$, $\mathcal{LT}(f)$ and $\mathcal{TD}(f)$, respectively. For an ideal $I \neq 0$ in $K[X]$, we have $\mathcal{LT}(I) := \{c'x^\alpha : \exists \ f \in I \text{ with } \mathcal{LT}(f) := c'x^\alpha\}$. And $\langle \mathcal{LT}(I) \rangle$:= ideal spanned by members of $\mathcal{LT}(I)$.

If we take a fixed monomial ordering, a finite subset $\mathcal{G} = \{g_1, g_2, \ldots, g_s\}$ of an ideal I is called as *Gröbner basis* [2, 5, 12, 13] if $\langle \mathcal{LT}(I) \rangle = \langle \mathcal{LT}(g_1), \mathcal{LT}(g_2), \ldots, \mathcal{LT}(g_s) \rangle$. Moreover, every ideal in $K[X]$ holds Gröbner basis. If one employs Gröbner basis for dividing polynomial $f \in K[X]$, then one can get rid of non-uniqueness of remainder in multivariate polynomial ring. A Gröbner basis [2] $\mathcal{G}$ for an ideal I is said to be *reduced* Gröbner basis [2] if, $\mathcal{LC}(g_j) = 1, \forall g_j \in \mathcal{G}$ and no term in g_i is divisible by $\mathcal{LT}(g_j)$ for $j \neq i$.

2.3 Linear Codes from Grassmann Variety and Schubert Variety

In this section, we define Grassmann and Schubert varieties and define linear codes associated with them through a geometric object called projective system in the set up described in [4, 14, 15] and generate the projective systems corresponding to Grassmann variety and Schubert variety.

By an $[n, k]_q$-*projective system* [4, 9], we mean a set of n points (points may be repeated) in the $(k-1)$-dimensional projective space $\mathbb{P}^{k-1}(GF(q))$. Also, there exists bijective correspondence between the equivalence classes of $[n, k]_q$ codes and the equivalence classes of non-degenerate $[n, k]_q$ projective systems [4].

Fixing ℓ, $1 \leq \ell \leq m$, the *Grassmannian* [4] is set of all ℓ-dimensional vector subspaces of an m-dimensional vector space $(GF(q))^m$ over $GF(q)$. It is denoted by $G_{\ell,m}(GF(q))$ [4]. Henceforth, whenever the context of Grassmannian is going to occur, we will denote it by $G_{\ell,m}$. This Grassmannian $G_{\ell,m}$ is an $[n, k]_q$-projective system [4, 9], along with below values of n, k:

$$n = |G_{\ell,m}(GF(q))| = \begin{bmatrix} m \\ \ell \end{bmatrix}_q := \frac{(q^m-1)(q^m-q)\cdots(q^m-q^{\ell-1})}{(q^\ell-1)(q^\ell-q)\cdots(q^\ell-q^{\ell-1})} \tag{1}$$

and $k := \binom{m}{\ell}$. Let $\mathbb{I}(\ell, m) := \{\alpha = (\alpha_1, \dots, \alpha_\ell) : 1 \le \alpha_1 < \cdots < \alpha_\ell \le m\}$. Let $W \in G_{\ell,m}(GF(q))$ and A_W be a matrix of size $\ell \times m$ possessing *rank* ℓ. Then through *Plücker embedding* [4, 14, 15] set up we obtain the corresponding Plücker coordinates $(p_\alpha(A_W))_{\alpha \in \mathbb{I}(\ell,m)} \in \mathbb{P}^{k-1}(GF(q))$, where $k = \binom{m}{\ell}$, $p_\alpha(A) := det(a_{ij})_{1\le i,j\le \ell}$. These Plücker coordinates satisfy certain quadratic polynomial.

The non-degenerate $[n, k]_q$-code related with projective system defined by $G_{\ell,m}(GF(q))$ along with Plücker map [4, 14, 15] is symbolized by $\mathscr{C}(\ell, m)$ which is named as *Grassmann code* [4] , where the parameters of this code are known to be $n = \begin{bmatrix} m \\ \ell \end{bmatrix}_q$, $k = \binom{m}{\ell}$ and $d = q^\delta$, where $\delta = \ell(m-\ell)$. Now we will take a look at sub-varieties of Grassmannians, viz. Schubert varieties. Let $C_1 \subset \ldots \subset C_\ell$ be sequence of subspaces of $(GF(q))^m$ related to α. Let $C_i = span\{e_1, \dots, e_{\alpha_i}\}$, $1 \le i \le \ell$, where $\{e_1, \dots, e_{\alpha_i}\}$ refers to canonical basis of $(GF(q))^m$. Then $\Omega_\alpha(GF(q)) = \{W \in G_{\ell,m}(GF(q)) : dim(C_i \cap W) \ge i \text{ for } i = 1, \dots, \ell\}$ is Schubert subvariety of $G_{\ell,m}(GF(q))$ related to α. As a projective variety, $\Omega_\alpha(GF(q)) := \{p \in G_{\ell,m}(GF(q)) : p_\beta = 0 \text{ for each } \beta \nleq \alpha\}$ [4] which gives a non-degenerate $[n_\alpha, k_\alpha]_q-$projective system [4, 9], through Plücker embedding [4] which is induced one, hence gives a linear code. Such code is termed as Schubert code [4] and symbolized by $\mathscr{C}_\alpha(\ell, m)$. The parameters of this code are known to be $n_\alpha = \#\Omega_\alpha(GF(q))$, $k_\alpha = \#\{\beta \in \mathbb{I}(\ell, m) : \beta \le \alpha\}$ [4] and $d(\mathscr{C}_\alpha(\ell, m)) = q^{\delta_\alpha}$, where, $\delta_\alpha = \sum_{i=1}^{\ell}(\alpha_i - i)$. In subsequent sections, we will take a look at different Schubert codes such as $\mathscr{C}_{(1,4)}(2,5)(GF(2))$, $\mathscr{C}_{(1,5)}(2,5)(GF(2))$, $\mathscr{C}_{(2,3)}(2,5)(GF(2))$, $\mathscr{C}_{(2,4)}(2,5)(GF(2))$.

3 Binomial Ideals of Linear Codes

In current section, we recollect some already proved results from [5] concerned with Gröbner bases for binary codes and thereby we will apply these results to find out Gröbner bases for these codes listed in Sect. 2. Moreover, we are proceeding to compute the error-correcting capability [1, 5, 12, 13] of Schubert codes in accordance with [5] and will discuss about their decoding. Denote $[X]$ to be the set $\{X^\alpha : \alpha \in \mathbb{Z}^n_{\ge 0}\}$ of monomials in $K[X]$. Further define the map ψ as:

$$\psi : [X] \to (GF(2))^n,$$

such that,

$$\psi(X^\alpha) = \psi\left(\prod_{i=1}^{n} x_i^{\alpha_i}\right) = (\alpha_1 \bmod 2, \dots, \alpha_n \bmod 2) \tag{2}$$

Let R be a relation on $(GF(2))^n$ elucidated below: For two vectors $\psi(X^\alpha), \psi(X^\beta) \in (GF(2))^n$, we have

$$\left(\psi(X^\alpha), \psi(X^\beta)\right) \in R \Longleftrightarrow \psi(X^\alpha) - \psi(X^\beta) \in \mathscr{C}. \tag{3}$$

Now for $\alpha, \beta \in \mathbb{Z}^n_{\geq 0}$, an ideal associated with linear code $\mathscr{C}$ which is denoted by $I(\mathscr{C})$, and known to be a binomial ideal is given as:

$$I(\mathscr{C}) := \langle \{X^\alpha - X^\beta : (\psi(X^\alpha), \psi(X^\beta)) \in R\} \rangle \subset K[X] \tag{4}$$

We can also define $I(\mathscr{C})$ in the set-up of generator matrix [5] of a linear code $\mathscr{C}$ as follows:

$$I = \langle \{X^{w_1} - 1, \ldots, X^{w_k} - 1\} \cup \{x_i^2 - 1 : i = 1, \ldots, n\} \rangle. \tag{5}$$

where $w_1, \ldots, w_k$ are rows of generator matrix of given $[n, k]_2$-linear code.

Theorem 1 *[5] $I(\mathscr{C}) = I$, where $I(\mathscr{C})$ as well as I are described in equation (4) and (5) above.*

Theorem 2 *[5] Suppose $\mathscr{C}$ is binary $[n, k, d]$-code, with $t = \lfloor (d-1)/2 \rfloor$ and let $\mathcal{G}$ be its reduced Gröbner basis w.r.t. $\prec$. For X^α in $[X]$, define $Can(X^\alpha, \mathcal{G})$ to be reduced form of X^α modulo $\mathcal{G}$. If $wt(\psi(Can(X^\alpha, \mathcal{G}))) \leq t$, then $\psi(Can(X^\alpha, \mathcal{G}))$ is error vector corresponding with $\psi(X^\alpha)$, i.e.*

$$\psi(X^\alpha) - \psi(Can(X^\alpha, \mathcal{G}))$$

is the nighest codeword to $\psi(X^\alpha)$. Also, if $wt(\psi(Can(X^\alpha, \mathcal{G}))) > t$, then $\psi(X^\alpha)$ contains more than t errors.

Theorem 3 *[5] Error-correcting capability t of a given linear code $\mathscr{C}$ over $GF(2)$ is given by*

$$t = min\{\mathcal{TD}(f) : f \in \mathcal{G} \setminus \{x_i^2 - 1 : i = 1, \ldots, n\}\} - 1. \tag{6}$$

Now, we will employ above theorems in order to find out ideal associated with corresponding Schubert codes and further compute their reduced Gröbner bases.

4 Gröbner Bases of Schubert Codes

Let us calculate Gröbner bases for different Schubert codes arising from Grassmannian $G_{2,5}$.

4.1 Gröbner Basis for the Schubert Code $\mathscr{C}_{(1,4)}(2,5)(\mathrm{GF}(2))$

Theorem 4 *The Schubert code* $\mathscr{C}_{(1,4)}(2,5)(GF(2))$ *is* $[7,3,4]$*-linear code over* $GF(2)$ *having a generator matrix which is in standard form given by:*

$$A_{(1,4)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}_{3\times 7}.$$

Proof Here $\alpha = (1,4)$, $\ell = 2$ and $m = 5$. Hence, length of this code is $n = 7$. Since $\mathbb{I}(2,5) = \{(1,2),(1,3),(1,4),(1,5),(2,3),(2,4),(2,5),(3,4),(3,5),(4,5)\}$, we get $k_{(1,4)} = \#\{\beta \in \mathbb{I}(2,5) : \beta \leq (1,4)\} = \#\{(1,2),(1,3),(1,4)\} = 3$. Further, $d(\mathscr{C}_{(1,4)}(2,5)) = 2^2 = 4$. Hence, this code is $[7,3,4]$-linear code. Now compute, generator matrix of this code. Here, $\Omega_{(1,4)} = \{p \in G_{2,5} : p_\beta = 0, \text{for all } \beta \nleq (1,4)\}$ Thereafter, we get points in $\Omega_{(1,4)}$, which are listed as follows: $P_1 = (1,0,0,0,0,0,0,0,0,0)$, $P_2 = (0,1,0,0,0,0,0,0,0,0)$, $P_3 = (0,0,1,0,0,0,0,0,0,0)$, $P_4 = (0,1,1,0,0,0,0,0,0,0)$, $P_5 = (1,1,1,0,0,0,0,0,0,0)$, $P_6 = (1,1,0,0,0,0,0,0,0,0)$, $P_7 = (1,0,1,0,0,0,0,0,0,0)$. Then, by using definition of generator matrix, we get generator matrix $A_{(1,4)}$ as above.

Theorem 5 *The error-correcting capability of Schubert code* $\mathscr{C}_{(1,4)}(2,5)(GF(2))$ *is* 1.

Proof By Theorem 1, we can write an ideal associated with this Schubert code, which is given by:

$$I = I(\mathscr{C}_{(1,4)}(2,5)(GF(2))) = \langle x_1x_5x_6x_7 - 1, x_2x_4x_5x_6 - 1, x_3x_4x_5x_7 - 1, x_1^2 - 1, x_2^2 - 1, x_3^2 - 1, x_4^2 - 1, x_5^2 - 1, x_6^2 - 1, x_7^2 - 1\rangle.$$

We denote reduced Gröbner basis of this ideal by $\mathcal{G}_{(1,4)}(2,5)$, which is given by:

$$\begin{aligned}\mathcal{G}_{(1,4)}(2,5) = \langle & x_1^2 - 1, x_1x_2 - x_4x_7, x_2^2 - 1, x_1x_3 - x_4x_6, x_2x_3 - x_6x_7, x_3^2 - 1,\\ & x_1x_4 - x_2x_7, x_2x_4 - x_1x_7, x_3x_4 - x_5x_7, x_4^2 - 1, x_1x_5 - x_6x_7, x_2x_5 - x_4x_6,\\ & x_3x_5 - x_4x_7, x_4x_5 - x_3x_7, x_5^2 - 1, x_1x_6 - x_5x_7, x_2x_6 - x_3x_7, x_3x_6 - x_2x_7,\\ & x_5x_6 - x_1x_7, x_6^2 - 1, x_7^2 - 1\rangle.\end{aligned}$$

By Theorem 3, we obtain $t = 1$. Therefore, the error-correcting capability of this code is 1. We also know the distance d of this code to be 4; therefore, value of t calculated here agrees with the value coming from the formula $t = [(d-1)/2]$.

4.2 *Decoding with* $[7, 3, 4]$*-linear Schubert Code* $\mathscr{C}_{(1,4)}(2, 5)(GF(2))$

By using Theorem 2, we can tackle the decoding problem up to certain extent by using Gröbner bases as follows (Table 1):

As far as decoding of remaining types of received words is concerned, one can employ classical decoding techniques like nearest neighbour decoding technique.

4.3 *Gröbner Basis for the Schubert Code* $\mathscr{C}_{(1,5)}(2, 5)(GF(2))$

Theorem 6 *A generator matrix of* $[15, 4, 8]$ *Schubert code* $\mathscr{C}_{(1,5)}(2, 5)(GF(2))$ *is given by:*

$$A_{(1,5)} = \begin{bmatrix} 1\,0\,0\,0\,0\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1 \\ 0\,1\,0\,0\,0\,1\,1\,1\,1\,1\,1\,1\,0\,0\,0 \\ 0\,0\,1\,0\,1\,0\,1\,1\,1\,1\,0\,0\,1\,0\,1 \\ 0\,0\,0\,1\,1\,1\,0\,1\,0\,1\,1\,0\,1\,1\,0 \end{bmatrix}_{4\times 15}$$

Proof In case of this Schubert code, we have $\alpha = (1, 5), \ell = 2$ and $m = 5$. Hence, we get $n = 15$, $k = 4$ and $d = 8$. Proceeding in the same way, as we have done in Theorem 4 above, we can get generator matrix of this code, given by $A_{(1,5)}$ as above.

Theorem 7 *The error-correcting capability of the Schubert code* $\mathscr{C}_{(1,5)}(2, 5)(GF(2))$ *is* 3.

Proof We can work out the reduced Gröbner basis $\mathcal{G}_{(1,5)}(2, 5)$ of the ideal associated with this code. However, in this case this basis is quite huge one, consisting of 490 binomials of the form $x_i x_j x_k x_l - x_{i'} x_{j'} x_{k'} x_{l'}$, where $1 \le i < j < k < l \le 15$ and $1 \le i' < j' < k' < l' \le 15$, and $\{i, j, k, l\} \cap \{i', j', k', l'\} = \phi$ and some remaining binomials. So, while writing this Gröbner basis, we will prefer to write few of them. Thus, by this convention, reduced Gröbner basis is given by: $\mathcal{G}_{(1,5)}(2, 5) = \langle x_1x_2x_3x_4 - x_8x_9x_{11}x_{13}, \ldots, x_{11}x_{12}x_{13}x_{14} - x_1x_9x_{10}x_{15}, x_1^2 - 1, x_2^2 - 1, x_3^2 - 1, x_4^2 - 1, x_5^2 - 1, x_6^2 -$

Table 1 Decoding of Schubert code $\mathscr{C}_{(1,4)}(\mathbf{2}, \mathbf{5})(\mathbf{GF}(\mathbf{2}))$

Received word	Canonical form of this word w.r.t. $\mathcal{G}_{(1,4)}(2, 5)$	Received word is decoded to
$x_1x_2x_3x_4x_5$	x_4	$x_1x_2x_3x_5$
$x_1x_2x_3$	x_5	$x_1x_2x_3x_5$
$x_2x_3x_4x_5x_6$	x_3	$x_2x_4x_5x_6$
$x_2x_3x_4x_5x_7$	x_2	$x_3x_4x_5x_7$
$x_3x_4x_5x_6x_7$	x_6	$x_3x_4x_5x_7$

$1, x_7^2 - 1, x_8^2 - 1, x_9^2 - 1, x_{10}^2 - 1, x_{11}^2 - 1, x_{12}^2 - 1, x_{13}^2 - 1, x_{14}^2 - 1, x_{15}^2 - 1\rangle$. The minimum of the total degree of each of the above binomial except binomials of the form $x_i^2 - 1$, for $i = 1, 2, \ldots, 15$ is 4 and hence by, Theorem 3, we get $t = 3$. Hence, the error-correcting capability of this code is 3. We have verified this value with the formula $t = \lfloor (d-1)/2 \rfloor$, as $d = 8$ in this case. Decoding with this code can be done in accordance with Theorem 2.

4.4 Error-Correcting Capability of Some Remaining Codes

We can perform calculations on similar lines, as we did in Theorem 4–6 and can arrive at error-correcting capability of below codes, as in Table 2:

Pictorially, we can give graphical representation of codes we studied and their error-correcting capability, as follows (Fig. 1):

In above graph, we have denoted codes $\mathscr{C}_{(1,4)}(2,5)(GF(2))$, $\mathscr{C}_{(1,5)}(2,5)(GF(2))$, $\mathscr{C}_{(2,3)}(2,5)(GF(2))$ and $\mathscr{C}_{(2,4)}(2,5)(GF(2))$ by $C1, C2, C3$ and $C4$, respectively.

Table 2 Error-correcting capability of some remaining Schubert codes

Schubert code	Number of binomials in Gröbner basis	Error-correcting capability
$\mathscr{C}_{(2,3)}(2,5)(GF(2))$	21	1
$\mathscr{C}_{(2,4)}(2,5)(GF(2))$	2127	3

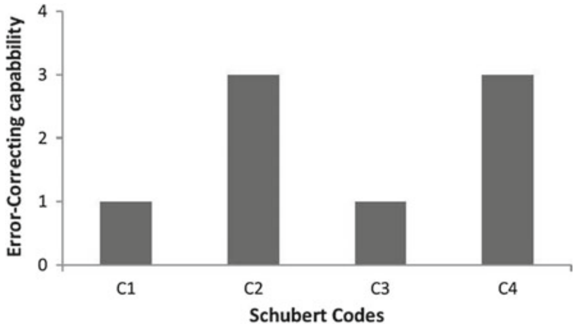


Fig. 1 Error-correcting capability of various Schubert codes

Notes and Comments. We have calculated Gröbner bases and have given decoding of Schubert codes in some special cases. It turns out that these codes have good error-correcting capability, which we always strive for. We have done all computations in a binary field, and it will be interesting to observe the scenario in field with higher orders than 2. In future, we wish to work on this problem for general Grassmann and Schubert codes.

References

1. Ling, S., Xing, C.: Coding Theory-A First Course, 1st edn. Cambridge University Press (2004)
2. Cox, D., Little, J., O'Shea, D.: Ideals, Varieties, and Algorithms, 3rd edn. Springer (2012)
3. MacWilliams, F.J., Sloane, N.J.A.: The Theory of Error Correcting Codes. 3rd reprint, North Holland (1981)
4. Ghorpade, S.R., Lachaud, G.: Higher weights of Grassmann codes. In: Buchmann, J., Hoeholdt, T., Stichtenoth, H., Tapia-Recillas, H. (eds.) Coding Theory, Cryptography and Related Areas, vol. 8, pp. 122–131. Springer, Berlin (2000)
5. Borges-Quintana, M., Borges-Trenard, M.A., Fitzpatrick, P., Martínez-Moro, E.: Gröbner bases and combinatorics for binary codes. Appl. Algebra Eng. Commun. Comput. **19**(5), 393–411 (2008)
6. Ryan, C.T.: An application of Grassmann varieties to coding theory. Congr. Numer. **57**, 257–271 (1987)
7. Ryan, C.T.: Projective codes based on Grassmann varieties. Congr. Numer. **57**, 273–279 (1987)
8. Nogin, D.Yu.: Codes associated to Grassmannians. In: Pellikan, R., Perret, M., Vlăduţ, S.G. (eds.) Arithmetic, Geometry and Coding Theory, pp. 145–154. de Gruyter, Berlin, New York (1996)
9. Tsfasman, M.A., Vlăduţ, S.G.: Algebraic-Geometric Codes, 1st edn. Kluwer (1991)
10. Ghorpade, S.R., Patil, A.R., Pillai, H.K.: Decomposable subspaces, linear sections of Grassmann varieties, and higher weights of Grassmann codes. Finite Fields Appl. **15**(1), 54–68 (2009)
11. Hansen, J.P., Johnsen, T., Ranestad, K.: Schubert unions in Grassmann varieties. Finite Fields Appl. **13**(4), 738–750 (2007)
12. Borges-Quintana, M., Borges-Trenard, M.A., Martínez-Moro, E.: On a Gröbner bases structure associated to linear code. J. Discrete Math. Sci. Crypt. **10**(2), 151–191 (2007)
13. Márquez-Corbella, I., Martínez-Moro, E., Suárez-Canedo, E.: On the ideal associated to a linear code. arXiv preprint: arXiv:1206.5124 (2012)
14. Patil, A.R.: Weight hierarchy and generalized spectrum of linear codes associated to Grassmann varieties, Ph.D. thesis, Indian Institute of Technology, Bombay (2008)
15. Hodge, W.V.D., Pedoe, D.: Methods of Algebraic Geometry—volume 2. Cambridge University Press (1952)

Interoperability Among Internet of Things (IoT) Components Using Model-Driven Architecture Approach

Kiranpreet Kaur and Anil Sharma

Abstract Internet of things (IoT) has become most rising technology in IT. The IoT world becomes very smaller, and the possibilities of IoT are endless. The center of attention of IoT is in the area of smart health, industrial (real-time analytical, automation) education, smart farming, connected home (security thermostats), connected cities (like smart meters). Although Internet of things (IoT) is becoming the essential aspect of every person's life and making the life comfortable, on the other side there are some challenges and problems which are also found in the Internet world. These issues make the smart world more unsafe and challenging like privacy issue, security, reliability, scalability, and interoperability. In these days, problem of incompatibility of devices is at high risk that makes things more expensive. So there is a need to tackle the interoperability problem. This paper proposed a model-driven approach (MDA) for making devices compatible. In the thesis, authors worked on platform independent model (PIM) and computation independent model. Computation independent model (CIM) is captured for the information and transferred into the platform independent model and Unified Modeling Language (UML) tool generate the Java code as a vocabulary.

Keywords Internet of things (IoT) · Model-driven architecture (MDA)
Computation independent model (CIM) · Platform independent model (PIM)
Unified Modeling Language (UML) · Zigbee

K. Kaur (✉) · A. Sharma
School of Computer Applications, Lovely Professional University,
Phagwara, Kapurthala (Jalandhar), India
e-mail: kphayer.577@gmail.com

A. Sharma
e-mail: anil.19656@lpu.co.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_52

1 Introduction

Today Internet of things (IoT) is one of the emerging technologies that not only enhancing the computing capabilities in little things but also require the development of new services and applications. Moreover, the term web/internet of things formed by two words "Internet" plus "Things", where internet refers to communication protocol (TCP/IP), while Things (IoT) are identifiable objects, so semantically Internet of things means overall system of interrelated things which exclusively depends on typical transmission protocols [1]. The Internet of things (IoT) is anything that can be connected or will be connected and communicated by using wireless sensor. Web/Internet of things refer to a widespread plan for the ability of system gadgets towards recognizing and gathering. The plenty of frame works used for information from a common environment, and thereafter exchange that data over the Internet. However, it is not enough for things to be linked to the Internet; these also need to be found, accessed, managed, and potentially associated with other "things." To make possible this interaction, a higher degree of interoperability is required that goes beyond simple protocol interoperability as provided by the Internet [2].

To address this issue, some have started to make large-scale centers that give a reliable and simple to utilize interface for both integrators and application engineers. While these hubs, by the simple method of aggregation of data and standard representation of "things" give a level of interoperability, these things normally do not interoperate with each other [3]. Authors argue that the lack of inter-hub interoperability may stifle the uptake of the IoT. Many of the things of interest to application and service developers will only be reachable using product or hub-specific APIs. As long term, this must be addressed through a standardization process, and indeed many in the Web and IoT community have started this process in various groups.

This paper presents the model-driven architecture approach for ensuring the interoperability among IoT components. Moreover, authors modeling the computational independent model (CIM) and platform independent model (PIM), and Zigbee protocol for exchanging data between devices. In this paper, Sect. 2 demonstrates about the basic information of model-driven architecture and then related work presented in Sect. 3. Our proposed work described in Sects. 4 and 5 illustrates the discussion and results. Then conclusion and references presented in Sect. 6.

2 Model-Driven Architecture (MDA)

In 2001, Object Management Group (OMG) adopted an approach called as model-driven architecture for better software development. Its main aim is to reduce the complexity, costs, quality of new application software. Moreover, interoperability among software is the main concern of model-driven architecture (MDA) [4]. MDA means model-driven approach used for specification that provides the protocols and guidelines for developing a system or architecture to develop a system before its problem. Demonstrate-Driven Engineering (MDA) as a way to deal with application plan and usage [5]. This is an extremely positive advancement for a few reasons. MDA energizes productive utilization of framework models in the product advancement process, and it sustains reuse of best practices while making groups of frameworks. As characterized by the protest administration gather (OMG), MDA is an approach to compose and supervise undertaking designs upheld via robotized devices and administrations for both characterizing the models and encouraging changes between various model sorts. This paper proposes architecture for better connectivity of heterogeneous devices [6] (Fig. 1).

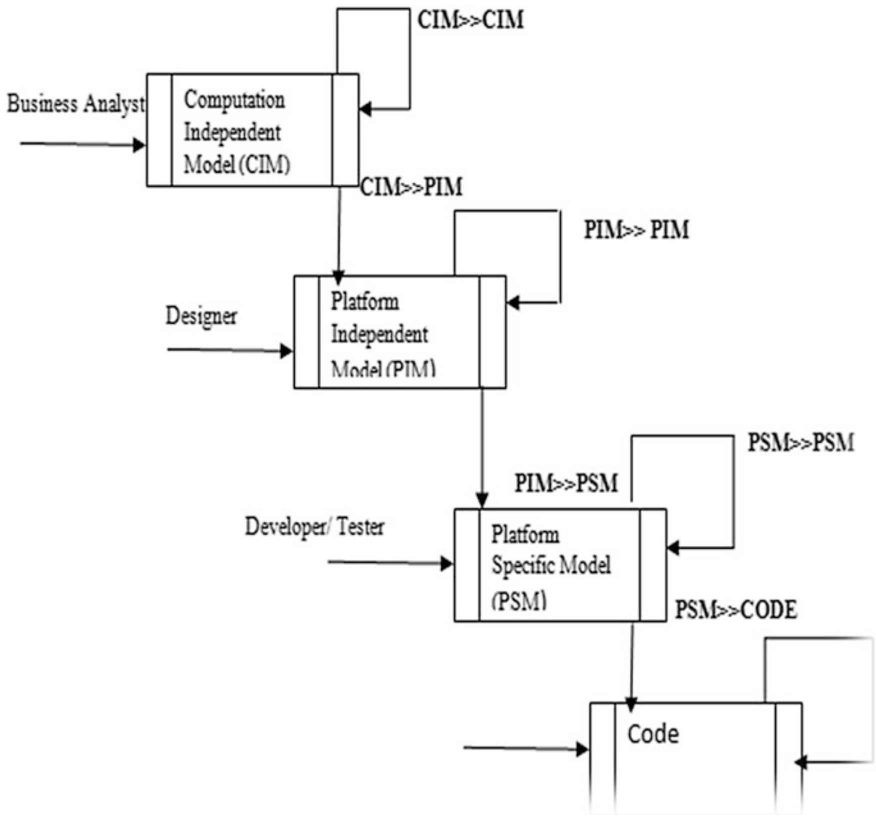


Fig. 1 MDA concept and models

2.1 Computation Independent Model

In MDA, computation independent model is basically used for capture the requirements of data of the system. Sometimes, it is called a domain model of the system. Moreover, it is estimated that the essential client of the CIM is the area specialist or business advisor. The CIM specifies the function or only external behavior of a system without knowing the internal functionality. CIM indeed work for data gathering of particular domain rather than implementation or analyzing [7].

2.2 Platform Independent Model

Platform independent models are the platform independent model that can run on any platform of same type of platform. The principle object of this model is to make the dynamic idea of the framework utilizing the grammar of the domain quit far. A product designer can present this idea by using this model which is distinctive displaying vocabulary. Platform independent model requests a general depiction to catch the semantics of wide range of spaces, which must to likewise be sufficiently exact to transformation change into end system.

2.3 Platform-Specific Model

A platform-specific model specifications join in the PIM with the facts of interest essential to specify how a framework uses a specific sort of stage. PSM does exclude the greater part of the detail of interest important to create an execution of that platform, and it is viewed as dynamic (implying that it depends on other express or verifiable models which do contain the essential elements).

3 Related Work

Ren et al. in 2016 give the information about the advancement of technology and computing availability that make it promising to insert rising computing power in little devices. Author also describes the combined advances in correspondence and results in a bionetwork of extremely interconnected gadgets called as the Web [8].

Koo et al. in 2016 proposed the bathroom safety enhancement through Internet of things. Author uses the new technology to provide the safety for elderly people in bathroom using wireless sensors in bathroom tub, toilet, shower, and floor. This paper is focus on the approach of Internet of things framework, improvement, and execution to upgrade bathroom safety enhancement. In this paper, author proposed sensor selection and application, integration of a wireless sensor local network system, design concept for implementation, and big data analysis system model. The study of this paper presents a design concept for a pilot test and big data system [9].

Gill et al. **in 2016** deliver sequence planning for elderly people using IoT components. Author uses the Resalert approach for information architecture for the effective delivery in the emergency situations. In this paper, author proposed a unique IoT enabled design and motivated approach Resalert to tackle the problem of tragedy information warning deliverance toward aged citizens. Author used LED projector, light bulbs, NBN connectivity, and GEO locations for efficient delivery of information to unnatural areas. The Raspberry pi system-based algorithm is used in system architecture [10].

Suryadevara et al. in 2015 proposed the normally contradicting necessities for QoS and EE provisioning in an incorporated way for an IEEE 802.15.4-based remote sensor arrangement (WSN). Author also introduces a versatile vitality productive calculation, named as ABCD that adjusts the center to get a control (MAC) parameters of IEEE 802.15.4 sensor hub. Author describes ABCD algorithm which is further improved by incorporating with a power collecting method, and another MAC parameter is known as the vitality back-off. The upgraded calculation called as EH-ABCD which offers main concern to various module of movements by enhancing the battery's life span and QoS values under factor activity stack conditions. At the end, author considers the arithmetical consequences which verify that proposed algorithms attain high EE and QoS values, whereas amplifying the life expectancy of sensor hubs for outside applications [11].

Karakostas in 2016 gives information about Web/Internet of things (IoT) scenarios, and author describes that there is a necessitate to forecast measures produced by items as a result of the dynamicity of IoT conditions, and it is hard to foresee with assurance if/when such occasions will happen. In this article, creator recommends an engineering that utilizes a Bayesian occasion expectation shows that utilizations verifiable experience information produced by the IoT cloud to process the likelihood of prospect measures. Moreover, author reveals the design by actualizing a model framework to anticipate outbound flight postpone occasions, in view of inbound flight delays, in light of recorded information made from flying insights databases [12].

Koo et al. in 2015 describe a graphic progress of IoT purpose for immense data compilation during an innumerable of water customers. Author proposed the plan comprises of downstream and upstream information gathering utilizing wireless sensor network (WSN) advancements linking to IoT. Downstream information might display irrigate utilization also execution information to customers and upstream information is like customary SCADA and automatic meter Reading (AMR) frameworks. Toward the end, this paper exhibits all information will be united to make a big data accumulation structure wherever information mining recognizes (1) neighborhood and framework exhibitions including weight and stream, (2) non-income and unlawful water utilization, and (3) areas and measure of stream breaks and water misfortunes. Author describes the objective of this advancement which is to permit both utilities and customers to proactively deal with their water use and accomplish larger amounts of maintainability in water supply [13].

Kun-Kun et al. in 2013 resent the product stage of the executive framework in smart home which is intended, and Chinese clients are utilized. Author break down the fundamental correspondence among clients and PC, schedule, climate anticipate, Web perusing, and various electrical models are intended to manage the acculturated connection among human and PCs, and creator executed a few elements, for example, investigation of enthusiastic models, outline of the passionate models, content feeling recognizable proof, and further essential capacities in human–PC communication. In this paper, author studies the emotional interaction system on smart home appliances and corresponds with users' after analyzing the specified emotions through text. At the end, author tells that the system with emotion is personalized, user-friendly and can be able to realize pleasant human–machine interaction [14].

Qiu, **Tie et al. in 2016** offer a directing convention for disaster reaction IoT depend on upon Global Information Decision (ERGID) to build up the exhibitions of steady at a communicate and proficient crisis reaction in IoT. Intentionally, creator outline and understand a technique refer to as delay iterative method (DIM), which depends on defer inference, for tackling the inconvenience of overlooking reasonable ways. In addition, author proposed sending system named as the residual energy probability choice (REPC) to adjust the heap of system by concentrating on the leftover force of hub. Creator gives data about the imitation outcomes and examination demonstrates that ERGID out performs EA-SPEED and SPEED as far as end-to-end (E2E) delay, parcel misfortune, and vitality utilization. Moreover, creator additionally does some sensible analyses with STM32W108 hub antenna and analyzes that ERGID can enhance the constant reaction capacity of system [15]. This paper presented the transformation from computation independent model to platform independent model in order to get the platform for interoperable devices. The proposed models show how a single platform can consolidate. Operations in ways are totally self-determining of WSDL, XML, UDDI, SOAP, Java, and other Web service execution innovations. Communication tools can then be used to create WSDL, UDDI, XML, SOAP and the innovation particular relics lastly the execution code from the plan input.

Yang, **Shuo**, **Jingzhi et al. in 2016** use semantic interoperability technique which consist of two regular stages including reliable data understanding and logical data usage. To implement semantic interoperability, author proposes a novel automatic tabular document exchange (Doc Ex) structure comprised of a new tabular document model (Tab Doc) and a semantic implication scheme to fit the two stages.

Author describes about this Tab Doc model that a new Tabular Document Language (Doc Lang) as an interaction medium between users and devices. Besides, this paper proposes a semantic inference algorithm (SIA) implementing the inference procedure on received tabular documents produced by a table designer system which integrates with SIA. The outcome shows that the planned technique in this article improves the act of information synthesis between different information systems on the Internet [3].

4 Proposed Work

In this section, authors discussed the proposed work in our specific area of research in Internet of things interoperability. It is clear that the advances are continuously developing. Rather than developing these smart gadgets with existing technologies, modeling these things at an upper level of generalization will decouple from the unwanted effects of technology modify and improve their endurance. An MDA-based elaboration of interoperability in smart things will enable defining these things in a technology-independent manner and will play a major role in rising up the quality of Internet of things, making these devices more robust, flexible, and agisle. Exemplifying business rationale in a way that is free of the specialized systems will formally catch the substance of the applications and will likewise make it conceivable to reuse these in an assortment of settings. Model-driven architecture approach is based on any technology and used Unified Modeling Language for reusing the vocabulary using models. In this paper, authors are using computation independent model (CIM) and platform independent model for transformation and exchanging information. Although CIM and PIM models are established in cloud services, our contribution is for smart devices and it is differentiate by its functionality and Zigbee protocol.

In the CIM, design shows the process of how data packet encodes into home automation profile as following:

1. In this diagram the data packet that is in different format as compared home automation send to the profile of Zigbee before the Zigbee profile it checks the coming data packet is valid, if it is valid send to the profile otherwise it goes back to the Phillips hue profile.
2. Then it checks that it is Zigbee lighting profile in Zigbee lighting library (ZLL) if yes it enter the data packet.
3. Again it checks the data and entered the data correctly according the file.
4. If the entered the data packet is valid, it sends the process to decode the data into Zigbee home automation file decoder.
5. After that it changes the type of data packet according to the requirements of home automation profile.

6. A data packet check in the home automation profile that decodes data which is match with the home automation profile for communication.
7. Is data packet is match with the HA profile send to smart things for connectivity otherwise goes back into decoder for convert the data format (Fig. 2).

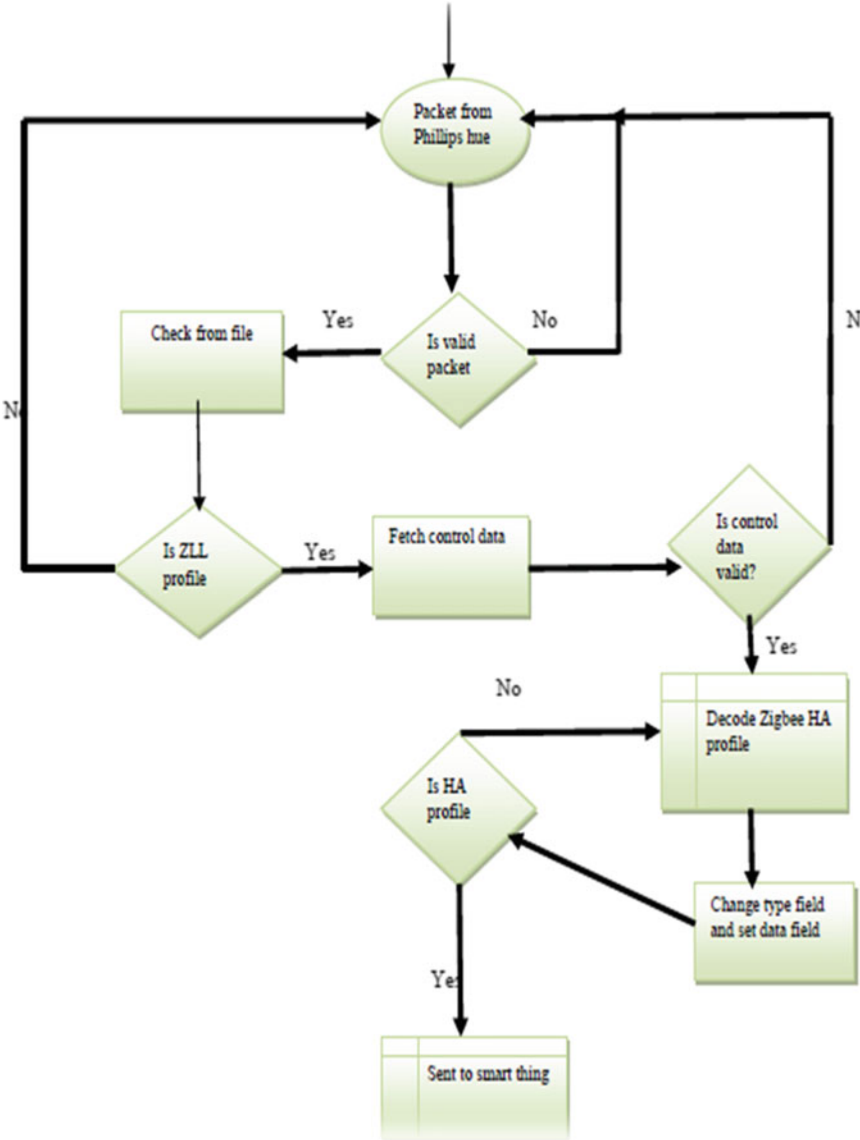


Fig. 2 CIM activity diagram

- Zigbee—Communication Protocol
- HA—Home Automation
- ZLL—Zigbee Lighting Library.

Above diagram of CIM presents only process of activity of its working. In this model, simple steps are presented and data is gathered for transformation into PIM model for analyzing it. CIM is the domain model in our proposed work (Fig. 3).

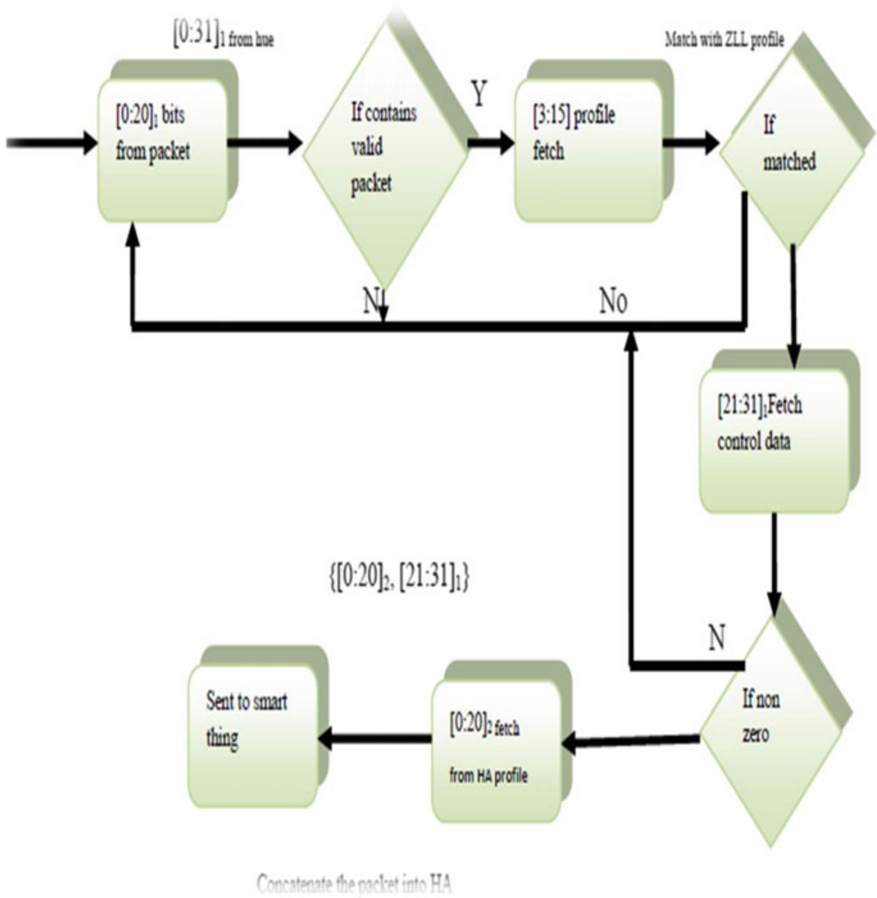


Fig. 3 PIM model

4.1 Platform Independent Model Design

In the design of PIM, the information from computation model transfers to platform independent model. In this platform, the diagram works according to the information and the code generate from this platform can run on the home automation profile. Let us discuss the diagram of PIM:

1. In the smart hue light, the data packets are used that are different from home automation things.
2. From the hue light, data packet is coming in bits form that 0 or 1 the total bits of hue smart light is 31 but it sends the bits in data packet that carries only 20 bits.
3. After that it checks the data packet that is coming from the smart light; if the address is valid, then it enters into the profile otherwise returns to the first step.
4. Then entered data match with the Zigbee library profile that is transmission protocol if it matched with the ZLL profile and fetch data for decoding; otherwise, it returns back.
5. The decoded bits of hue light is ready to used in home automation, but firstly it checks the address of hue light; if the address is not empty, then it send to the home automation profile where it exchange the data packets.
6. The decoded data packets converted into smart things that can run on any platform and can control by any device.

A PIM indicates the framework at a higher level of generalization as compared to a PSM. While the PIM is independent of any technology utilized for executing the system, it can be reused over some different platforms. So our proposed PIM design can execute on any platform using the UML. In this thesis, PIM used for home automation and hue lights for compatibility between each other.

4.2 ZigBee

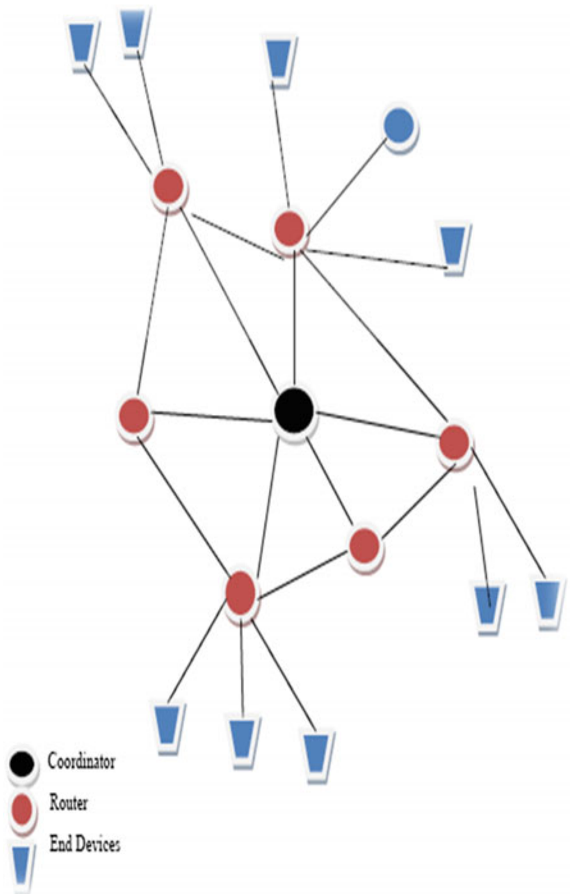
Zigbee is a remote systems administration standard that is designed for remote control and sensor applications which is reasonable for operation in cruel radio conditions and in confined areas. Zigbee innovation expands on IEEE standard 802.15.4 which characterizes the physical and MAC layers. Zigbee characterizes the application and security layer particulars empowering interoperability between items from various producers [16].

Zigbee works at 900–928 MHz and 2.4 GHz and speed of Zigbee is 250 kbp communication protocol [17]. There are some reasons for using the Zigbee protocol for communicating the two devices:

- If user have to send information over wide vicinity, where it crosses various hops and every object one need to control are low-controlled or even battery-consumed. Consequently, Zigbee is mainly utilized in home automation system manufacturing and ZigBee's capacity to maintain interconnect networking that can help information broadcast range and supply better constancy (when only linked hub fails and does not work).

- Zigbee devices can be coordinators, switches, and end gadgets. Coordinators set up the system and store data like protection keys, while switches perform as middle hubs and transfer information from different devices. A common illustration is the point at which you claim a keen light (i.e., Zigbee-empowered) and a brilliant light switch (i.e., additionally Zigbee-empowered), and you need the light change to speak with and control the light. With Zigbee, the two gadgets—regardless of the possibility.
- From various makers—talk a typical vocabulary. Zigbee does not concentrate on the indicate point market, for example, Bluetooth, where one powerful gadget sends information to another powerful gadget over a short range (Fig. 4).

Fig. 4 Zigbee transmission protocol for communication



Zigbee protocols is best suited for the communication and transformation proposes in smart home automation because it can work for a long time and used this protocol for connecting the smart gadgets to achieve the interoperability in the Internet of things world.

5 Discussion and Analysis

The model-driven architecture approach seems useful to the efficient incorporation among smart devices at a high. This research suggests the need for interoperability for two communication devices or Internet of things components for the communication purposes of exchange the data from one device to another device (e.g., hue light and smart thing and how the information is exchanged between each other IoT device). This paper presented the transformation from computation independent model to platform independent model in order to get the platform for interoperable devices. The proposed models show how a single platform can consolidate. Moreover, operations in ways that is totally self-determining of WSDL, XML, UDDI, SOAP, Java, and other Web Service execution innovations. Communication tools can then be used to create WSDL, UDDI, XML, SOAP and the innovation particular relics lastly the execution code from the plan input.

This screenshot shows the class diagram of platform independent model. Class diagram of this model contains the attributes and variables for interlinked with each other generate the code which is useful for communication in IoT devices. It is a tool that is useful to create classes and generate the vocabulary for exchanging the information between smart devices (Fig. 5).

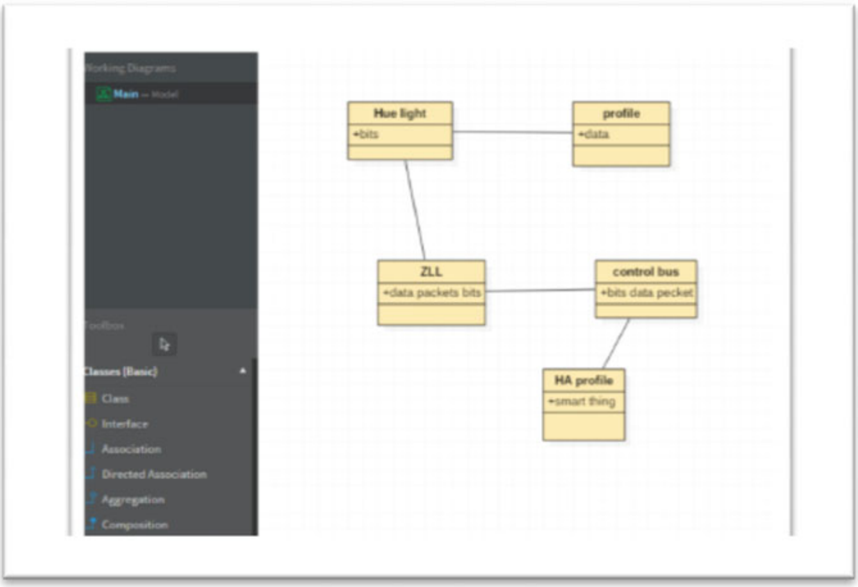


Fig. 5 PIM class diagram

A PIM determines the framework of higher point of generalization when contrasted with another models. Since the PIM is autonomous of any innovation utilized for executing the framework, it can be reused over a few dissimilar platforms.

The output of this model generates the code as Java language called the vocabulary in MDA environment that helps to understand and useful to communicate between devices. The PIM class diagram creates attributes and relations between these entities and attributes that execute this model and produce the language that can execute and run on any platform with different programmability and helps to devices to connect with each other.

In addition to, a PIM indicates the framework at a higher level of generalization as compared to a PSM. While the PIM is independent of any technology utilized for executing the system, it can be reused over some different platforms. So our proposed PIM design can execute on any platform using the UML. In this thesis, PIM used for home automation and hue lights for compatibility between each other (Fig. 6).

File Edit Format View Help

```
import java.util.*;
       public class InterOp
       {
          public static ArrayList<String> available_zll_profiles = new ArrayList<String>();
          public static ArrayList<String> available_ha_profiles = new ArrayList<String>();
          public static void addProfilesToDB()
        {
            available_zll_profiles.add("010111000011010");
          available_ha_profiles.add("001010111001010");
      }
          public static boolean isNonZero(String  pkt)
     {
      int result = 0;
          for(int i=0;i<20;i++)
        {
        result=result+Integer.parseInt(""+pkt.charAt(i));
      }
        return result==0?false:true;
      }
         public static boolean belongsToZLLProfile(String  zll)
      {
          String substr = zll.substring(5,20);
          if(available_zll_profiles.contains(substr))
      {
       return true;
      }
      else
       {
         return false;
   }   }
          public static boolean belongsToHAProfile(String ha)
      {
        String substr = ha.substring(5,20);
          if(available_ha_profiles.contains(substr))
      {
        return true;
     }
```

Fig. 6 Output of PIM model

File Edit Format View Help

```
        return false;
        }        }
      public static void main (String[] args)
         {
           addProfilesToDB();
         String ha="01110001010111001010001011001011";
          String zll="10100010111000011010101111110101";
          Scanner sc = new Scanner(System.in);
          System.out.println("Enter HA packet");
           ha = sc.next();
           System.out.println("Enter ZLL packet");
           zll = sc.next();
            if(ha.length()==32 && zll.length()==32)
          {
              if(isNonZero(ha) && isNonZero(zll))
          {
          if(belongsToZLLProfile(zll) && belongsToHAProfile(ha))
          {
             String final_output = ha.substring(0,20)+zll.substring(20, 32);
            system.out.println("Output Packet :\n"+final_output);
          }else
         {
        System.out.println("doesn't belongs to profiles");
        }                        }
        else
        {
          System.out.println("non zero check is not satisfied");
       }         }
else
 {
     System.out.println("Packet length not equal to 32");
 }    }
    public InterOp()
    {
super();
}
}
```

Fig. 7 Output 2 of PIM model

This is the output of PIM model that designed in Unified Modeling Language (UML). It is the code generate by platform independent model during the execution time of this model. The generated code is the Java languages that is flexible and platform independent (Fig. 7).

This produced Java language explained the whole working of class diagram of PIM model, and this output shows the completeness and correctness of platform independent model. It shows that when user entered the data or bits of hue light, this model changes or type of that data according the device requirements.

Figure 8 shows the final result to prove the interoperability. In this when author fetch the HA packet and ZLL packet for compilation, it matches both packets and combines these packets and the generated output used for the interoperable objects.

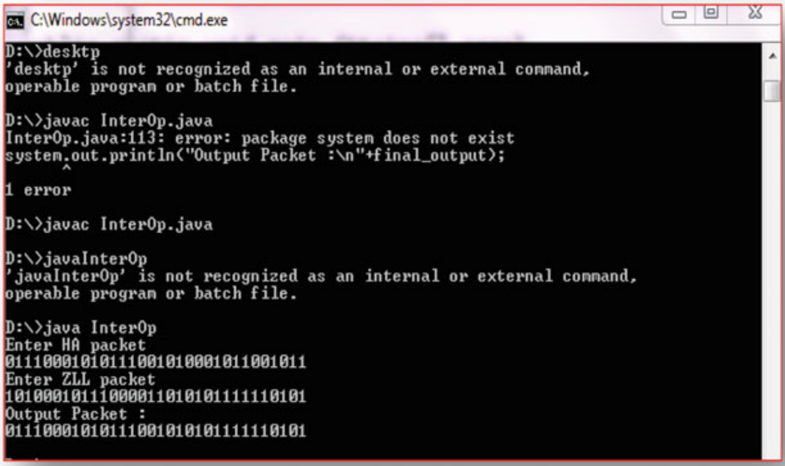


Fig. 8 Result

6 Conclusion and Future Scope

Toward the end, it is concluded that interoperability in the IoT is essential to achieving the capability of the applications and administrations that can collaborate with smart things in our physical world. At that point, the development of IoT centers that aggregate IoT gadgets utilizing Web protocols, application designers can get to individual hub-hosted facilitated physical assets, for example, atmosphere sensors, home automation equipment, home appliances, and different things in a uniform way. The problem of interoperability becomes clearly one of unifying the introduction of center point lists and information positions.

In our coordinated effort, authors proposed an approach for better interoperability arrangements in the continually growing technologies. This paper needs to create MDA approach for Internet of things, and in this approach authors utilized CIM and PIM models for reusability in IoT environment. MDA in view of model-driven engineering way to deal with software development is generally new and particularly quick creating technique. So MDA is an approach that is most appropriate for the arrangement of interoperability. This paper takes a home automation gadget that can associate with the smart hue light utilizing MDA models approach. Later on MDA can be utilized for product quality and portability also. The conclusion that it is possible to produce an innovation/platform subordinate executable code from the PIM naturally using the ideas of MDA and that MDA can efficiently deal with the difficulties of evolving advancements, various environment, interoperability, and so on, reusing the models. There is a need to find more solutions for the interoperability issue and should be improve the approach for better communications of devices. Also, there is other challenges need to solve for Internet of things environment.

References

1. Domingo, M.: An overview of the Internet of Things for people with disabilities. J. Netw. Comput. Appl. **35**(2), 584–596 (2012)
2. IERC, A.C.: IoT semantic interoperability: research challenges, best practices, recommendations and next steps. Technical Report 8. European Commission Information Society and Media (2013)
3. Yang, S., Jingzhi G., Ran W.: Semantic interoperability with heterogeneous information systems on the internet through automatic tabular document exchange. Inf. Syst. (2016)
4. Soley, R.: Model driven architecture. OMG white paper 308.308,5 (2000)
5. Truyen, F.: The fast guide to model driven architecture the basics of model driven architecture. Cephas Consulting Corp (2006)
6. Singh, Y., Manu, S.: Models and transformations in MDA. In: First International Conference on Computational Intelligence, Communication Systems and Networks CICSYN'09. IEEE (2009)
7. Garrido, J.: Definition and use of computation independent models in an MDA-based groupware development process. Sci. Comput. Program. **66**(1), 25–43 (2007)
8. Ren, C., et al.: When biometrics meet IoT: a survey. In: Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation. Atlantis Press (2016)
9. Koo, D.D., et al.: An internet-of-things (IoT) system development and implementation for bathroom safety enhancement. Procedia Eng. **145**, 396–403 (2016)
10. Gill, A.Q., et al.: IoT-enabled emergency information supply chain architecture for elderly people: the Australian context. Inf. Syst. **58**, 75–86 (2016)
11. Suryadevara, N.K., et al.: WSN-based smart sensors and actuator for power management in intelligent buildings. IEEE/ASME Trans. Mechatron. **20**(2), 564–571 (2015)
12. Karakostas, B.: Event prediction in an IoT environment using naïve Bayesian models. Procedia Comput. Sci. **83**, 11–17 (2016)
13. Koo, D., Kalyan, P., John, M.: Towards sustainable water supply: schematic development of big data collection using internet of things (IoT). Procedia Eng. **118**, 489–497 (2015)
14. Kun-kun, D., Zhi-liang, W., Mi, H.O.N.G.: Human machine interactive system on smart home of IoT. J. China Univ. Posts Telecommun. **20**, 96–99 (2013)
15. Qiu, T., et al.: ERGID: an efficient routing protocol for emergency response internet of things. J. Netw. Comput. Appl. **72**, 104–112 (2016)
16. Al-Fuqaha, A., et al.: Internet of things: a survey on enabling technologies, protocols, and applications. IEEE Commun. Surv. Tutor. **17**(4), 2347–2376 (2015)
17. Jain, R.: Wireless Protocols for Internet of Things: Part II–ZigBee. (2014)
18. Aissani, M., et al.: EA-SPEED: energy-aware real-time routing protocol for wireless sensor networks. Int. J. Inf. Commun. Technol. **5**(1), 22–44 (2013)

Hand Phone Open-Air Face Image Retrieval for Varying Focal Length and Occlusion from Cloud Storage

V. Elanangai

Abstract The rapid growth of smartphone instrument apps like fluke thermal app, medical device apps replaces the existing instruments; in such scenario, the smartphone camera plays a vital role in replacing vision instruments. The smartphone camera image and image retrieval needs effective algorithms compete to the existing algorithms. In this paper, I evaluate the different smartphone images for enhanced face image retrieval. The smartphone face image retrieval id developed with multiple facial and embedded sparse code words attributes for better results of the images acquired under different environmental conditions such as indoor, outdoor, illumination. The 50 images are acquired from four different smartphone face images and publicly available datasets such as PubFig to compare the efficiency of proposed algorithm. From the comparison, dynamic indexing in the algorithm retrieves the image from database in milliseconds. From the results, I conclude that the performance of algorithm varied for smartphone and dataset images.

Keywords Face recognition · Open-air face image · Cloud storage
Dictionary learning

1 Introduction

Face image search plays a major trend in image processing person identification. However, with advancement in smartphones, make the people upload the photo of themselves and others in the Internet. The uses of smartphone rapidly increase. The image-based search for person identification from the database is more suitable

V. Elanangai (✉)
Sathyabama Institute of Science and Technology, Chennai, India
e-mail: elanangai123@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_53

when compared to text-based search for person identification. Resolution relativity between images plays a major role in face recognition. Recognition algorithms often fail to recognize face due to the difference in resolution between images. Local octal pattern algorithm recognizes face images with different resolution effectively. A novel method of representing high-resolution face images into high discriminative spaces and lower dimensionality. Resolution orthogonal locality preserving projection method places landmarks by sparse representation to preserve important details in high-resolution images [1]. The video recorded by video surveillance cameras are of low resolution. Face recognition for low-resolution images yields better result by use of singular value decomposition. The effectiveness of singular value decomposition in representing a face image is more than other methods. The proposed model generates a matrix for high-resolution images for every low-resolution image. By mapping process and pixel, SVM analysis, the image characteristics are learned more accurately [2]. A hyperspectral face recognition algorithm is proposed which extends the limitation of 13 different face recognition techniques. Gray scale and RGB, face recognition system, are compared with hyperspectral face recognition system, and their advantages are shown [3]. Benchmarking for different face recognition systems is widely done, namely Video-to-Still (V2S), Still-to-Video, and Video-to-Video (V2V). A Point-to-Set Correlation Learning (PSCL) is proposed for face recognition on COX faces DB [4]. Expression detection along with face recognition is a major trend. The different poses and expressions are detected using probabilistic facial recognition model. Each expression has corresponding feature library matrix represented in the 3D model. By comparing FLM data with available data, the expression and feature matched. All these model sets together form a database, which detects the expression changes in an image [5]. For the face recognition by description, a compact binary face descriptor (CBFD) is proposed. The descriptor produces different pixel values for images with different resolutions. The difference in pixel values is determined using different pixel vectors. Binary codes are formed using feature mapping technique to detect the difference in binary codes. Along with this, coupled CBFD (C-CBFD) is implemented to make heterogeneous face recognition easier [6]. Face recognition by deep architectures is a success. A supervised autoencoder and standardized autoencoder along with deep architecture extracts feature in an image. The proposed method outperforms deep Lambertian network without any training data or domain information [7]. However, conventional methods do not recognize face images distorted by occlusion and reflection. In addition, the system needs more resources to perform face recognition. The proposed work is a novel approach, which performs face recognition on handheld device with assistance from cloud environment. Section 2 explains the related work, Sect. 3 explains the proposed methodology, and remaining section explains the result and discussion.

2 Related Work

Dictionary learning method of face image search considers illumination and resolution in recognition. The trained classifier can compare input image with the sparse codes to recognize images. The data sets together to form a dictionary, which will be useful in making recognition easier [8]. Eye coordinates play an important role in face recognition. Eye coordinates matching is the first step in face recognition. The output of the system will vary completely if eye coordinates differ and all other features match. Different eye coordinates and the ambiguity caused due to their placement are investigated [9]. Frontal face recognition often requires good illumination and correct pose for the designed algorithm to work effectively. Patch-based face representation scheme represents an image with a partial frontal face. Multitask learning scheme represents these patches in a transformation dictionary. The dictionary transforms different poses into discriminative subspaces. The face at different poses can be recognized using discriminative subspace, which increases algorithm efficiency [10]. Data-driven method of face recognition recognizes the face of people images taken at a different lifetime. The images collected form a dataset, which consists of cross-age images. A data-driven cross-age reference coding (CARC) method encodes low-level features of the face image for face recognition across all ages [11]. Face images when represented in subspaces lose important data, which helps the algorithm in face recognition. The multi-kernel multi-metric learning method differentiates facial features effectively for accurate face recognition. The algorithm learns multiple features to eliminate varying results in face recognition [12]. Occlusion of face images affects the face recognition considerably. A new partial face recognition scheme recognizes images with a partial face. To produce accurate results from the partial face, detection of key points and local textural features play a major role. A point set matching method matches key points and textural features from the query image to the images in the gallery [13]. A matching algorithm evaluates the performance of three-dimensional holistic face recognition and different de-noising methods. A model method determines the face that is more prone to noise and determines the level of post-processing after the image acquisition phase for accurate face recognition [14]. Pose, illumination, and occlusion pose a major barrier for accurate face recognition. Local octal pattern [LOP] extracts various features from the query image and generates sparse code words. The indexing of images starts with most similarity in features and sparse words. The methodology section explains each phase of face recognition and retrieval.

3 Workflow

The proposed system uses various attributes for face recognition. The system is designed to work in both online and off-line mode. In online mode, the image captured by a mobile camera is processed in the cloud storage (Fig. 1).

The image taken from a mobile camera is saved in the local phone directory. Phone hardware typically consists of low memory space. The local octal pattern [LOP] algorithm needs memory resource which the phone has limited of, so the entire processing is done in the MATLAB drive. The image taken from a mobile camera is uploaded to MATLAB drive via the Internet. The feature extraction by local octal pattern method executes in the MATLAB drive. Multiple attributes such as pose, age, illumination, skin tone, and color are considered when performing face retrieval. These attributes vary for the same faces taken by different mobile cameras. Different phones have cameras with different resolutions, which pose a significant problem in face recognition and retrieval. In feature extraction, the image with different resolution will produce varying results, which will affect efficient face retrieval. For efficient face retrieval, the files in the MATLAB drive can be accessed anywhere by any device such as mobile, tablet, and computer. MATLAB driver connector enables users to upload files to the cloud. To access files on the cloud, a MATLAB drive connector is needed. When working with files on MATLAB drive folder, only the local copy of the file is changed and when the change is done the file will be automatically uploaded to MATLAB drive. During the upload process, no changes can be done to it. Only on uploading the file to the drive, the file can be deleted or accessed. To upload a file on a MATLAB drive, the particular file needs to be pasted on a local MATLAB drive folder. The MATLAB drive connector will synchronize the files automatically when a new file is detected. The updated file will be automatically synced between devices by MATLAB drive connector (Fig. 2).

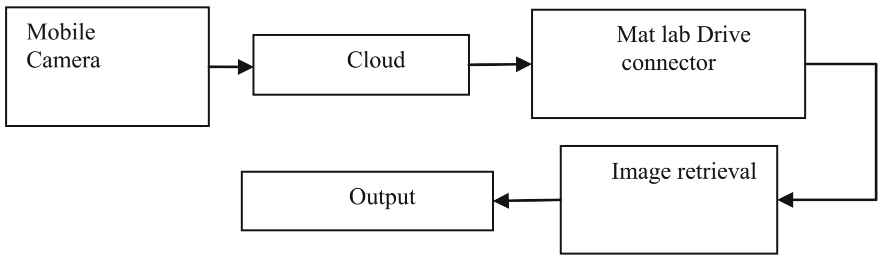


Fig. 1 Block diagram of workflow

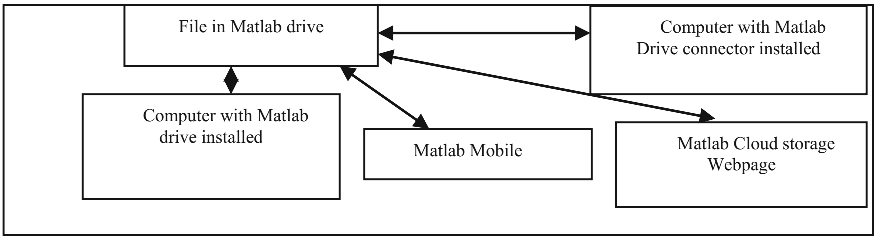


Fig. 2 MATLAB drive connector for the proposed algorithm access

4 Implementation of the Retrieval of Images

Traditionally, haar cascade method is used for face matching. This method uses feature matching to detect similar faces. The multiple attributes of the face are not considered, but only the features such as (i) the eye region are darker than nose and cheeks. (ii) The other feature is eyes are darker than the bridge of the nose. These features' similarity will vary for the same image taken at different resolutions. The illumination on the face image and the pose of a face will also greatly affect the outcome of face retrieval. To overcome these barriers, sparse code words are used for image retrieval. Sparse coding is used to represent data efficiently from over complete bases. Depending on the attributes from an input image, face retrieval is done. The more the attributes, the better the results. So to represent attributes effectively from an input image, embedded sparse code words are used. Sparsity is a technique used in embedded sparse code words to reduce the degeneracy. The local features are extracted, and similarity between the images and probe image is found using local octal pattern [LOP] algorithm. Initially, landmarks selected at points and matrices are formed around the center pixel. The matrix samples are compared with images in the datasets for similarity. Images with similarity selected, indexed, and ranked to group similar images for efficient face retrieval. The same method is applied to train the system in off-line mode. In off-line mode, the image results are produced from the available images in the gallery. The application is created specifically using MIT app inventor. The opened app will have a button to select the image and a window to display the selected image. Another button "Get result" is used to display the results with similar images from the gallery. Local octal pattern algorithm uses attributes such as skin tone, color, pose, and age to provide matching results. The outcomes of similar images are displayed on the next screen of the app (Figs. 3 and 4).

The PubFig database consists of images of celebrities taken at different poses with varying resolutions. Over 58,797 images of 200 people are taken out of which 30 images of 15 people are used for analysis. By use of LOP algorithm, the system was able to retrieve similar images by resolving barriers such as resolution, pose, age, illumination, skin tone, and color. The outcome of local octal pattern [LOP] algorithm feature extraction is compared with Local Binary Pattern [LBP] feature

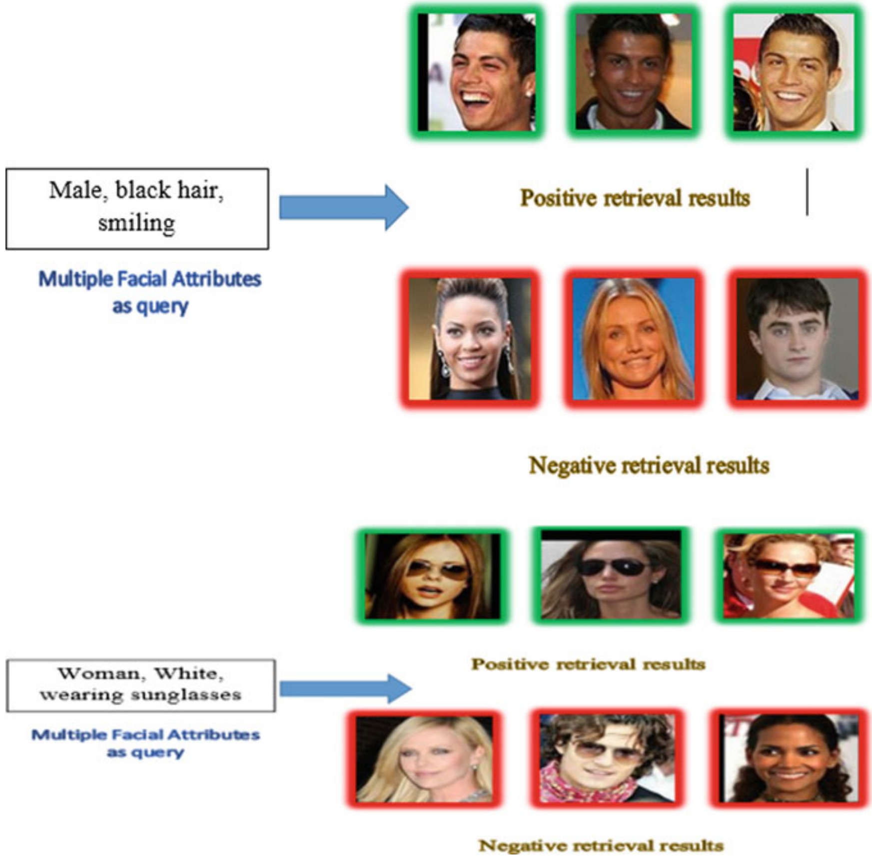


Fig. 3 Multiple attributes and retrieval results

extraction. Local Binary Pattern [LBP] uses a 3 × 3 grid space to match the features with other images, but local octal pattern [LOP] uses a threshold matrix of 5 × 7 grids for local feature extraction. Single attribute matching uses only certain checklist such as color, age, whereas multiple attribute matching compares skin tone, color, and illumination with the query image for face retrieval. The preprocessing step involves face detection, face landmark detection, and face alignment as shown in Fig. 5. For example, Fig. 6 shows the output of an aligned face after the preprocessing step. Figure 7 shows the feature extraction steps.

The LOP algorithm picks up a place to initially fix a matrix and converts the corresponding matrix to binary numbers. The process repeats for the adjacent pixels until all the pixels in the face are covered. The image initially converted to gray scale and the adjacent cells first-order derivatives calculated in all directions as shown in Eqs. (1), (2), and (3).

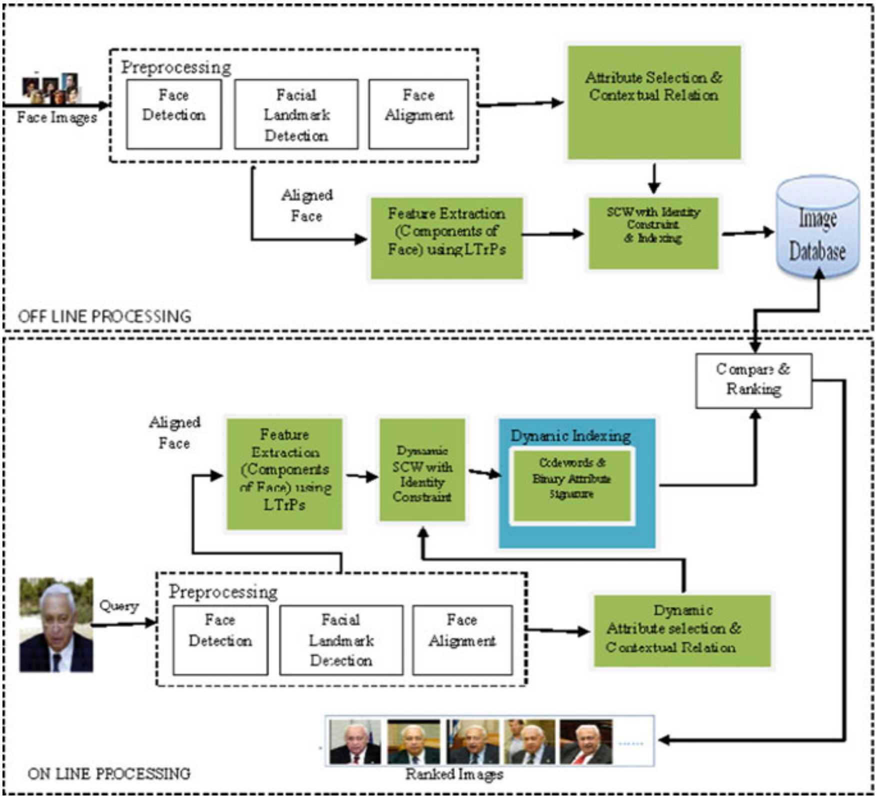


Fig. 4 System architecture in enhanced face image retrieval—multiple facial attribute and attribute-embedded sparse code words

$$\mathrm{I}^1_{0^\circ}(\mathrm{g_c}) = \mathrm{I}(\mathrm{g_h}) - (\mathrm{g_c}) \tag{1}$$

$$\mathrm{I}^1_{45^\circ}(\mathrm{g_c}) = \mathrm{I}(\mathrm{d}) - (\mathrm{g_c}) \tag{2}$$

$$\mathrm{I}^1_{90^\circ}(\mathrm{g_c}) = \mathrm{I}(\mathrm{g_v}) - (\mathrm{g_c}) \tag{3}$$

where gc, gd, gv, and gh represent vertical and horizontal directions. Figure 8 shows the output of feature extracted from the query image using local octal pattern algorithm.

The histogram is calculated for each image depending on the magnitude of the pixels as shown in Eq. (4).

$$M^1_{l(gp)} = \sqrt{(I^1_0 g^2_p) + (I^1_{45^0} g^2_p) + (I^1_0 g^2_p)} \tag{4}$$

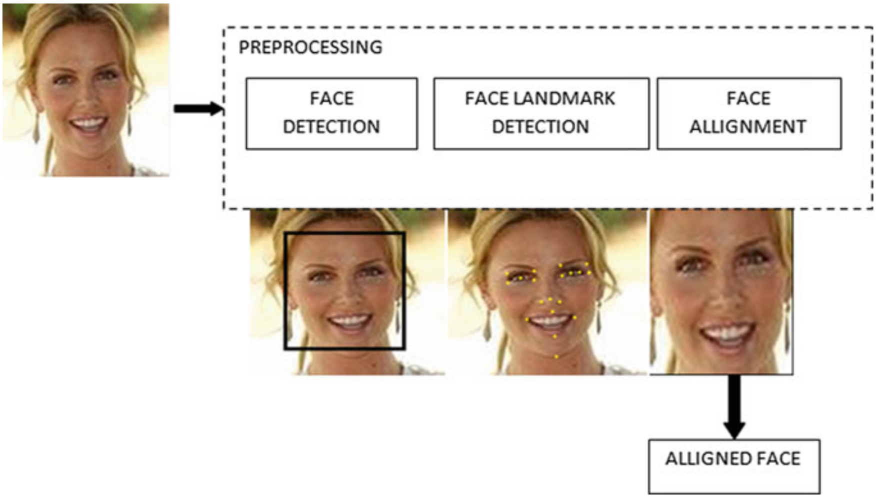


Fig. 5 Preprocessing steps

Fig. 6 Aligned face

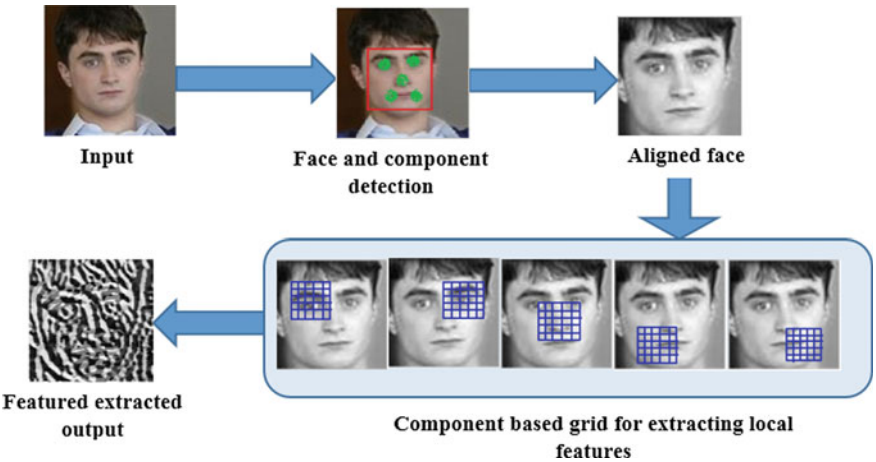


Fig. 7 Feature extraction steps

Fig. 8 Feature extraction using LOP

Each query image will have its attributes. This will help in determining the appropriate match and successful image retrieval. Each image is unique in some way; all the face image attributes will change depending on race, religion, gender, birth, and nativity. For successful image retrieval, multiple attribute dictionary learning for sparse coding is done.

4.1 *Multi-attribute Dictionary Learning for Sparse Coding (MASC)*

Sparse coding is to represent a signal as a linear combination of a few atoms in a given dictionary. Multi-attributed dictionary learning for sparse coding is proposed based on the unified distance measure of data and attributes. Instead of learning the dictionary for the entire database, can learn K-category dependent sub-dictionary D (1), …, D(K) denoted as D. Dictionary learning needs to be compact, reconstructive, and label-consistent.

Compact term. Compact terms evaluate the similarity in dictionary atoms by data distance and attribute similarity. Depending on the level of similarity, attributes are shared closer to the centroid. The compact term is defined as in Eq. (5)

$$C(D) = \sum_{K=1}^{K} \sum\nolimits_{\forall v_q \in D^{(k)}} d(v_q, \overline{v}^{(k)}) \tag{5}$$

$\overline{v}(k)$—centroid of atom in dictionary $D^{(k)}$, v_q—atom. An atom v_q is assigned to dictionary $D^{(k)}$ if it satisfies Eq. (6)

$$k^* = \arg\ min_k\ d(v_q, \overline{v}^{(k)}) \tag{6}$$

Reconstructive term. The reconstructive term is used to make dictionary selection with minimal reconstruction error during a training process as in Eq. (7)

$$R(D) = \sum_{k=1}^{k} \sum\nolimits_{\forall vq \in D(k)} \left\| v_q - \alpha_q^{(k)} \right\| 2 \tag{7}$$

An atom v_q is assigned to dictionary $D^{(k*)}$ if it satisfies Eq. (8)

$$k^* = \arg\ min_k \left\| v_q - D^{(k)} \alpha_q^{(k)} \right\| 2 \tag{8}$$

Label term. Label term emboldens dictionary atom labels to be similar. Label term is given as in Eq. (9)

$$L(D) = \sum_{k=1}^{k} * \left(\frac{1}{r} \sum_{i=1}^{r} \frac{N_{i,j}^{k}}{\sum_{j=1}^{ni} N_{i,j}^{k}} \right) \tag{9}$$

where $j* = \arg\max N_{i,\ j}^{k}$, J*—label type for attribute i, in dictionary k with the maximal number, number of labels with type j of attribute ai in dictionary k, $N_{i\ \ j}^{k}$—number of labels with type j of attribute a_i in dictionary k.

Optimization. The objective function for the multi-attribute dictionary learning and minimizing the object function would be the solution for the dictionary learning + problem given as in Eq. (10)

$$min_D C(D) + R(D) - L(D) \tag{10}$$

Proposed MASC Algorithm Steps
Minimizing the object function using K-method and clustering method is as follows:

i. *Calculate centroid for a cluster by using learned dictionary matrix.*
ii. *Data samples assigned based on the centroid.*
iii. *Centroids and cluster are updated in each iteration.*
iv. *The process terminates when there is no change in clusters.*

4.2 Dynamic Indexing and Retrieval

Dynamic indexing. Dynamically Weighted Attributed–Embedded Invert Index. From the sparse coding output, the images with the most feature, matching listed and the results indexed. Attribute importance score is then transformed using Linear Transformation in the Interval of Variation for current attribute. The transformed score is calculated using the formula as in Eq. (11)

Fig. 9 Image retrieval for a male input image

$$l_{hi} + \frac{(h_i - L_i)}{(H_i - L_i)} * (UB_i - LB_i) \tag{11}$$

i. h_i = Attribute importance score
ii. L_i = Lowest of attribute value importance scores for attribute i in data model
iii. H_i = Highest of attribute value importance scores for attribute i in data model
iv. UB_i = Upper bound of Interval of Variation for attribute i
v. LB_i = Lower bound of Interval of Variation for attribute i.

Attribute normalizes the attribute importance scores of all the attributes so that they sum up to number of attributes. The normalized values are the dynamic weights of each of the attributes depending on the input image and features extracted, the corresponding images with similarity indexed. Sparse code words for features and attributes are generated for face retrieval in online mode. Figure 9 shows the image retrieval for a male input image. Figure 10 shows the image retrieval for a female input image.

Dynamic Indexing Algorithm Steps
For dynamic indexing, VP tree (Vantage Point Tree) is constructed. The steps are given below

Fig. 10 Image retrieval for a female input image

Fig. 11 Aligned face (mobile)

Fig. 12 Feature extraction using LOP (mobile)

i. *Select point v(Vantage point) according to which is farthest from center of gravity.*
ii. *Calculate distance from point v to database point.*
iii. *Sort database point according to distance in ascending order.*
iv. *Calculate median distance.*
v. *All points below median distance area assigned to the left subspace of v, above median distance, occupy right subspace of v.*

The searching complexity is (O log2 N). Here, similarity of the region is considered as a distance between the feature vectors. Before constructing the VP tree, it must construct feature vector of all invariant region from each image in the database. At each leaf of VP tree, it is not only having feature vector but also the invariant region and database image from which it is extracted. In an off-line mode of face retrieval, the input images features and attributes are extracted, and sparse code words are generated to match appropriate images from the gallery. Figure 11

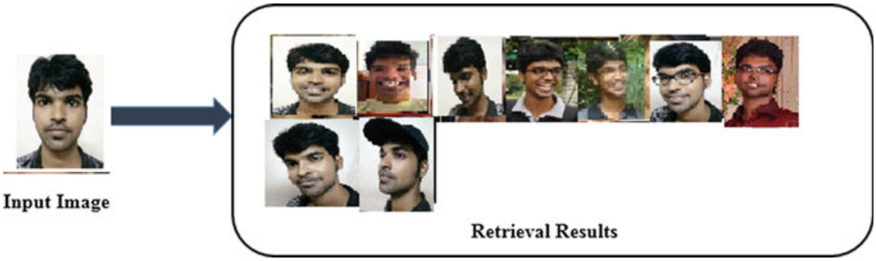


Fig. 13 Sample retrieval results (mobile)

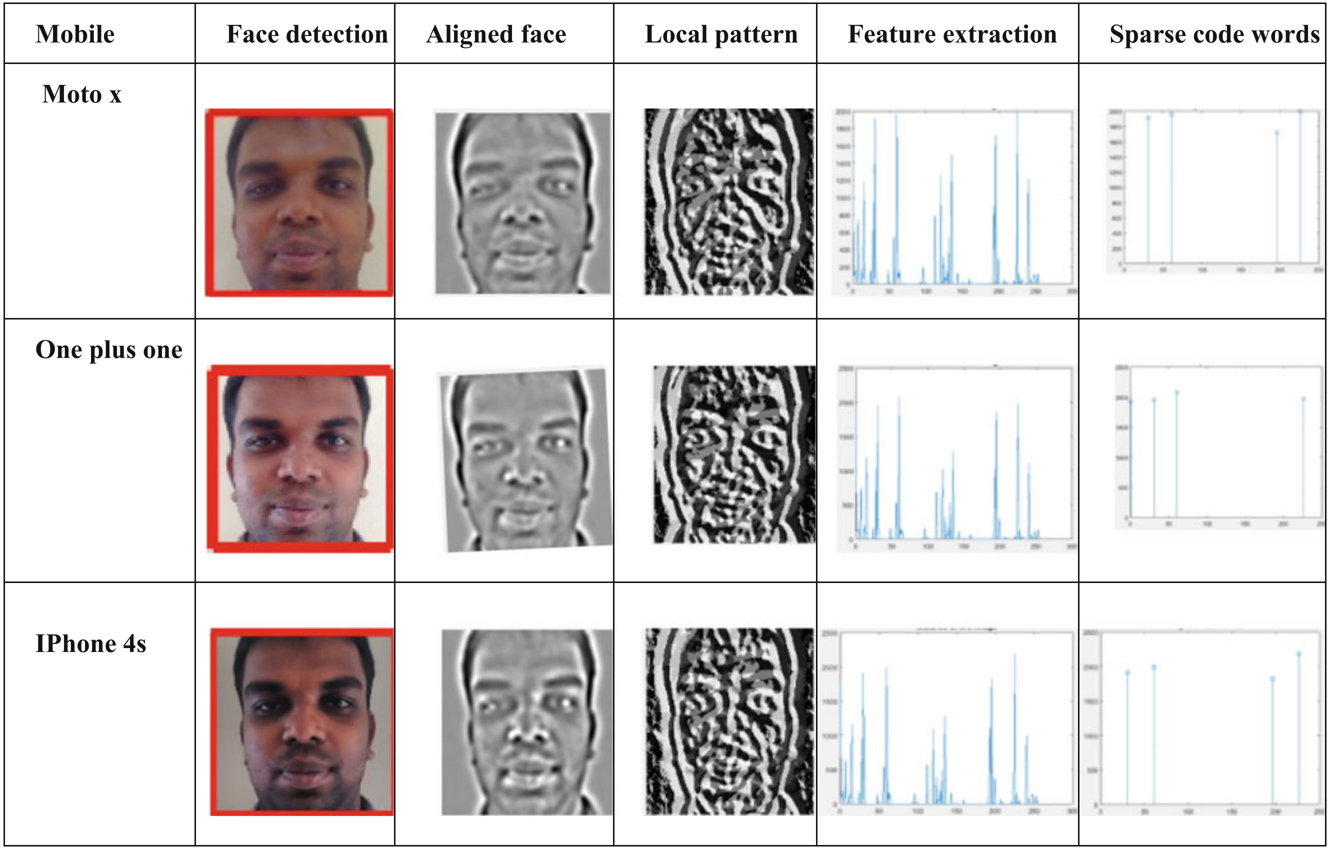


Fig. 14 Images of face taken with different cameras and their corresponding images

Table 1 Comparison of various megapixel mobiles

Multi-attribute for face	Various illumination			
	Moto x (%)	One plus one (%)	iPhone 4 s (%)	DSLR (%)
Male/wearing glass/mustache	80	82	84	83
Female/wearing earrings/brown hair	81	83	81	82
Female/teeth visible/black	80	80	83	82
Smiling/black hair/youth	83	82	83	82
Black/wearing lipstick/Asian	81	82	84	82

Table 2 Performance of multiple faces attributes

Attributes	PubFig (%)
Male/wearing cooling glass/mustache	22.5
Female/wearing earrings/brown hair	21.7
Female/teeth visible/black	22.9
Smiling/black hair/youth	23.8
Black/wearing lipstick/Asian	21.4

shows the aligned face (*mobile*). Figure 12 *shows the feature extraction using LOP* (*mobile*), Fig. 13 *shows the sample retrieval results* (*mobile*) (Fig. 14).

Their specifications are as follows: Moto × −10 megapixels @4320 × 2432 pixels with retrieval time is of 3 s with the proposed algorithm; the one plus one mobile with −13 megapixels @4128 × 3096 pixels with 4.5 s with retrieval time for proposed algorithm; and the iPhone 4s −8 megapixels @3264 × 2448 pixels of 3.3 s with retrieval time for proposed algorithm. DSLR 24.3 megapixels @3216 × 2136 pixels images are used for the proposed and existing algorithms. Table 1 shows the comparison of various megapixel mobiles, and Table 2 shows the performance of multiple faces attributes from PubFig.

The retrieval time varies with respect to the number of images in the database.

5 Conclusion

In this paper, a system with automatic, matching face retrieval is proposed. The result for query image is obtained based on matching attributes such as pose, age, illumination, skin tone, and color. The proposed system was able to retrieve images with matching attributes using local octal pattern [LOP] by overcoming the pointed out barriers. Comparative chart between different phones shows that retrieval time varies with respect to different phones. The proposed system was able to retrieve results with 96% accuracy than compared to Local Binary Pattern algorithm [LBP].

Acknowledgements The work was supported By Dr.N.R. Shanker., Chase Research and Development Solutions, Chennai, India.

References

1. Soldera, J., Alberto, C., Behaine, R., Scharcanski, J., Member, S.: Customized Orthogonal Locality Preserving Projections With Soft-Margin Maximization for Face Recognition, pp. 1–10 (2015)
2. Jian, M., Lam, K., Member, S.: Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition, vol 8215, no. c, pp. 1–14 (2015)
3. Uzair, M., Mahmood, A., Mian, A.: Hyperspectral Face Recognition With Spatiospectral Information Fusion and PLS Regression, vol. 24, no. 3, pp. 1127–1137 (2015)
4. Huang, Z., Member, S., Shan, S., Lao, S., Kuerban, A.: A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database, vol. 24, no. 12, pp. 5967–5981 (2015)
5. Moeini, A., Moeini, H., Member, S.: Real-World and Rapid Face Recognition towards Pose and Expression Variations via Feature Library Matrix, vol. 6013. no. c (2015)
6. Lu, J., Liong, V.E., Zhou, X., Zhou, J.: Learning Compact Binary Face Descriptor for Face Recognition, vol. 8828, no. c, pp. 1–16 (2015)
7. Gao, S., Zhang, Y., Jia, K., Lu, J., Zhang, Y.: Single Sample Face Recognition via Learning Deep Supervised Autoencoders, vol. 10, no. 10, pp. 2108–2118 (2015)
8. Qiu, Q., Chellappa, R.: Compositional Dictionaries for Domain Adaptive Face Recognition, vol. 24, no. 12, pp. 5152–5165 (2015)
9. Dutta, A., Günther, M., El Shafey, L., Marcel, S., Veldhuis, R., Spreeuwers, L.: Impact of Eye Detection Error on Face Recognition Performance, pp. 137–150 (2014)
10. Ding, C., Member, S., Xu, C., Tao, D.: Multi-task Pose-Invariant Face Recognition, vol. 24, no. 3, pp. 980–993 (2015)
11. Chen, B., Chen, C., Hsu, W.H., Member, S.: Face Recognition and Retrieval Using Cross-Age Reference Coding With Cross-Age Celebrity Dataset, vol. 17, no. 6, pp. 804–815 (2015)
12. Lu, J., Wang, G., Moulin, P.: Localized Multi-Feature Metric Learning for Image Set Based Face Recognition, vol. 8215, no. c, pp. 1–13 (2015)
13. Weng, R., Lu, J., Member, S., Tan, Y., Member, S.: Robust Point Set Matching for Partial Face Recognition, vol. 7149, no. c, pp. 1–14 (2016)
14. Emambakhsh, M., Gao, J., Evan, A.: Noise Modelling For Denoising and Three-Dimensional Face Recognition Algorithms Performance Evaluation, pp. 741–749 (2015)
15. Vasanth, K., Tena, J., Elanangai, V., Amuthan, A.: Cascaded algorithm for the elimination of impulse noise variants and artifacts in images, Vol. 9, no. 19, pp. 5779–5796 (2014)
16. Vasanth, K., Jawahar Senthil Kumar, V., Elanangai, V.: Unsymmetrical trimmed midpoint as detector for salt and pepper noise removal, Vol. 2, pp. 813–822 (2012)
17. Vasanth, K., Elanangai, V., Saravanan, S., Nagarajan, G.: FSM-Based VLSI architecture for the 3 $\times$ 3 Window-Based DBUTMPF algorithm, Vol. 398, pp. 235–247 (2015)

Survey of Different Approaches Used for Food Recognition

Sandhya Arora, Gauri Chaware, Devangi Chinchankar, Eesha Dixit and Shevi Jain

Abstract Food recognition is an ever-growing field gaining rapid momentum in the past couple of years. Various approaches have been implemented to get accurate results by correctly identifying the food item. Traditional methods like the implementation of neural networks, SVMs, HMMs utilizing hand-crafted features of the large data-sets of food images are one way of developing food recognition systems. To improve the accuracy, modern methods using newer concepts of convolutional neural networks and deep learning which avoid the use of hand-crafted features are being implemented to build even better food recognition systems. These newer methods require huge data-sets of images of food items to work with to obtain good results. Besides approaches based on image recognition, other innovative images are also being explored for recognizing food images. Food items are being recognized using the cutting sounds, acoustic sensors, electronic tongues and so on.

Keywords Convolutional neural network · Deep learning · Food recognition Image recognition · Neural network

S. Arora · G. Chaware · D. Chinchankar · E. Dixit (✉) · S. Jain
Cummins College of Engineering, Pune, India
e-mail: eesha.dixit@cumminscollege.in

S. Arora
e-mail: sandhya.arora@cumminscollege.in

G. Chaware
e-mail: gauri.chaware@cumminscollege.in

D. Chinchankar
e-mail: devangi.chinchankar@cumminscollege.in

S. Jain
e-mail: shevi.jain@cumminscollege.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_54

1 Introduction

Food is the cornerstone of one's life. In this increasingly diet conscious, curious world, food recognition has been rapidly gaining popularity over the past couple of years. Food recognition is being implemented for various reasons ranging from monitoring the calorie intake to simply identifying the food item and getting additional information regarding the same. Food recognition systems are being developed for quite some time now and various approaches have been used to implement these systems. Multiple approaches are being developed using different technologies in order to attain maximum possible accuracy over a broad range. Not only are the systems being developed aiming for high accuracy rates in correctly identifying the food item but also aiming to cover a wide range of food items and cuisines.

These food recognition systems being developed are mainly working in either of the two ways: using an image of the food item as an input or, without using the image of the food item as an input but using the cutting sounds, acoustic sensors, electronic tongue and so on as an input to recognize the food item. Either a classical approach or a modern approach can be used for recognizing the food-item from the image.

2 Literature Survey

2.1 Classical Approach for Food Recognition from Images

Classical approaches for food recognition from images require explicit selection, extraction and encoding of features which then are consecutively passed to classifiers.

Feature selection and feature extraction may be done using **color histograms**, **SIFT and SURF or histogram of oriented gradients** (**HOG**). Color is one of the prominent features of food images, but it can be said to be a trivial feature selection approach for food identification due to the overlapping of color attributes of different food items. The red, green and blue channel values of the image are used as the feature vector and the Gaussian distribution can be used to represent the color data, as stated by Martin et al. [1]. SIFT and SURF are low level feature extraction methods which work on extracting primitive features like edges, corners and so on. SURF is an extension of SIFT and is faster than SIFT [2]. HOG can be said to be like SIFT as both algorithms use gradient histogram to describe the patterns found in the image. But HOG can describe local patches much faster [2]. In the work by Kawano et al. they found HOG to be about 12 times faster than SIFT and SURF [2]. Many times, combinations of one or more such methods are adopted as features (feature fusion).

Low level features like SIFT, SURF, Color descriptors and Texture descriptors (Gabor texture [2]), which are often hand crafted, are further encoded with feature encoding [2] methods like Bag-of-Features (BoF) [3, 4], Locality Constrained Linear Coding (LLC) [5] or Fisher Vectors (FV) [6]. BoF stores the feature information as a histogram. The four typical steps in BoF are: key-point detection, local descriptor extraction, codebook learning and histogram representation [3]. Relative positioning of ingredients plays a pivotal role in food recognition but BoF fails to capture such information [7]. FV is a method of representing a set of local features [2]. It proves to be better than BoF as it reduces computational cost.

Few of the approaches that use the features extracted using the methods mentioned above as an input to the classifier are as follows.

Mining descriptive components with super-pixel and random forests: In food recognition, location attributes play a vital role in the recognition task. Thus, this task calls for methods that can capture local patterns. For extracting discriminative image regions, a method based on Random Forests (RFs) proposed by Bossard et al. can be used [8] which works as follows: the searching for the discriminative parts is restricted to the image patches aligned with super-pixels [8, 9] which helps in extracting more effective visual features. An RF is then used to hierarchically cluster super-pixels of the training set which are then used to train the component models. Lastly, for classification, all the super-pixels of the input image are ranked using the component models. Afterwards, a multi-class SVM with spatial pooling can predict the final class.

Using pair-wise local features: Relative positions of food ingredients play a crucial role in food recognition [7, 10]. For example, bread slices stacked upon one another have a higher probability of being classified as a sandwich than two bread slices placed side by side [7]. To understand these spatial relationships, pairwise local features based on distance and orientation between ingredients can be used. A multidimensional histogram can be used to represent how these features are distributed, as stated by Yang et al. [10]. SVM is further used as a classifier.

Multi-view recognition using multiple kernels: He et al. [7], considers the food image from multiple view points before classifying it through a multi kernel SVM. This consideration is important because a single viewpoint would not be able to capture all the information of food [7]. Under each viewpoint, He et al. used multiple kernels to classify food from different feature channels. Such an implementation of multi-view compared with other baseline methods proved to be superior. The proposed "DietCam" in [7] gave 90% accuracy for general food items.

2.2 *Deep Learning Based Food Recognition*

Traditional food recognition methods make use of hand-crafted features [11, 12]; the problem with these features is that they may be suitable to recognize a certain food type well but may fail to generalization [11]. Deep learning has an advantage over this. It is a collective term for algorithms having a deep, multi-layered architecture that solves complex problems by learning more and more information as the network deepens. The most important characteristic of a Deep Convolutional Neural Network (DCNN) is that, better image features are automatically extracted via training [12]. Convolutional kernels are responsible for extracting these features. A DCNN combines all the typical object recognition tasks like local feature extraction and encoding and learning [13].

The number of iterations [14] the network is trained on and the number and size of convolutional kernels largely affects the network's accuracy [12]. DCNN will give the most accurate results when trained on a large dataset, but this poses limitations on its practical usage. To overcome this difficulty, Yanai et al. proposed two methods namely pre-training and pre-training with fine tuning. A pre-trained DCNN is one which is trained on a large-scale dataset like ILSVRC or ImageNet and then which does the feature extraction for small-scale data. This basically enables transfer learning [13]. In the first method proposed, the pre-trained DCNN is used only for extracting features and then later passed on to a classifier like SVM. In the second method, fine tuning of parameters is done, which is basically making the network adaptable to the small-scale data set which is initially trained on a large-scale data set. In this case, the DCNN is used both for extracting the features and as a classifier. According to Oquab et al., it is important that the training data for pre-training and fine tuning are related to better the performance [15]. Yanai et al. found out that a pre-trained DCNN which has undergone fine tuning drastically improves the classification performance [13].

2.3 Data-Sets

Different data-sets used by various different developed food recognition systems are as follows.

- UEC-256 and UEC-100 are dataset of Japanese dishes. These data sets consist of more than 100 images for each category and food location within each food photo. UEC-100 includes 100 food categories with a total of 8643 images with around 90 images in each category. UEC-256 encompasses 256 categories with about a total of 28,375 images with around 110 images in each category [9, 11, 13].
- Food-101 dataset comprises of 101 categories and is publicly available. For each food class, there are about 1000 images. It offers food type information for each image. Most of the images in this data set are popular western food images [8, 11, 14].
- The Pittsburgh Fast Food Image Dataset (PFID) is a dataset of American fast food images. The data was collected by obtaining three instances of 101 foods from 11 popular fast food chains. It is not widely used because it requires a significant amount of cleaning up [7, 12, 16].

2.4 Food Recognition Using Techniques Other Than Image Recognition

The solution to food recognition does not only pertain to recognizing the food item from images. Working upon this thought, many other innovative ways of recognizing food items were developed.

- *Drink recognition using an electronic tongue*: Liu et al. used the measurements of an electronic tongue to perform the recognition of two types of beverages by using an electronic tongue-based approach. Features considered are the data from the 7 sensors on the tongue which are then inputted to a classifier. Liu et al. [17], found out that Random Forests outperformed Neural Networks and SVMs in classifying the food items.
- *Using acoustic sensors*: The study in [18] works on the basic idea of recognizing food items based on chewing information. As stated by Bi et al. the process of food recognition consists of several steps: framing of acoustic signals, processing of the frames by HMMs to detect chewing events and detecting of fluid intake by swallowing events, then each such event is processed to extract several key features in both time and frequency domain and lastly a light-weight decision tree based algorithm is to recognize the type of food intake [18].
- *Using cutting sounds*: In "CogKnife", in [19], food is identified from the cutting sounds. Kojima et al. used a small microphone on a knife to record the cutting sounds of food and then extract the spectrograms from these sounds to be used as features to train a classifier. The study found that SVM approach was the most accurate compared to K-NN and CNN [19].

3 Proposed Approach

There are many food recognition systems available in today's world catering to various cuisines; but there exists no well-known system that recognizes Indian cuisine. India is famous for its tasty delicacies. Tourists visiting India get a whiff of the diverse culture through its food. This system will most certainly aid these tourists who are hungry to get to know India. Besides tourists, the locals who really enjoy food and want to explore the widths and depths of Indian cuisine and those who want to add to their repertoire new Indian food items' recipes will definitely find use in the system we propose. The proposed system will be inputted with image of the food item to be recognized. The image will then undergo some image

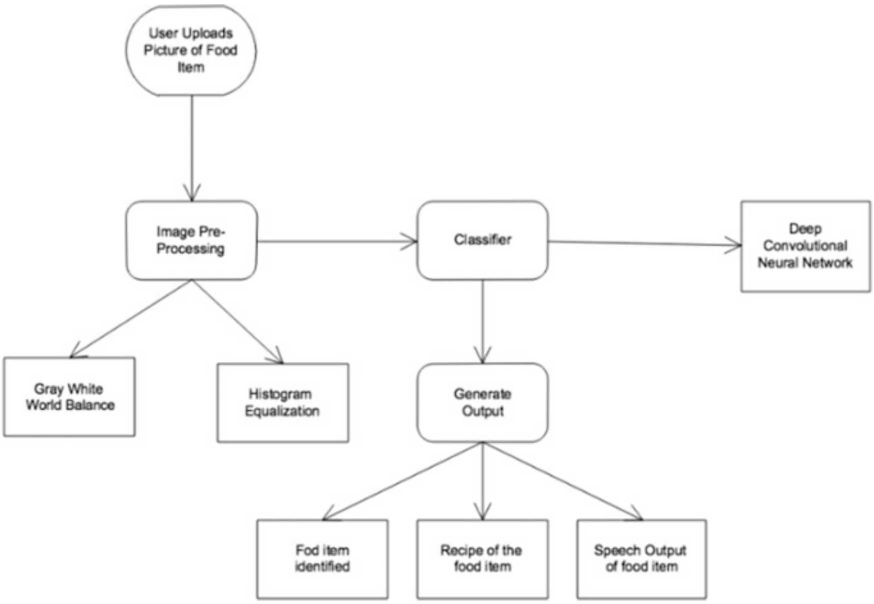


Fig. 1 Proposed Indian food recognition system's block diagram

processing like that mentioned by Yu et al. [14] to obtain a balanced image to input to the classifier which will be a DCNN. The DCNN will output a class label which will be the name of the recognized food item. Along with the food item's name, the proposed system will also output a pronunciation of the food item in the native language and will provide the user with a recipe for the food item along with a calorie content (Fig. 1).

4 Future Enhancements

In the future, the proposed frameworks can also be implemented on mobile devices. As the processing power of the mobile phone is significantly lower comparing to the GPUs, the amount of pre-trained DCNN parameters can be reduced for this [13]. Most of the times a pre-trained model is used, however, the use of GPUs can be done to train the network, this would require certain changes to the network and may take a long time but by finding the best network structure and parameters that fits the problem well, better results can be achieved [14]. Construction of 3D structures by investigating multiple images to have more accurate estimate of the gram amounts of the food can be done [1]. Users' food history, GPS location data and time information and increasing the number of food categories could make the system more practical. In the future, users could be allowed to add new training images and food categories [2].

5 Results and Discussions

Different approaches yielded different results, that is, each food recognition system varied in their accuracy rates. These accuracy rates were contingent to the technology used, data-set used and many other factors. Various results yielded are as follows.

5.1 Classical Approaches' Results (Using Images)

Employing the random forests approach, Bossard et al. implemented methods yielding 54.40% and 58.36% average accuracy for the restricted and full training set, respectively on classes having 50 components each [3]. Employing a combination of bag of color features and texture features, Sasano et al. achieved recognition rates as depicted in the graph (Fig. 2).

Categorizing the food items into different categories based on the difficulty level determined by the number of dominant ingredients, He et al. compared their approach which uses multi-kernel SVM with other popularly used approaches, with respect to the precision obtained. It could be seen for each of the 6 difficulty categories, the approach suggested by He et al. [7] outperformed in all cases.

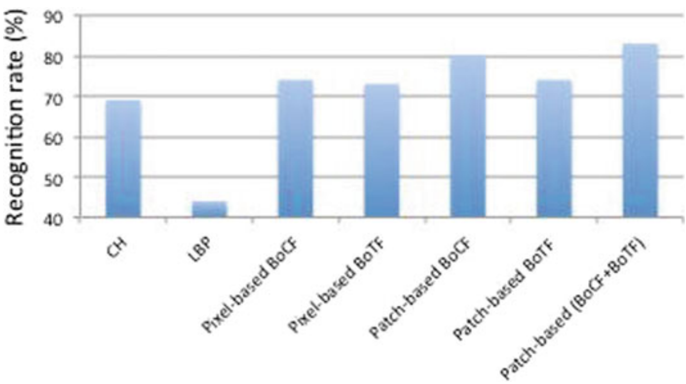


Fig. 2 Comparison of the results of the combined BoCF and BoTF and other features by Sasano et al. [3]

5.2 Modern Approaches' Results (Using Images)

Kagaya et al. compared their approach of food recognition which employed convolutional neural networks with a baseline method that used SVMs. For the experiment using SVMs, they randomly selected the same number (1,200) of food images and non-food images, with 1,000 being used for training the SVM and the remaining 200 being used for testing. They computed the average result over 15 trials [12] (Table 1).

Yu et al. compared various models for food recognition. It could be seen that the approaches that used modern methods outperformed classical methods. While Inception-ResNet (full layers training) achieved 72.55% top-1 accuracy on the Food-101 data-set, classical methods achieved a top-1 accuracy below 50% on the same dataset. Inception-ResNet (full layer training) achieved 91.31% top-5 accuracy.

Liu et al. implemented a new deep learning-based food recognition. The deep learning based algorithms they implemented showed substantially better recognition accuracy. With an increase in the number of iterations, the accuracy increased as well. Starting from 5000 iterations yielding 88.7% top-5 accuracy, on having 60,000 iterations, the top-5 accuracy plummeted to 94% [11].

Yanai et al. proposed a food recognition system using DCNN with pre-training and fine tuning. While the baseline methods attained a top-5 accuracy of 79.88% on the UEC-FOOD100 data-set, DCNN combined with pre-training and fine-tuning methods attained a top-5 accuracy of 94.39% on the same data-set [13].

5.3 Results of Other Approaches (Without Using Images)

- Electronic tongue: In this approach, the data used for the system was split into 4 datasets: the 1st dataset was used for recognition of orange juices of 2 concentration levels & had 216 samples in total. The other 3 were used to recognize Chinese vinegars of 2 quality grades, 4 kinds of materials and 12 different brands. They consisted of 432, 414 and 108 samples, respectively [17]. It was observed that neural networks achieved a classification accuracy of 86.68%, SVM achieved 66.45% and random forests achieved 99.07% [17].
- Using acoustic sensors: Experimental data were collected from 12 subjects by Bi et al. 7 different types of food were used to evaluate their systems. In total, they collected 171 samplings which were composed of bite, chewing and swallowing

Table 1 Comparison between baseline method and CNN method [12]

Method	Accuracy (%)
Baseline	89.7 ± 0.73
CNN	93.8 ± 1.38

events. On an average, the system proposed by Bi et al. yielded an accuracy of 86.6% [18].

- Using cutting sounds: In 'CogKnife', the system proposed by Kojima et al., they evaluated the accuracy of their technique and verified the best set of parameters and performed leave-one- out and 20-foldout validation by using six types of foods. They observed that SVM approach provided them with the highest classification accuracy of 95% followed by CNN of 89% and lastly, K-NN of 83% [19].

6 Conclusion

In conclusion, food recognition, as a domain, is a promising field gaining more and more importance with each passing day. Although multiple food recognition systems have been developed employing traditional, modern, hybrid or completely different approaches, the crux of the problem remains the same- extending the scope to cover larger varieties of food-items, to include as many cuisines as possible and to try and improve the accuracy rates. The field of food recognition will keep evolving and developing till desired optimality and accuracy is attained.

References

1. Martin, C., Kaya, S., Gunturk, B.: Quantification of food intake using food image analysis. In: 31st Annual International Conference of the IEEE EMBS, Minneapolis, Minnesota, USA, 2–6 Sept 2009, pp. 6869–6872 (2009)
2. Kawano, Y., Yanai, K.: FoodCam: a real-time food recognition system on a smartphone. Multimedia Tools Appl. **74**, 5236–5287. Springer (2015)
3. Sasano, S., Han, X., Chen, Y.: Food recognition by combined bags of color features and texture features. In: 9th International CISP-BMEI, Datong, China (2016)
4. O'Hara, S., Draper, B.: Introduction to the Bag-of-Features Paradigm for Image Classification and Retrieval (2011)
5. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: IEEE CVPR, pp. 3360–3367 (2010)
6. Perronnin, F., S´anchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: ECCV, pp. 143–156. Springer (2010)
7. He, H., Kong, F., Tan, J.: DietCam: multi-view food recognition using a multi-kernel SVM. IEEE J. Biomed. Health Inf. **20**, 848–855 (2016)
8. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101—mining discriminative components with random forests. In: ECCV, Zurich, Switzerland, pp. 446–461 (2014)
9. Zheng, J., Jane Wang, Z., Ji, X.: Superpixel-based image recognition for food images. In: IEEE CCECE, Vancouver, Canada, 15–16 May 2016, pp. 1–4
10. Yang, S., Chen, M., Pomerleau, D., Sukthankar, R.: Food recognition using statistics of pairwise local features. In: IEEE CVPR, San Francisco, USA, pp. 2249–2256 (2010)

11. Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y., Chen, S., Hou, P.: A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. IEEE Trans. Serv. Comput. (2017)
12. Kagaya, H., Aizawa, K., Ogawa, M.: Food detection and recognition using convolutional neural network. In: 22nd ACM International Conference on Multimedia, Orlando, Florida, Nov 2014, pp. 1085–1088
13. Yanai, k., Kawano, Y.: Food image recognition using deep convolutional neural network with pre-training and fine tuning. In: IEEE ICMEW, Turin, Italy, 29 June–3 July 2015, pp. 1–6
14. Yu, Q., Mao, D., Wang, J.: Deep Learning Based Food Recognition
15. Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Learning and transferring mid-level image representations using convolutional neural networks. In: IEEE Computer Vision and Pattern Recognition (CVPR), Columbus, USA, pp. 1717–1734 (2014)
16. Baxter, J.: Food recognition using ingredient-level features. http://jaybaxter.net/6869_food_project.pdf (2012)
17. Liu, M., Wang, M., Wang, J., Li, D.: Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and chinese vinegar. Sens. Actuators B Chem. (2013)
18. Bi, Y., Lv, Y., Song, C., Xu, W., Guan, N., Yi, W.: AutoDietary: a wearable acoustic sensor system for food intake recognition in daily life. IEEE Sens. J., **16**(3), 806–816 (2016)
19. Kojima, T., Ijiri, T., White, J., Kataoka, H., Hirabayashi, A.: CogKnife: food recognition from their cutting sounds. In: ICEMW, Seattle, WA, USA, July 2016

A Novel Deep Learning Framework Approach for Natural Calamities Detection

Rahul Nijhawan, Megha Rishi, Amit Tiwari and Rajat Dua

Abstract The term “natural calamities” has now emerged to be a global threat all over the world because of their steady mayhem and chaos in the form of human and economic loss. As these catastrophic events increase in severity and destruction, their classification based on their wide diversity is an important task from the disaster management aspect. However, it is yet an unpursued and unexamined topic in the branch of computer vision. This paper proposes a novel framework to detect natural calamities and thereby classify them in accordance with their class. A miscellany of ten different natural disasters, namely avalanche, cyclone, drought, earthquake, landslide, thunderstorm, tsunami, tornado, wildfire, and volcano was considered for classification. The framework solely relies on a hybrid of the convolutional neural network (CNN) for feature extraction. Because of the unavailability of a scrupulous dataset, a new dataset was constructed for testing the capability of the framework. The model performs exceptionally better when compared with other state-of-the-art algorithms (SVM, KNN, RF, and CNN + SVM). The results demonstrated that the model performs considerably better than average human intelligence in terms of different calamity and is able to recognize different classes of natural disasters with an accuracy of 82.23%.

Keywords Convolutional neural networks · Deep learning · Natural calamities Feature extraction

R. Nijhawan (✉)
Indian Institutes of Technology Roorkee, Roorkee, India
e-mail: rahulnijhawan2010@gmail.com

M. Rishi · A. Tiwari
College of Engineering Roorkee, Roorkee, India
e-mail: megharishi5396@gmail.com

A. Tiwari
e-mail: tiwariamit6250@gmail.com

R. Dua
Graphic Era University, Dehradun, India
e-mail: rdua3561@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_55

1 Introduction

A natural calamity is a catastrophic phenomenon arising out of the geological variations in the earth. These phenomena prove to be fatal for both human life and resources and in the recent time, many life-risking catastrophic events have resulted in major economic damage, some examples are hurricane, floods, typhoon, droughts, and other, therefore, the development of a robust classification framework is of utmost importance as it can be of great help for surveillance purpose in catastrophic prone regions, satellite integrated with the image classification query system can be used for cartography, oceanography, geology, biodiversity planning, and many other purpose. However, due to a wide diversity of natural calamity and despite the several attempts made at image recognition, recognition performance is not yet remarkable. In the recent years, natural calamity has been an integral part of the discussion and its classification has been carried out using several state-of-the-art algorithms. For obtaining high-level conceptual information such as the vision of an object, it is requisite to convert the raw-level input to a higher more "relevant" depiction that can reinforce the image detection process [1]; however, a remarkable effort is required to achieve the same. The traditional algorithms are best suited for a limited amount of dataset but for a considerable amount of a large data the performance degrades. Deep learning, gaining an immense popularity among the researchers within a short period, is portrayed as a substantial tool for image classification. It is just an emerging property of a neural network that contains subsequently more layers and being well estimated that accuracy of image recognition comes out to be significantly large if you heap on multiple neurons covering many number of layers. The traditional setup is to extract handcrafted features and then implements the state-of-the-art techniques. On the contrary, deep learning schemes also optimize the features that are extracted and subsampling layers give better results by the use of CNN and autoencoders. Large network depth is obtained as an output from inconsiderable amenable fields of convolutional winner-take-all neurons, followed by sparsely connected layers, out of which the training is provided to only the winner neurons [2]. Deep learning uses multilayer processing that reduces the time and provides better accuracy as in resemblance with state-of-the-art algorithm. It uses calculation models that contribute to multiprocessing layers of abstraction and have composed better results in the case of image and video recognition as compared to conventionally used methods [3]. It recognizes advanced structures in large datasets integrated with backpropagation algorithm that indicates how the internal parameter is changed which is used to figure out the representation in each layer from representation in the previous layer [4]. Although a considerable amount of work has been done in detection and recognition of natural hazards, the classification of a wide diversity of natural calamity is still recondite. Our approach deploys CNN for the classification and recognition of a wide category of the catastrophic events, namely avalanche, cyclone, drought, earthquake, landslide, thunderstorm, tsunami, tornado, volcano, and wildfire. In this paper, we have proposed a novel DL + SVM integrated framework approach composed of the ensemble of CNN. This ensemble framework has been hybrid with SVM model for classification. The proposed framework outperformed the state-of-the-art algorithms. Accuracy assessment has been performed by means of several statistic measures (sensitivity, specificity, and kappa coefficient).

The paper is organized as follows: Introduction (Sect. 1), Data Collection (Sect. 2), Methodology—classifiers, proposed approach (Sect. 3), Results and Discussion (Sect. 4), and finally Conclusion (Sect. 5).

2 Data Collection

In order to acquire an appreciable accuracy via deep learning, we provided a substantial dataset to our model. Because of the unavailability of an open-source dataset for natural calamities, we constructed a dataset of the images manually, the images were taken from the disaster-prone regions where the natural disaster is more likely to strike. Avalanche: The Arctic and temperate region of the Alps mountain range. Cyclones: The northwest Australian coastline between Exmouth and Broome. Drought: The most severe drought hit region in India, namely Maharashtra, Odisha, and Rajasthan. Earthquake: Countries falling within the seismic zone, Turkey, Nepal, and Philippines. Landslide: In the Appalachian Plateau Province and Fayette Country. Thunderstorm: The southern region of the United States, namely Arkansas, Alabama, Mississippi, and other prone regions. Tornado: The southern region of the USA, the Midwest, The Great Plains, and The Mississippi Valley, that is vulnerable to tornadoes including Alaska and Hawaii. Tsunami: In the coastal regions of Albania, Chile, China, Indonesia, and Pakistan. Volcano: The countries that have undergone numerous volcanic eruptions, Philippines, Ethiopia, Japan, and Indonesia. Wildfire: Santa Cruz Mountains, the part of the Pacific coast ranges in the central and northern California. The dataset comprises a diverse class of around 150–200 images (Fig. 1 and Table 1) approximately which was fine-tuned on the pre-trained AlexNet architecture.

The dataset was further divided for training and testing purpose, 70% of images were selected on a random basis to construct the training set and the remaining 30% images were used to create the testing set.

3 Methodology

3.1 Support Vector Machine

SVM is a supervised machine learning algorithm satisfying the purpose of classification and regression. SVM training algorithm is mainly in use for constructing algorithms allocating new example to the probabilistic binary linear classifier [5–7].

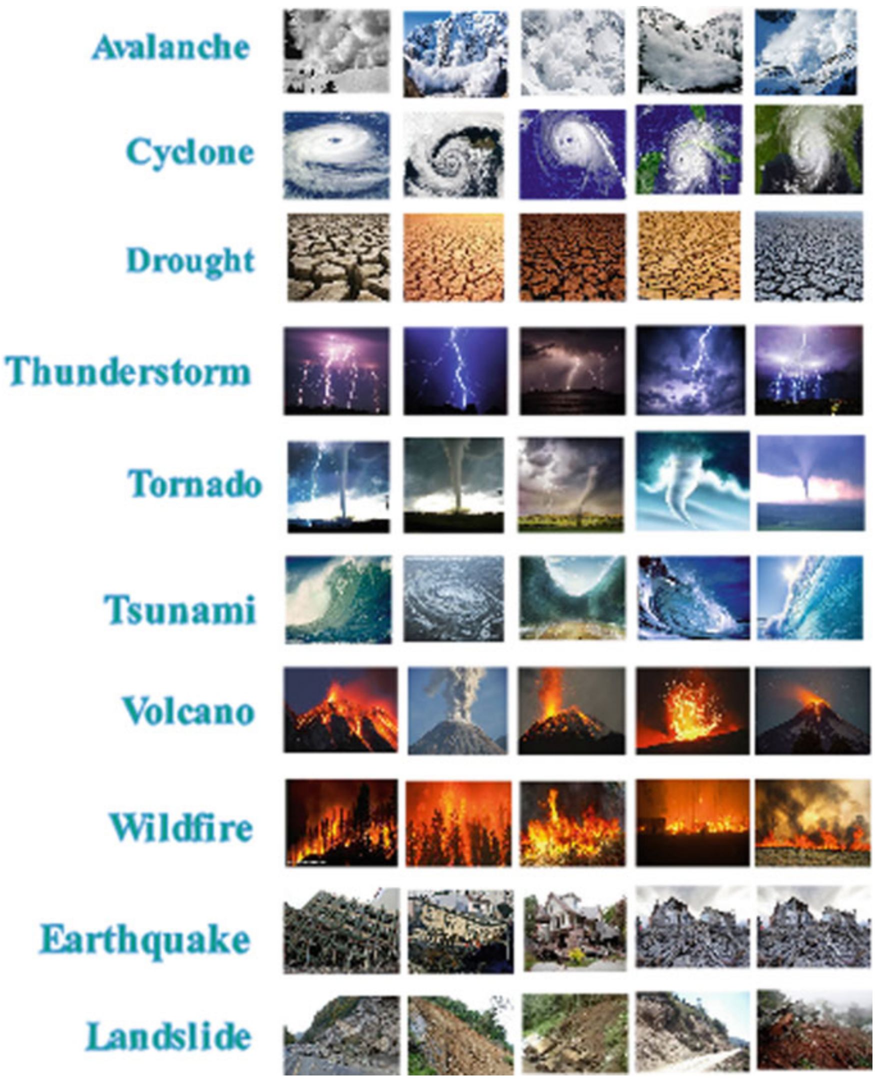


Fig. 1 Dataset samples for different natural calamities

3.2 K-Nearest Neighbor (KNN)

It is a nonparametric method used for classification and regression. The output is dependent on whether KNN is used for regression or for classification purpose. In it, object classification is done in accordance with the majority of K neighbors [8].

Table 1 Table shows the class of natural calamity and the total

Type of calamities	Number of images
Earthquake	143
Cyclone	165
Drought	127
Earthquake	175
Landslide	122
Thunderstorm	128
Tornado	187
Tsunami	109
Volcano	176
Wildfire	186
Total	**1518**

Numbers of images present corresponding to it.

3.3 Random Forest (RF)

It is a group of learning method used for the purpose of regression, classification, as well as other tasks. In a decision tree, that are present in groups are comparable to weak learners. While working with the approach of decision trees, input moves from the top traversing down the tree drawing out to be into consequently smaller and smaller sets [9, 10].

3.4 Convolutional Neural Network (CNN)

It is a classification of artificial neural networks that have been prosecuted to scrutinize visual imagery. CNN uses a variation of multilayer perceptron designed to require minimal processing. It compares image by breaking the image into overlapping images, and then each image is fed into a small neural network. It is well estimated that deep learning has given the best results in terms of accuracy factor provided that a large dataset is considered. Previous works [11, 12] have proved that CNN has outperformed other state-of-the-art methods.

3.5 Proposed Architecture

Pre-processing. In order to optimize the robustness of our framework, we increased the size of the dataset. The idiosyncrasy of our dataset proved to be a constraint to our framework; this was followed by random rotation and shifting of the images which eventually help us to ameliorate the generalization of our framework. This procedure is known as data augmentation.

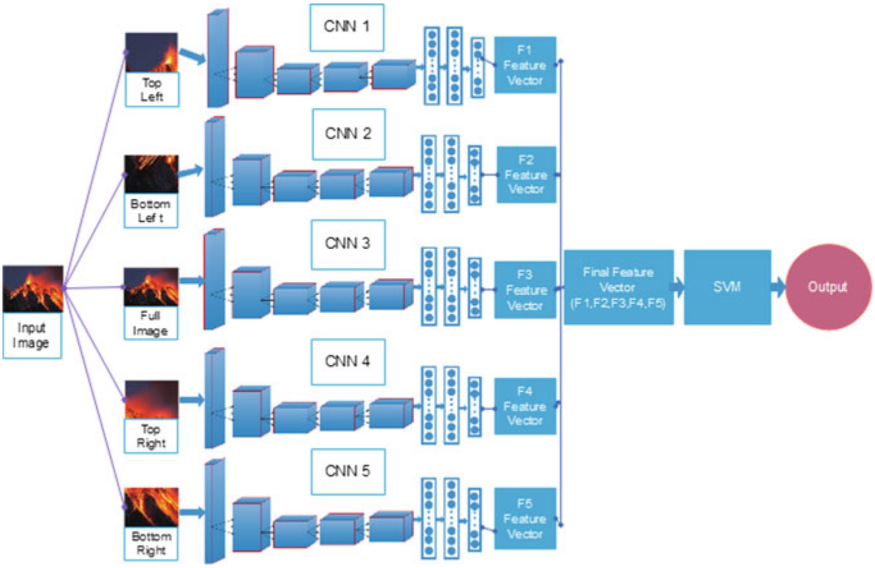


Fig. 2 Proposed DL framework for classification of natural calamities

Proposed Framework. Our proposed approach uses the deep convolutional neural network for identification and classification of ten different classes of natural calamities. Our model consists of five individual CNN grouped together, following are the input assigned to the CNN architecture, CNN 1: Top left of the image, CNN 2: Bottom left of the image, CNN 3: Full image, CNN 4: Top right of the image, CNN 5: Bottom right of the image. Each CNN model exploits a pre-trained AlexNet [13] architecture on ImageNet dataset which is finely tuned on our constructed dataset. Figure 2 shows our proposed deep learning framework with a hybrid of five CNN.

Post-processing. Over and above in comparison with raw output after labeling by convolutional neural network, we also require post-processing for cleaning up the result. Post-processing is referred for improving the quality of classified output and removal of noise. It involves three steps:

Filtering. It involves the filtering of classified output by using majority filter. This step is useful in removing of noise or isolated pixels.

Smoothing. It involves smoothing class boundaries and clumping classified output further by using building clean tool. This step involves smoothing the ragged class boundaries and clumping the classes.

Generalization. It involves classifying output by removing small isolated regions by using region group, nibble tools, and set null. This step involves the reclassification of small isolated regions of pixels respectively to the nearest classes.

4 Results and Discussion

We have made use of distinct scenarios for the optimization of results. Further, we have shown four different scenarios. Our aim is to select the best scenario that results in optimal training and accuracy.

4.1 Experimental Scenarios

Scenario 1 (**2CNNs + RF**). In the first scenario, we have used an ensemble of 2 CNN for feature extraction such that one CNN is for the left portion of the image and second is for the right portion of the image. We performed the classification using several state-of-the-art techniques with best results obtained with RF classifier.

Scenario 2 (**3CNNs + SVM**). In the second scenario, we have used an ensemble of 3 CNN for feature extraction such that one CNN is for the left portion image, second for the full image, and third for the right portion of the image with SVM model for classification.

Scenario 3 (**4CNNs + SVM**). In the third scenario, an ensemble of 4 CNNs have been used for the top left part of the image, second for the top right part of the image, third for the bottom left part of the image, and fourth for the bottom right part of the image with SVM model for classification.

Scenario 4 (**5CNNs + SVM**). In the fourth scenario, ensemble of 5 CNNs have been used for feature extraction such that first CNN is for the top left part of the image, second for the bottom left part of the image, third for the full image, fourth for the top right part of the image, and fifth for the bottom right part of the image with highest classification accuracy achieved with SVM classifier. It was observed that the highest classification accuracy (82.23%) is observed for the fourth scenario (PA) with a value of kappa coefficient 0.81. The highest percentage (84.32%) of sensitivity is observed for the proposed approach, while the lowest (57.32%) percentage is observed for the KNN classifier (Table 2). We generated a huge library of a combination of parameters for the state-of-the-art algorithms. A total of 156 trials are attempted for the state-of-the-art algorithm, and the results with highest classification accuracy (SVM: 67.11%, KNN: 54.12%, RF: 69.11%, CNN + SVM: 74.21%) are compared with our proposed scenarios. We further attempted 119 trials with different combinations of input and architectures for our proposed scenarios. We found that RF classifier gave the best performance when hybrid with two-layered CNNs (71.23%), SVM with three-layered CNNs (75.21%), and SVM with four-layered CNNs (79.77%) (Table 2).

Table 2 Accuracy assessment

Classification	Accuracy (%)	Specificity (%)	Sensitivity (%)	Kappa coefficient
SVM [5–7]	67.11	66.54	70.12	0.66
KNN [8]	54.12	52.11	57.32	0.52
RF [9, 10]	69.11	66.13	73.12	0.71
CNN + SVM [14]	74.21	71.22	77.23	0.72
S1 (2CNNs + RF)	72.23	68.43	75.32	0.71
S2 (3CNNs + SVM)	75.21	72.12	78.98	0.73
S3 (4CNNs + SVM)	79.77	77.23	81.11	0.78
PA (5CNNs + SVM)	**82.23**	**76.32**	**84.32**	**0.81**

5 Conclusion

In this paper, we have proposed a novel DL framework approach with a hybrid of CNNs for deep feature extraction and SVM for classification of different natural calamities. Further, we described four scenarios and selected the best-suited one, thereby obtaining accuracy 82.23%. Our proposed framework was observed to outperform the state-of-the-art algorithms. Our proposed model can be employed further by satellite aerial real-time image processing systems for locating the geographical locations of regions victimized by these natural disasters. This framework also provided a generalized architecture for efficient classification performance of similar classification problems.

References

1. Agarwal, S., Roth, D.: Learning a sparse representation for object detection. In: European Conference on Computer Vision, pp. 113–127. Springer, Berlin, Heidelberg (2002)
2. McCoppin, R., Rizki, M.: Deep learning for image classification. In: SPIE Defense + Security. International Society for Optics and Photonics, pp. 90790T–90790T (2014)
3. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)
4. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 248–255. IEEE (2009)
5. Cortes, Corinna, Vapnik, Vladimir: Support vector machine. Mach. Learn. **20**(3), 273–297 (1995)
6. Huang, F.J., LeCunm, Y.: Large-scale learning with svm and convolutional for generic object categorization. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 284–291. IEEE (2006)
7. Ishii, T., Nakamura, R., Nakada, H., Mochizuki, Y., Ishikawa, H.: Surface object recognition with CNN and SVM in Landsat 8 images. In: 14th IAPR International Conference on Machine Vision Applications (MVA), pp. 341–344. IEEE (2015)

8. Lee, Y.: Handwritten digit recognition using k nearest-neighbor, radial-basis function, and back propagation neural networks. Neural Comput. **3**(3), 440–449 (1991)
9. Ho, T.K.: Random decision forests. In: Proceedings of the Third International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. IEEE (1995)
10. Ho, T.K.: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. **20**(8), 832–844 (1998)
11. Ouyang, W., Wang, X.: Joint deep learning for pedestrian detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2056–2063 (2013)
12. Deng, J, Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: a large-scale hierarchical image database. In: CVPR. IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
14. Niu, X.-X., Suen, C.Y.: A novel hybrid CNN–SVM classifier for recognizing handwritten digits. Pattern Recogn. **45**(4), 1318–1325 (2012)

Comprehensive Analysis of Nonlocal Means Algorithm on Wide Set of ECG Signals

Supraja Kulkarni and Layak Ali

Abstract One of the important analyses of electrical activity of the heart is through electrocardiogram (ECG). The major and important step of this analysis is the extraction of pure ECG signals from noise which is an ongoing challenge. During recent years, many algorithms have been used for denoising the ECG signals. One of the well-known algorithms is nonlocal means (NLMs). The NLM is based on self-similarity of the patches. It is also based on the periodicity of the signal. Since ECG signal repeats its characteristics for regular intervals of time, NLM can be used for denoising ECG signals. This paper presents the comprehensive analysis of NLM algorithm on a wide set of well-known ECG data. The various ECG signals from ECG 101 to ECG 109 have been examined at different dB of the noise level. The analysis of NLM algorithm is carried out using various parameters such as signal-to-noise ratio (SNR), mean square error (MSE), and percent distortion (PRD). The paper further presents the numerical and graphical results to show the effectiveness of NLM algorithm.

Keywords ECG · NLM · Denoising · SNR · MSE · PRD

1 Introduction

The electrocardiography (ECG) is the noninvasive method of detecting various arrhythmias. It is also an observation of the electrical movements of the heart. ECG contains important pointers for detecting various kinds of diseases affecting the heart. It is essential for screening and diagnosing in order to detect various diseases and also for identifying and analyzing fatal arrhythmias correctly as early as possible. It is one of the vital tools used by the health examiner to observe the

S. Kulkarni · L. Ali (✉)
School of Engineering, Central University of Karnataka, Gulbarga 585367, India
e-mail: informlayak@gmail.com

S. Kulkarni
e-mail: Suprajabk31@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_56

physiological state of the heart. Due to its importance, an ECG signal should be extracted in a clean form, i.e., without any kind of noise or disturbances. Exact abstraction of health factors from the noisy biomedical signals is a serious and demanding task. Signals can be contaminated due to various reasons such as high-frequency noise, muscle artifacts, power line interference, and data collecting device noise, signal processing artifacts and baseline wander overlap signals of medical interest in both time and frequency [1]. In order to remove these sources of noise from the signal, standard filters are used. Low-pass filters are used to remove high-frequency noise from the source signals. Many algorithms are being implemented to get the pure ECG signal [2–8].

Denoising is also essential in image processing. The major factor here is to preserve sharp edges which are much serious in understanding the image. One of the recent approaches for addressing edge ruining in image denoising is through nonlocal means filter, which was first introduced by Buades et al. [9]. The NLM filter denoises image by considering patches from various parts of the image that possess related regional structure, considering the point that natural images regularly repeat its characteristics. Even ECG signal has regular morphological characters with slightly different variations. Hence, NLM algorithm can be used for denoising ECG signals. This paper presents the comprehensive analysis of NLM algorithm on the wide set of ECG data by introducing noise at different depts. The denoising capability of NLM algorithm is evaluated based on important parameters such as signal-to-noise ratio (SNR), mean square error (MSE), and percent distortion (PRD). The results obtained are presented in mainly two forms—numerical and graphical. The presented results support the effectiveness of NLM algorithm for denoising on wide ECG datasets. The paper is organized as follows: Sect. 2 presents the detailed NLM algorithm. Section 3 presents ECG dataset. Section 4 discusses the results, and paper is concluded in Sect. 5.

2 Nonlocal Means Algorithm

The nonlocal means (NLMs) algorithm can also be called as a statistical algorithm which was started by Buades et al. [9]. Their work [9] has been a reference to many citations and extension to the original algorithm [9] and is extensively used for denoising ECG signals [10]. The NLM algorithm is based on self-similarity concept and periodicity. Since the ECG possesses the said properties, NLM algorithm may be the best choice for denoising. The main aim of this technique is to filter by fixing the problems associated with local smoothing through calculating the smoothed value as a weighted average of other values in time series, which is based upon the similarity among the neighborhoods around the time series value [10]. The NLM considers the points which have higher similarity rather than the distance between them.

Nonlocal means of denoising gives solution for recovering of exact signal "u" specified in a field of noisy signal "v = u + n," where "n" is additional noise added to the signal "u". By considering an assumed sample "s", the estimated value $\hat{u}(s)$ is computed as a weighted average of all the points in the signal.

$$\hat{u}(s) = \frac{1}{Z(s)} \sum_{t \in N(s)} w(s,\ t) v(t) \tag{1}$$

where $Z(s) = \sum_t w(s,\ t)$

$$\text{And}\, w(s,\ t) = \exp\left(- \frac{\sum_{\delta \in \Delta} (v(s+\delta) - (v(t+\delta))^2}{2 L_\Delta \lambda^2} \right) \tag{2}$$

$$\equiv \exp\left(- \frac{d^2(s,t)}{2 L_\Delta \lambda^2} \right)$$

In the above equations, "λ" represents bandwidth parameter, and "Δ" represents local patches surrounding "s" and "t", totally containing L_Δ samples. Here, d^2 represents summed, squared point-by-point difference between samples in the patches centered on "s" and "t".

3 ECG Dataset

The dataset used for comprehensive evaluation of NLM algorithm is taken from [11]. The MIT-BIH Arrhythmia Database comprises 48 half-hour extracts of two-channel ambulatory ECG records, acquired from 47 subjects studied by the BIH Arrhythmia Laboratory in the vicinity of 1975 and 1979. Twenty-three records were picked arbitrarily from an arrangement of 4000 24-hour ambulatory ECG records gathered from a varied populace of inpatients (around 60%) and outpatients (around 40%) at Boston's Beth Israel Hospital; the residual 25 records have been picked from the similar set to incorporate less normal however clinically critical arrhythmias that would not be very much spoken to in a little arbitrary example. These records have been digitized at 360 samples for each second per channel through 11-bit resolution above a 10 mV range. At least two cardiologists individually interpreted all the records; differences were fixed to obtain the computer-readable reference explanations for every beat (about 110,000 observations in all) comprised in the databases [11]. To evaluate the NLM algorithm, authors have considered ECG101–ECG109 [11].

4 Experimental Result and Discussion

This section presents the detailed algorithmic parameters used for comprehensive comparison and test bench. Further, this section presents the experimental results using NLM algorithm.

4.1 Selection of the Parameter

The parameters used in the algorithm are mainly the patch size which is indicated by "half-width" "P" (hence $L_{\Delta} = 2P + 1$), the length of "N(s)", defined as a half-width "M", bandwidth". The bandwidth "λ" is an important parameter that controls the measure of smoothing connected. A very lower "λ" may usually arise noise variations to get great effect in de-weighting distinctive patches, bringing lack of averaging; an excessively expansive "λ" will cause disparate patches to appear similar, which results in blurring. The patch half-width "P" chooses the points where patches are to be compared, and should generally be related to the size of signal. Considering ECG signals, a sensible result to "P" is nothing but the "half-width" of the "higher-amplitude" "R" ECG signal. Normally, expanding the area "half-width" "M" (resulting in a "less local" look) would prompt improved performance [1].

4.2 Results

The comprehensive analysis of NLM algorithm was carried out on a computer with 4 GB RAM and Windows 10, in MATLAB R20017a platform. The NLM algorithm is evaluated on wide range of ECG dataset. Information from the physionet MIT-BIH Arrhythmia Database (www.physionet.org), for the heart signals ranging from ECG100 to ECG109 has been used with the introduction of signals with the various level of noise varying from 2 to 24 dB level to the original signal. The quantification of NLM algorithm is in terms of mean square error (MSE), percent distortion (PRD), and signal-to-noise ratio (SNR) with the help of following equations.

$$SNR = 10 \lim_{10} \frac{\sum_{n=1}^{N} (v[n] - u[n])^2}{\sum_{n=1}^{N} (\hat{u}[n] - u[n])^2}$$

$$MSE = \frac{1}{N} \sum_{n=1}^{N} (\hat{u}[n] - u[n])^2$$

Table 1 MSE

dB	ECG								
	101	102	103	104	105	106	107	108	109
2	400	350	650	850	710	1000	6000	200	1000
4	264	224	421	469	450	763	3940	123	890
6	150	127	262	265	303	515	2504	78	517
8	101	76	181	163	176	333	1499	53	327
10	062	49	118	93	103	231	976	39	198
12	39	29	78	54	62	151	599	28	129
14	26	19	49	36	41	101	379	19	78
16	17	13	32	22	23	63	241	12	50
18	12	9.03	21	15	16	42	156	81	33
20	08	6.2	13	10	11	28	103	4.8	19
22	5.2	3.8	9.6	7.6	8.3	19	62	2.9	132
24	3.2	2.4	6.4	5.1	58	12	39	1.4	9.4

$$PRD = 100\sqrt{\frac{\sum_{n=1}^{N}(\hat{u}[n]-u[n])^2}{\sum_{n=1}^{N}u^2[n]}}$$

The obtained results are presented in two ways: (1) numerical and (2) graphical.

1. *Numerical results*: Numerical results such as values of MSE, PRD, and SNR over different set of ECG signals with the addition of noise at different levels ranging from 2 dB to 24 dB are represented in Tables 1, 2, and 3 respectively.

Table 2 PRD

dB	ECG								
	101	102	103	104	105	106	107	108	109
2	2.09	1.94	2.62	2.97	2.79	3.51	7.99	1.46	3.75
4	1.70	1.54	2.10	2.20	2.17	2.78	6.39	1.14	3.08
6	1.28	1.14	1.66	1.65	1.78	2.28	5.10	0.91	2.45
8	1.05	0.89	1.38	1.29	1.35	1.84	3.94	0.75	1.95
10	0.83	0.69	1.11	0.98	1.04	1.53	3.18	0.65	1.51
12	0.65	0.57	0.90	0.74	0.81	1.24	2.49	0.54	1.22
14	0.53	0.45	0.71	0.61	0.65	1.01	1.98	0.46	0.95
16	0.44	0.37	0.59	0.47	0.49	0.80	1.58	0.36	0.76
18	0.36	0.31	0.47	0.39	0.42	0.65	1.27	0.29	0.62
20	0.29	0.25	0.38	0.33	0.34	0.54	1.03	0.25	0.48
22	0.24	0.20	0.31	0.28	0.29	0.44	0.83	0.17	0.39
24	0.18	0.16	0.26	0.23	0.25	0.35	0.63	0.12	0.33

Table 3 SNR

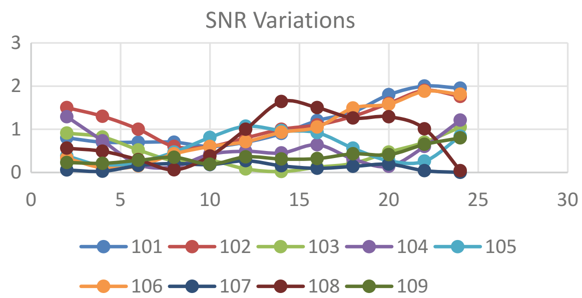
dB	ECG								
	101	102	103	104	105	106	107	108	109
2	0.80	1.50	0.91	1.29	0.38	0.33	0.056	0.56	0.23
4	0.70	1.30	0.82	0.73	0.16	0.107	0.026	0.49	0.21
6	0.70	1.00	0.52	0.18	0.24	0.15	0.159	0.293	0.28
8	0.70	0.60	0.27	0.13	0.47	0.43	0.206	0.06	0.35
10	0.60	0.60	0.28	0.42	0.81	0.60	0.180	0.38	0.19
12	0.70	0.80	0.08	0.49	1.07	0.71	0.27	1.00	0.36
14	0.90	1.00	0.02	0.45	0.97	0.93	0.152	1.64	0.31
16	1.20	1.10	0.12	0.64	0.92	1.05	0.095	1.50	0.32
18	1.40	1.30	0.21	0.31	0.56	1.49	0.139	1.26	0.43
20	1.80	1.60	0.47	0.14	0.27	1.58	0.19	1.29	0.41
22	2.00	1.90	0.70	0.60	0.26	1.88	0.04	1.01	0.65
24	1.95	1.76	1.05	1.21	0.84	1.81	0.00	0.04	0.80

2. *Graphical results*: Since the graphical results available as an outcome of experiment were more and were having almost similar trend, graphs are presented only for ECG101 with noise level of 2, 12 and 24 dB.

4.3 Discussion

This paper has shown an application of NLM algorithm to the various sets of ECG signals with addition of white Gaussian noise of different dB level. From Figs. 4, 5, and 6, it can be noticed that NLM denoises the true signal and signal with the addition of 2 dB noise signal and hence successfully meets the objective values. From Figs. 1, 2, 3, variations of SNR, PRD, and MSE, respectively, for different

Fig. 1 SNR variations of different ECG signals with different dB of noise levels



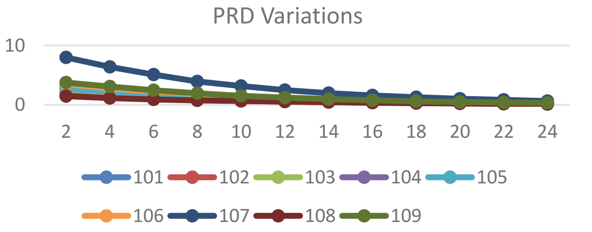


Fig. 2 PRD variations of different ECG signals with different dB of noise levels

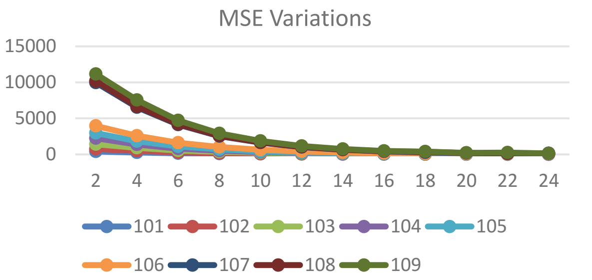


Fig. 3 MSE variations of different ECG signals with different dB of noise levels

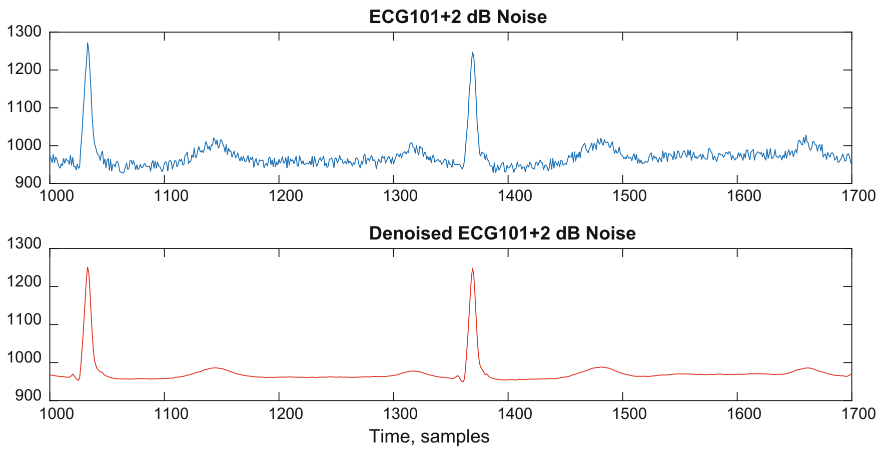


Fig. 4 Noise and denoised signal for 2 dB noise

ECG signals with different dB of noise level can be observed. NLM preserves the edges competently and removes the noise excellently.

The essential NLM strategy of considering for the signal patches comparative characteristics is probably going to be lower effective in suppressing interference

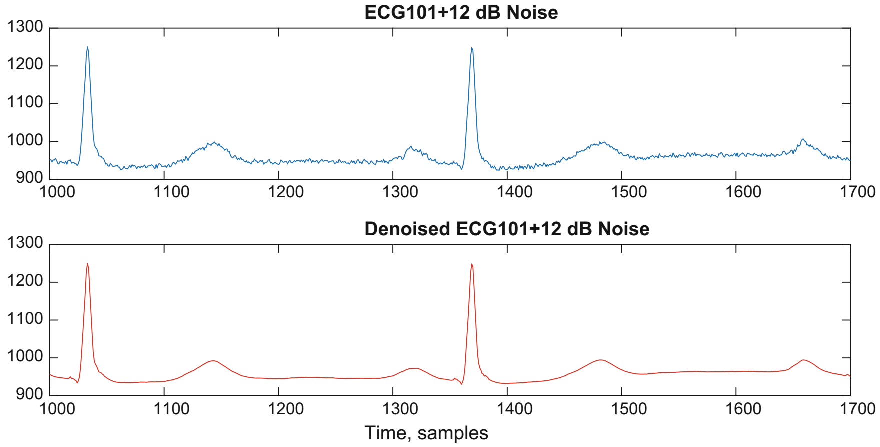


Fig. 5 Noise and denoised signal for 12 dB noise

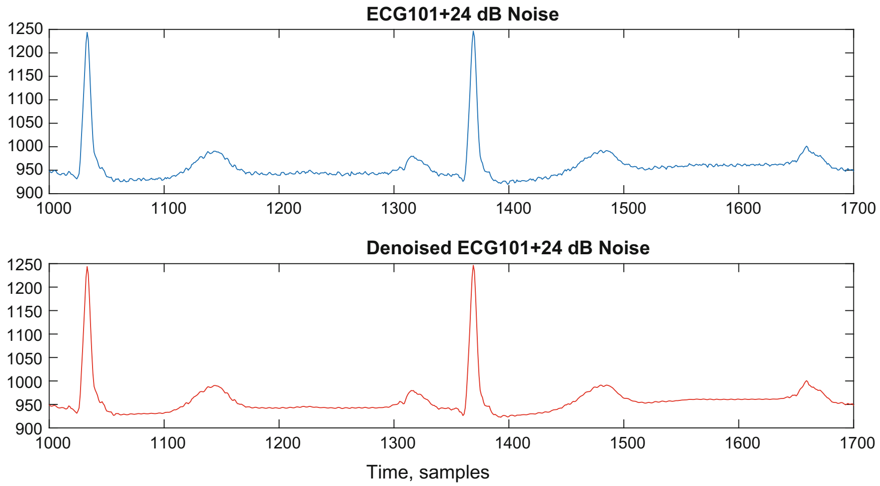


Fig. 6 Noise and denoised signal for 24 dB noise

artifacts. NLM has slight impact on the structure of the ECG signal with higher-amplitude portions. This may be clarified naturally, as uniform minor changes in the state of higher-magnitude wave patches surrounding "s" and "t" can prompt a great variance "d^2(s, t)" and along these lines to small weights. Signals with lesser amplitude may have similar patches, higher-amplitude signals do not match with other region, and hence, they are lightly averaged. Though, a drawback is that signal regions that are closer to a high-amplitude region, but which are not some portion of it, might be under found the middle value so that signal departures may not be as perfect as wanted [10]. Numerous clarifications are existing to the above issue; for instance, locally versatile bandwidth parameter determination [12], shape versatile patches [13], or their mix [14] then can give favorable opportunities for further task.

5 Conclusion

This paper demonstrates comprehensive analysis of nonlocal means (NLMs)-based filtering algorithm on various sets of ECG signals, i.e., from ECG 101 to ECG 109 with the addition of noise with different dB levels. It is possible to remove noise with the help of this algorithm in much satisfactory level without altering the original signal characteristics. This algorithm is one of the effective ways for denoising the signals in general and ECG signals in particular. The results are obtained by calculating means square error, percent distortion, and signal-to-noise ratio. It can be concluded that MSE decreases with the increase in the noise level in signals. Percent distortion also decreases with the increase dB level in the signal. SNR follows the similar trend. Thus, the numerical and graphical results conclude the effectiveness of NLM algorithm for denoising ECG signal at various degree of noise level. Authors are working toward deploying NLM algorithm with ECG system based on Arduino.

References

1. He, T., Clifford, G., Tarassenko, L.: Application of independent component analysis in removing artefacts from the electrocardiogram. Neural Comput. Appl. **15**(2), 105–116 (2006)
2. Barros, A.K., Mansour, A., Ohnishi, N.: Removing artifacts from electrocardio-graphic signals using independent components analysis. Neurocomputing **22**, 173–186 (1998)
3. Alfaouri, M., Daqrouq, K.: ECG signal denoising by wavelet transform thresholding. Am. J. Appl. Sci. **5**(3), 276–281 (2008)
4. Poornachandra, S.: Wavelet-based denoising using subband dependent threshold for ECG signals. J. Digit. Signal Process. **18**, 49–55 (2008)
5. Li, N., Li, P.: An improved algorithm based on EMD-wavelet for ECG signal de-noising. In: Proceedings of International Joint Conference on Computational Sciences and Optimization, pp. 825–827 (2009)
6. Suchethaa, M., Kumaravel, N.: Empirical mode decomposition based filtering techniques for power line interference reduction in electrocardiogram using various adaptive structures and subtraction methods. Biomed. Signal Process. Control **8**, 575–585 (2013)
7. Kopsinis, Y., McLaughlin, S.: Development of EMD-based denoising methods inspired by wavelet thresholding. IEEE Trans. Signal Process. **57**(4), 351–1362 (2009)
8. Yan, Y., Zhanzhong, C.: Noise and zero excursion elimination of electrostatic detection signals based on EMD and wavelet transform. In: IEEE International Congress on Image and Signal Processing, pp. 1–5 (2009)
9. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)

10. Tracey, B.H., Miller, E.L.: Nonlocal means denoising of ECG signals. IEEE Trans. Biomed. Eng. **59**(9) (2012)
11. Moody, G.B., Mark, R.G.: The impact of the MIT-BIH Arrhythmia Database. IEEE Eng. Med. Biol. **20**(3), 45–50 (2001) (PMID: 11446209)
12. Ville, D.V.D., Kocher, M.: SURE-based nonlocal means. IEEE Signal Process. Lett. **16**(11), 973–976 (2009)
13. Deledalle, C.A., Duval, V., Salmon, J.: Nonlocal methods with shape adaptive patches (NLM-SAP). J. Math. Image Vis. **43**(2), 103–120 (2012)
14. Ville, D.V.D., Kocher, M.: Nonlocal means with dimensionality reduction and SURE-based parameter selection. IEEE Trans. Image Proc. **20**(1), 1732–1735 (2011)

Periocular Region-Based Biometric Identification Using Local Binary Pattern and Its Variants

K. Kishore Kumar and Movva Pavani

Abstract Biometric systems are gaining importance significantly in the present day automated systems especially in the areas of authentication, access control, security, and forensic applications for the identification of criminals. Existing biometric systems using the face and iris regions had reached the state of maturity with almost having high performances of 100% accuracy provided the images of the subjects are acquired in the cooperative scenarios. The periocular region is one of the most promising biometric traits providing better robustness and high discrimination ability. Periocular region-based biometric recognition systems are well suited for the wild environments where the subjects are not cooperative. In the proposed paper, the pixel-based LBP and patch-based LBP variants are used as local descriptors for the feature extraction of discriminative features from the full face and periocular regions. Euclidean distance is used to find the matching score between two extracted feature vectors. The experimentation is performed on FRGC, FERET, and Georgia Tech face databases to compare the performance of both periocular and face biometric modalities. It showed that the periocular region has almost the same level of performance of the face region using only 25% data of the complete face.

Keywords Periocular region · Pixel-based approach · Patch-based approach
Euclidean distance · Rank-1 recognition rates · Unconstraint biometrics

K. Kishore Kumar (✉) · M. Pavani
Faculty of Science & Technology, Department of ECE, IFHE, Hyderabad, India
e-mail: kkishore@ifheindia.org

M. Pavani
e-mail: pavanimovva@ifheindia.org

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_57

1 Introduction

Biometric systems have become an integral part of our modern life and the present traditional biometric systems [1] using the face and iris had reached the state of maturity. The constraint biometric systems demand the cooperative subjects for registering the data and they work very well under controlled and ideal circumstances like high-resolution images taken from the short distance, well-illuminated conditions. Under the circumstances, when the images are captured from longer distances when the subject is on the move or from the noncooperative subject the performance of these systems will degrade when they operate in non-ideal conditions. Research is now focused on improving the biometric systems suited for the wild environments under the non-ideal conditions. In the design of the framework for the unconstraint scenarios, the periocular region has become the best alternative with better robustness and high discrimination ability. The periocular region is the sub-portion of the face in the vicinity of an eye with eyebrows, eyelids, and eye folds.

In the unconstrained scenarios, periocular region [2] is gaining significance as the biometric trait as it is the most discriminating portion of the face, due to its flexibility and easy to acquire the images in the adverse conditions. Periocular region [3]-based biometric systems are capable of providing the better recognition rates even with the subjects imaged at a distance, on the move, under irregular lighting condition with cameras working in the visible wavelength. During the biometric acquisition process, the periocular region is imaged along with the complete face and iris region which doesn't require any extra storage memory space which is a cost effective [4, 5]. Better recognition accuracies can be obtained when the periocular region is used in fusion with face/Iris modalities. Major advantage with the periocular region is: it is less affected by the aging process compared to face, and periocular region is used in the design of age-invariant face recognition systems [6, 7]. Periocular biometrics use small periocular region [3] templates which use only 25% of the full face image when compared with the large face models makes them faster in the recognition process.

2 Related Work

Park et al. [2] explored the discriminating capability of the periocular region and established periocular region of the face as a useful biometric when iris recognition fails and proposed the periocular-based recognition system using the local appearance-based approaches local binary patterns and gradient orientation histograms as feature extraction techniques for the periocular region data and Euclidean distance is used as a classifier. Experimental results showed that the periocular-based recognition under occlusion provided the better performance than the face recognition. Sato et al. [8] presented the first partial face recognition using the sub-regions of the face like nose, eyes, and ears. Periocular region was found in the literature for the first with the context of partial face recognition. Experimentation was performed on 720 images from 120 subjects, and it showed that the eye region achieved better recognition rates (92%) compared to others regions demonstrating that the periocular region provides better discriminative features. Savvides et al. [9] performed partial face recognition using the sub-regions of the face like nose, eyes, and mouth on the facial recognition grand challenge (FRGC) data set, and the results showed the eye region achieved better recognition rates compared to others. Teo et al. [10] compared the performance of full face recognition with the eye-based partial face recognition. In the proposed paper, the nonnegative matrix factorization (NMF)-based techniques are used to extract features from the data and the experimentation is performed on Essex University Face Database of 3060 images from 153 subjects. Miller et al. [11] explored the effects of data quality in a periocular biometric system. Santos and Hoyle [12] showed the fusion of iris, and periocular modalities had improved the performance of the recognition systems. Lyle et al. and Merkow et al. presented the use of the periocular region for the soft biometric recognition systems, and periocular region had been used for the gender classification [3].

3 Proposed Methodology

Face/periocular biometric system [13] requires the following sequences of steps as shown in the figure are pre-processing (removal of noise, geometric normalization, illumination correction, and histogram equalization), testing/training sets, feature extraction, and matching techniques (Fig. 1).

3.1 Periocular Region Extraction

Periocular region is extracted from normalized and equalized facial images, which is accomplished by placing a square bounding box around each eye, centered on the post-geometric normalization eye center locations.

3.2 Feature Extraction

Each feature extraction technique transforms a two-dimensional image into a one-dimensional feature vector through its unique process. Feature extraction techniques employed in the face and periocular region are uniform LBP [14], 3P-LBP, and its variant hierarchical three-patch local binary pattern (H-3P-LBP) [15]. The uniform LBP is a pixel-based while 3P-LBP and H-3P-LBP are

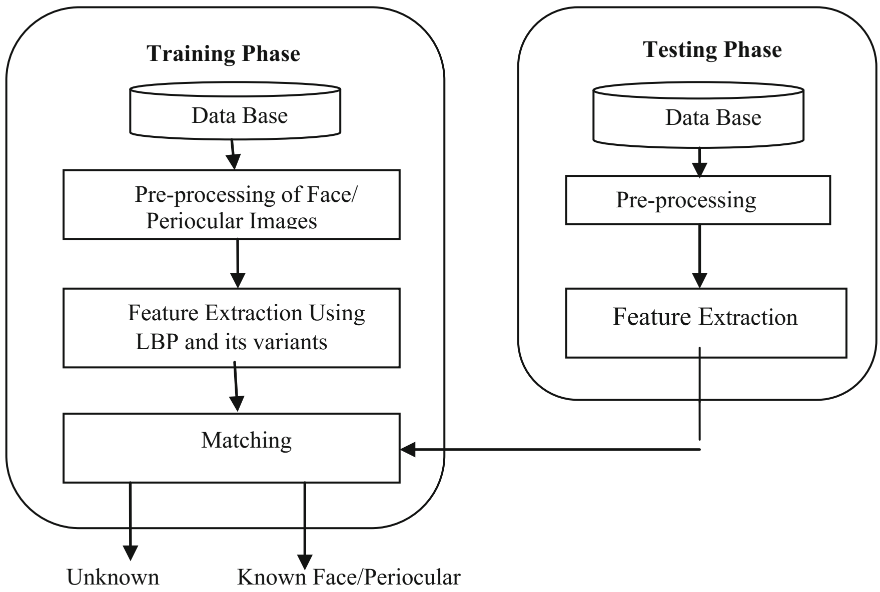


Fig. 1 Framework of the proposed approach

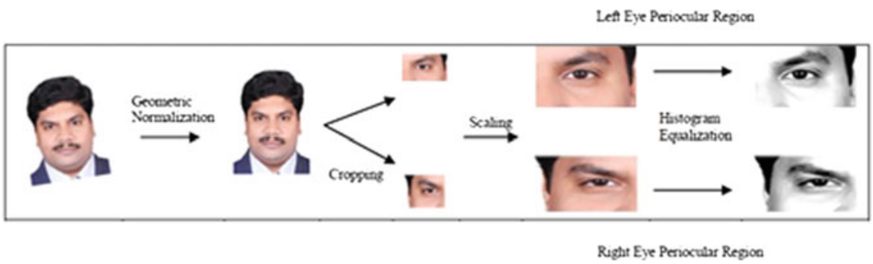


Fig. 2 Process flow for the periocular region extraction

patch-based methods. Patch-based approaches encode the similarities across pixels of the neighboring patches. Unlike pixel-based techniques, the patch-based approaches consider colored regions, complex textures, edges, and lines in a unified way (Fig. 2).

3.2.1 Local Binary Pattern (LBP)

Local appearance-based approaches [16] are the feature extraction techniques which collect statistics within local neighborhoods around each pixel of an image providing one-dimensional feature vectors. A local binary pattern (LBP) is a texture classification method that was developed by Ojala et al. [17]. LBP is used extensively for both facial recognition [15] and periocular recognition [18]. The LBP value of the pixel of concern P_k is a function of intensity changes in the neighborhood of M sampling points on a circle of radius r, then the LBP operator is given by,

$$LBP_{M,r} = \sum_{n=0}^{M-1} s(g_n - g_c)2^n \tag{1}$$

where $\mathrm{S(p)} = 1$ if $\mathrm{p} > 0$
$= 0$ if $\mathrm{p} < 0$ g_c intensity of the pixel of concern at the center of the pixels on the circumference of a circle with values of g_n, where n = 0, …, M − 1. In the proposed work, all LBP calculations are made from a circle of radius one pixel with eight pixels along the circumference of the circle.

3.2.2 3P-Local Binary Pattern (3P-LBP)

In 3P-LBP, the center patch value is compared with the pair of patches which are β patches apart along the circle and the value is set based upon the similarities between of the two patches with the center patch.

The 3P-LBP is given by

$$3\mathrm{P} - \mathrm{LBP}_{r,n,w,\beta}(p) = \sum_{i}^{m} (f(d(C_i,\ C_p) - d(C_{i+\beta\, mod\, m},\ C_p))2^i \tag{2}$$

where C_p is the central patch, C_i and $C_{i+\beta}$ are two patches along the ring and d is the distance function between patches and f is defined as

$$f(x) = \begin{cases} 1, & x \geq \tau \\ 0, & x < \tau \end{cases}$$

where τ is set to a value slightly greater than zero to provide stability in uniform regions.

3.2.3 Hierarchical 3P-Local Binary Pattern (3P-LBP)

Patch-based approach H-3P-LBP is an extension of 3P-LBP operator [15] computed over various scales (multi-resolution) of an image. H-3P-LBP descriptor $H(I)$,

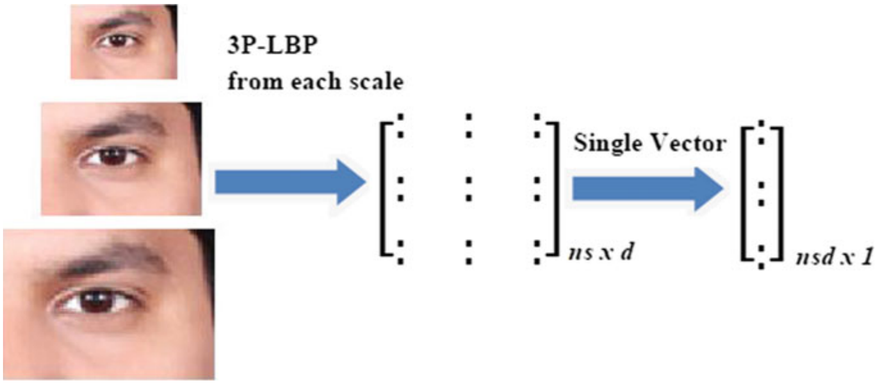


Fig. 3 Hierarchical 3P-LBP code for an image

final feature matrix is obtained by combining the 3P-LBP descriptors from each level (Fig. 3).

3.3 Classification Using Euclidean Distance Metric

Euclidean distance is employed to measure the matching score between the pair of feature vectors. It is the straight-line distance between two pixels and is evaluated using the Euclidean norm.

$$ED_{\ i,h} = \sqrt{\sum_{i=1}^{p} \left(a_{i,j} - a_{h,j}\right)^2} \qquad ((3))$$

4 Experiment

For the experimentation FRGC, FERET and Georgia Tech face database are used to compare the performance of both periocular and face biometric modalities. These databases are publicly available and have various subject face images with variations in pose, occlusions, expression, illumination, etc., taken under unconstrained conditions.

Face Recognition Grand Challenge (*FRGC Data Set*)

Face recognition grand challenge (FRGC) database consists of large face images, and it is significantly used for extracting the periocular regions from the face images. It consists of 16,029 still frontal face images, high-quality resolution faces of size 1200 × 1400 with different sessions, and variable expressions.

Face Recognition Technology (*FERET Data Set*)

The FERET database consists of gray scale and color images of faces. The database has subjects with the different poses and with different facial expressions. Dataset consists of 1,980 frontal face images taken from 990 subjects.

Georgia Tech Database

Georgia Tech face database consists of 750 images of 15 images per subject with the image resolution of 640 × 480 pixels.

5 Results and Discussion

In the proposed work, local descriptors uniform local binary patterns LBP, 3P-LBP, and its variant hierarchical three-patch local binary pattern (H-3P-LBP) are used for the feature extraction of discriminative features from the regions of the periocular [19] and full face. Rank-1 recognition rate is the performance measure used to define the robustness of the designed system, and it illustrates the successfulness of the scheme in identifying the best match for a subject. In most of the cases, face region performs better when compared with the periocular regions as it has more information in terms of nodal points. Table 1 shows the performance statistics [20, 21] for the LBP and its variants in terms of Rank-1 accuracy for the three databases. Periocular region showed significant accuracies compared to face with only using 25% of the full face information [22] (Figs. 4, 5, and 6).

Table 1 Results obtained from local binary pattern and its variants on FRGC, FERET, and Georgia Tech dataset, respectively

Method	Region	Rank-1 accuracy FRGC	Rank-1 accuracy FERET	Rank-1 accuracy Georgia Tech
LBP	Left eye	80.89	83.32	63.87
	Right eye	80.57	82.18	64.91
	Face	82.14	91.64	66.98
3P-LBP	Left eye	95.13	87.64	82.02
	Right eye	95.18	84.38	86.13
	Face	95.32	93.23	89.78
H-3P-LBP	Left eye	95.34	89.10	84.77
	Right eye	94.86	85.18	88.65
	Face	95.91	94.38	90.23

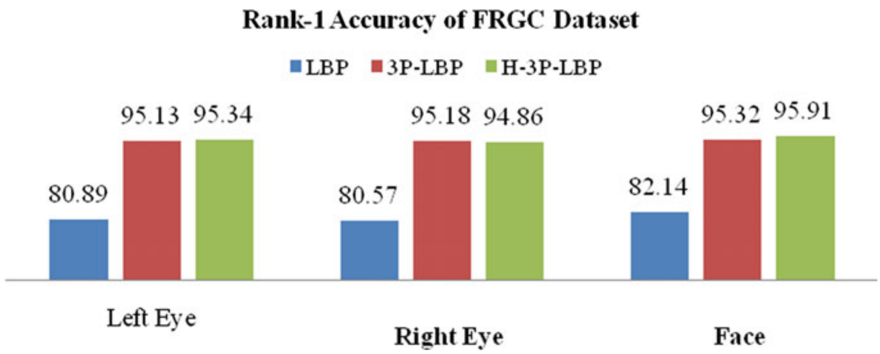


Fig. 4 Rank-1 recognition rates for the LBP, 3P-LBP, H-3P-LBP methods on FRGC dataset

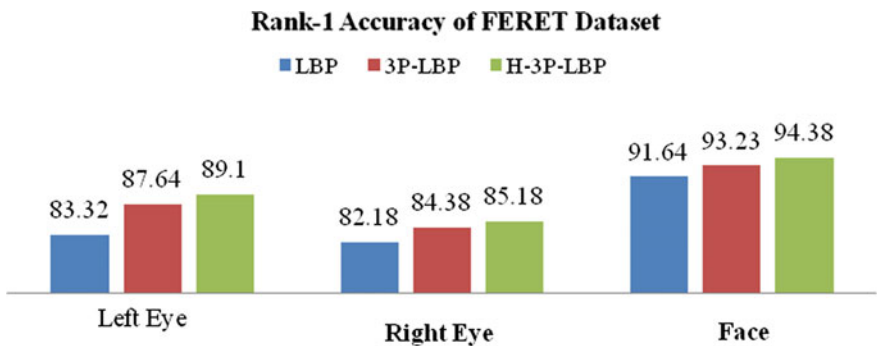


Fig. 5 Rank-1 recognition rates for the LBP, 3P-LBP, H-3P-LBP methods on FERET dataset

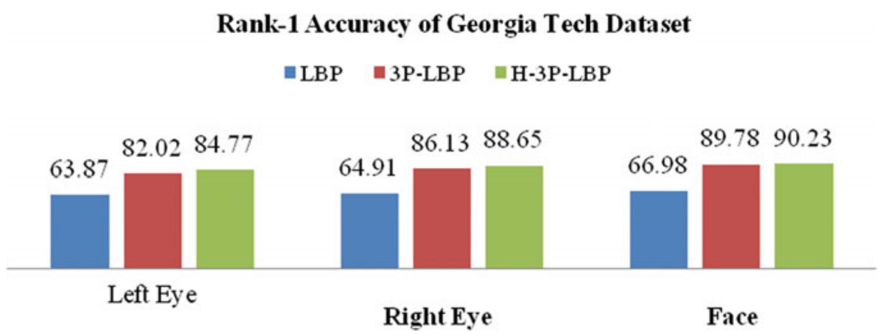


Fig. 6 Rank-1 recognition rates for the LBP, 3P-LBP, H-3P-LBP methods on Georgia Tech dataset

6 Conclusion

In the proposed paper, the pixel-based LBP and patch-based LBP variants are used as local descriptors for the feature extraction of discriminative features from the regions of the full face, periocular and Euclidean distance are used as a classifier to find the matching score between two feature vectors on three distinctive and challenging datasets. FRGC, FERET, and Georgia Tech databases are used for the experimentation to compare the performance of both periocular and face biometric modalities, and it showed that the periocular region has a similar level of performance of the face region using only 25% data of the complete face.

References

1. Jain, A., Ross, A., Prabhakar, S.: An introduction to biometric recognition. IEEE Trans. Circuits Syst. Video Technol. **14**(1), 4–20 (2004)
2. Park, U., Jillela, R., Ross, A., Jain, A.: Periocular biometrics in the visible spectrum. IEEE Trans. Inf. Forensics Secur. **6**(1), 96–106 (2011)
3. Merkow, J., Jou, B., Savvides, M.: An exploration of gender identification using only the periocular region. In: Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–5, Sept 2010
4. Juefei-Xu, F., Savvides, M.: Unconstrained periocular biometric acquisition and recognition using COTS PTZ camera for uncooperative and non-cooperative subjects. In: Proceedings of the IEEE Workshop on ACV, pp. 201–208, Jan 2012
5. Juefei-Xu, F., Luu, K., Savvides, M., Bui, T., Suen, C.: Investigating age invariant face recognition based on periocular biometrics. In: Proceedings of the International Joint Conference on Biometrics, pp. 1–7, Oct 2011
6. Ling, H., Soatto, S., Ramanathan, N., Jacobs, D.W.: Face verification across age progression using discriminative methods. IEEE Trans. Inf. Forensics Secur (2010)
7. Ramanathan, N.: Face verification across age progression. IEEE Trans. Image Process (2006)
8. Sato, K., Shah, S., Aggarwal, J.: Partial face recognition using radial basis function networks. In: Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, pp. 288–293, Apr 1998
9. Savvides, M., Abiantun, R., Heo, J., Park, S., Xie, C., Vijayakumar, B.: Partial holistic face recognition on FRGC-II data using support vector machine. In: Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition, p. 48, June 2006
10. Teo, C., Neo, H., Teoh, A.: A study on partial face recognition of eye region. In: Proceedings of the International Conference on MV, pp. 46–49, Dec 2007
11. Miller, P., Lyle, J., Pundlik, S., Woodard, D.: Performance evaluation of local appearance based periocular recognition. In: Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–6, Sept 2010
12. Santos, G., Hoyle, E.: A fusion approach to unconstrained iris recognition. Pattern Recogn. Lett. **33**(8), 984–990 (2012)
13. Hollingsworth, K., Bowyer, K., Flynn, P.: Identifying useful features for recognition in near-infrared periocular images. In: Proceedings of the IEEE International Conference on Biometrics: Theory Applications and Systems, pp. 1–8, Sept 2010
14. Kumar, K.K., Rao, P.T.: Face verification across ages using the discriminative methods and see 5 classifier. In: Proceedings of First International Conference on ICTIS, vol. 51, pp. 439–448. Springer SIST series

15. Mahalingam, G., Ricanek, K.: LBP-based periocular recognition on challenging face datasets. EURASIP JIVP **2013**(1), 1–13 (2013)
16. Joshi, A., Gangwar, A., Sharma, R., Saquib, Z.: Periocular feature extraction based on LBP and DLDA. In: Advances in Computer Science, Engineering & Applications, Volume 166 of Advances in Intelligent and Soft Computing, pp. 1023–1033. Springer (2012)
17. Ojala, T., Pietikainen, M., Maenpaa, T.: A generalised local binary pattern operator for multiresolution gray-scale and rotation invariant texture classification. In: Second International Conference on Advances in Pattern Recognition, pp. 397–406 (2001)
18. Mahalingam, G., Chandra, K.: Face verification with ageing using AdaBoost and local binary patterns. In: Proceedings of the Seventh Indian Conference on Computer Vision Graphics and Image Processing—ICVGIP 10 ICVGIP 10 (2010)
19. Kumar, K.K., Rao, P.T.: Biometric identification using the periocular region. In: Proceedings of 2nd International Conference on ICTIS, vol. 84, pp. 619–628. Springer, SIST series
20. Bharadwaj, S., Bhatt, H., Vatsa, M., Singh, R.: Periocular biometrics: when iris recognition fails. In: Proceedings of the IEEE International Conference on Biometrics: Theory, Applications, and Systems, pp. 1–6, Sept 2010
21. Kumar, K.K., Rao, P.T.: Periocular region based biometric identification using the local descriptors. In: Proceedings of 2nd International Conference on Intelligent Computing and Communication ICICC 2017, vol. 673, pp. 341–351. Springer, AISC series
22. Kumar, K.K., Pavani, M.: LBP based biometric identification using the periocular region. In: Proceedings of IEEE 8th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 204–209. Vancouver, BC (2017)

Hand Geometry- and Palmprint-Based Biometric System with Image Deblurring

Hiren J. Galiyawala and Rupal Chaudhari

Abstract Due to peg-free system for image acquisition, it is obvious to have a motion of user's hand. Motion activity creates the blurring effect in captured images, which leads to the false identification of a person. In such cases, image deblurring is required to restore the image. This paper proposes hand geometry- and palm texture-based person identification system with motion blur deblurring of palm images. The deblurring process may generate the ringing artifacts. The quality of the restoration of a motion blurred image is depending on the estimation of motion blur parameters, blur angle and blur length. In this paper, Radon transform and cepstral method have been used for the blur parameter estimation. Point spread function (PSF) can be constructed using blur parameters. For the deblurring of the palm images, L2/TV regularization based on augmented Lagrangian method has been used. It deblurs the image with the removal of ringing artifacts. Restored images are further used in the biometric system to analyze the robustness of the system against the motion blur. This system is tested on a database collected from 100 people. 99.5% genuine acceptance rate (GAR) is achieved for bimodal system against motion blur for discrete cosine transform (DCT)-based features. Equal error rate (EER) achieved is 1.1719.

Keywords Hand geometry · Palmprint · Multimodal biometric system
Blur parameter estimation · Image deblurring · PSF estimation

H. J. Galiyawala · R. Chaudhari (✉)
Department of Electronics & Communication Engineering, CGPIT,
Uka Tarsadia University, Surat, India
e-mail: rupalec19@gmail.com

H. J. Galiyawala
e-mail: hire16@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_58

1 Introduction

In recent years, many of the industries, academic institutions etc., are using biometric-based user authentication system. Researchers have constantly adapted the user authentication technology to improve the security of the system for the society and to secure the people from various antisocial elements. Among the various biometric traits, fingerprint, hand geometry, and palmprints are the hand-based biometric used for security. Hand geometry and palmprint can be acquired from the single image of hand [1]. It is obvious to have various physical movements like rotation, motion in the various directions of hand, which may create blurring effect in the image. Thus, it affects the security of the system. This paper is focused on motion blur artifact.

Image restoration is a challenging issue in image processing. Uniform image degradation model [2] is given by $b = i * h + \eta$, where b is the observed blurry image, i is original image, h is the blur kernel (PSF) that causes the image degradation, '*' is the convolution operator, and η is the additive noise. According to this model, deconvolution of PSF and blur image will give the deblurred image. Deblurring can be accomplished using either blind or non-blind deblurring technique. In non-blind deblurring, the original image and blur kernel is already known, but in blind deblurring, both are unknown. That is considered as an ill-posed problem [3, 4].

Radon transform [5–7], cepstral method, Haar wavelet transform (HWT), support vector machine (SVM), and Hough transform [3, 8–10] have been proposed for estimation of blur angle and blur length. In the PSF construction, blur angle is estimated first, and using the value of the angle, blur length is computed. Once the blur parameters are accurately estimated, PSF (blur kernel) can be easily constructed. Previously, various methods were used for the deconvolution such as direct inverse filtering, Wiener filter [2, 4, 7, 11], Richardson–Lucy (RL) method [12, 13], regularized least squares minimization [14], and TV regularization techniques for image and video restoration [8, 15, 16]. Feature extraction techniques such as wavelet and Gabor filter, fusion of palmprint features with finger geometry and directional energy were analyzed. k-Nearest neighbor (k-NN) classifier [17], Hamming distance, and Euclidian distance classifiers [18, 19] were proposed for feature matching.

This paper proposes person identification system using the bimodal biometric system with one novel approach of palmprint deblurring. Radon transform and cepstral method have been proposed to estimate the blur parameters. PSF is constructed using estimated blur angle and length. L2/TV minimization algorithm has been applied for the restoration of palm images. Restored images were further used for the biometric identification process. The paper is organized as follow. In Sect. 2, basic block diagram of palmprint-based person identification with motion deblurring of palm image is explained. Section 3 gives the brief information on quality assessment parameters. Results of blur parameters and deblurring algorithm are discussed in Sect. 4. In next section, identification results are given, and the final section gives the conclusion of presented work.

2 Person Identification System

Figure 1 shows the block diagram of hand geometry- and palmprint-based person identification system with the deblurring of a motion blurred palm images. Original palm images are filtered by the weighted averaging filter and converted into gray scale. Uniform motion blur filter is applied on the resultant grayscale image. Blur parameters are estimated, and PSF is constructed, which is injected into the deblurring algorithm. It gives the restored image with no ringing artifacts. Deblurred images are passed through the filtering process for the reduction of noise. Hand geometry and palmprint features are extracted to identify the person.

2.1 Blur Parameter Estimation

Hand motion of a person or camera shake produces the motion blurring effect in captured image. Motion blurring function which is often called PSF is the function of two blur parameters—1. blur angle (θ) and 2. blur length (L). Uniform blurring is caused by L and θ, which results in blurring of the image [3]. More accurate estimation of blur parameters gives the quality restoration of an image. For angle estimation, Radon transform technique was used, and for length estimation, cepstral method was used.

Blur Angle Estimation

Radon transform of an image $I(x, y)$ is given by equation, $R(I)(x, \theta) = \int_{-\infty}^{\infty} I(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta)dy$. Blur angle is estimated using Radon transform with some preprocessing steps [5, 8]. Motion blur filter is applied to the

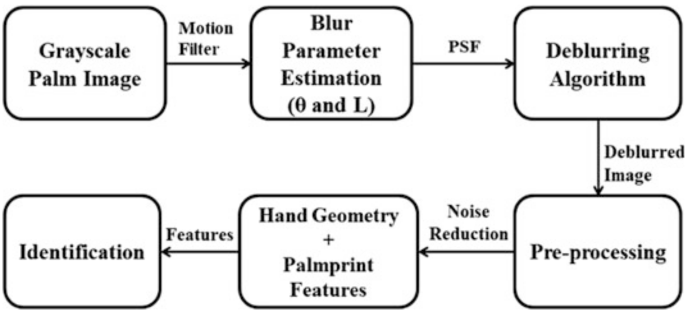


Fig. 1 Block diagram

original grayscale image. After that, Fourier transform of the image is computed. Apply log transform three times on the Fourier spectrum. Resultant log spectrum is passed through the averaging filter to remove the noise. This filtered spectrum is converted into binary using Otsu's method. These preprocessing steps give more clear boundary lines in a spectrum of the image, which helps in accurate estimation of blur angle. Radon transform of the binary spectrum is computed for the values of theta ranging from 0 to 180°. One maximum value from each column of the Radon matrix is taken which gives the array of size 1×180. Five maximum values are found from the array, and their locations in the array will give five possible motion blur angles of the observed blurry image.

These values are the possible blur angles. The highest peak of the Radon transform is the actual value of blur angle. In some cases, the actual value of the angle is observed at nearby maximum values of Radon transform. Thus, for the accurate angle estimation, five maximum values of the Radon transform were recorded.

Blur Length Estimation

The cepstrum transform of the image $f(x, y)$ is given by equation $C\{f(x, y)\} = F^{-1}(\log(|F(f(x, y))|))$, where '$F$' indicates the Fourier transform of an image. For the blur length calculation, cepstral method [5] has been used. Calculation of blur length is depending on the previously estimated blur angle because in the blur length estimation, cepstrum is rotated by the estimated blur angle in inverse direction. Fourier transform of the motion blur image is computed. After that, logarithmic transform is applied on the spectrum. To convert the image into the cepstral domain, inverse Fourier transform of log spectrum is computed. This cepstral is rotated by the blur angle in inverse direction, and mean of the array is computed; the first negative value from the origin will give the value of blur length. Negative values generated only when the zero crossing of the sinc function is occurred, otherwise blur length is estimated by the lowest peak [8].

2.2 Deblurring Algorithm

Motion blurred images are deblurred using L2/TV regularization technique based on augmented Lagrangian method. It is an algorithm for solving total variation constrained least squares problems. Chan et al. [15, 16] give a deconvtv algorithm for the video restoration. For the restoration of blur palm images, deconvtv is used which gives resultant images with no ringing artifacts and also preserved the texture of palm. Ringing artifacts are generating due to an error in blur parameter estimation. L2/TV least squares minimization with total variation regularization is given by: L2 = $minimize_f \ \frac{\mu}{2}||Hf - g||^2 + ||Df||_2$, where g is the observed blurry image/video, H is the convolution matrix, and D is three-dimensional forward finite difference operator. The vector f is optimization variable. μ is the regularization parameter. In this deblurring process, $\mu = 10^5$ is chosen when blur length is less than 100, and $\mu = 10^6$ is chosen for blur length greater than 100.

2.3 Preprocessing

Basic steps for preprocessing of deblurred images are shown in Fig. 2. Preprocessing majorly contains: 1. noise reduction by average filtering, 2. gray-to-binary conversion by thresholding method, 3. smoothing by the morphological opening operation, 4. hand boundary extraction also by the morphological operation, 5. fingertips and valley point identification by tracing the hand boundary, 6. palm ROI extraction using fingertips and valley points.

2.4 Feature Extraction

Various features are extracted from hand geometry and palmprints for the further identification process. Hand geometry features are based on Euclidian distance. Four finger lengths, 12 finger widths and one palm width, i.e., 17 geometrical features, are extracted from hand geometry. Palmprint features are based on standard deviation and various transforms like discrete Fourier transform (DFT), discrete wavelet transform (DWT), and discrete cosine transform (DCT). A total of 36 standard deviation values are extracted as palmprint features.

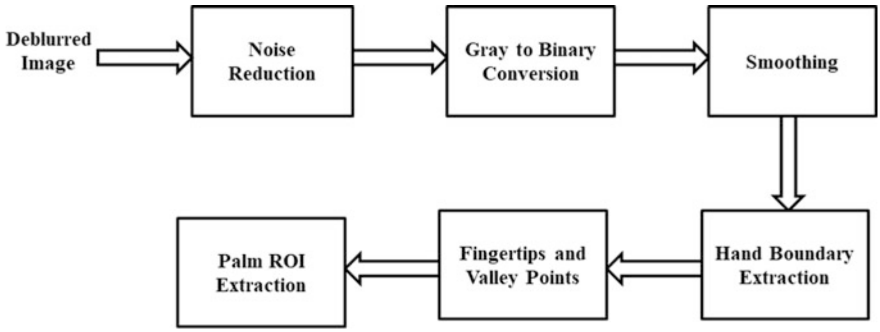


Fig. 2 Basic steps for preprocessing and ROI extraction

2.5 Identification

Person identification algorithm calculates the minimum distance between features stored in the database and extracted feature set. For identification, different training and testing sets are prepared, which are discussed in the results section. The details of preprocessing, feature extraction, and identification process can be found in [1].

3 Quality Assessment Parameters

The quality of restored images was analyzed by two quality assessment parameters: peak signal-to-noise ratio (PSNR) [8] and structural similarity index (SSIM) [20] of restored image. PSNR is the ratio between the maximum possible power of the signal to the noise which affects the image. PSNR is defined by Eq. (1).

$$PSNR = 10log_{10}\left(\frac{peak\,value^2}{MSE}\right) \tag{1}$$

where peak value is maximum possible pixels in the image, and MSE is mean square error. Value of PSNR is measured in decibels (dB).

SSIM is the structural similarity between the original and deblurred images. SSIM index is measured in between the range of 0–1. It is a measure of luminance, contrast, and structure of an image. More the SSIM, more similar the image to the original image. It is defined by Eq. (2).

$$SSIM(x, y) = \frac{\left(2\mu_x\mu_y + C_1\right) + \left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{2}$$

where μ_x and μ_y are the arithmetic mean, σ_x and σ_y are the values of variance, and σ_{xy} is the covariance of x and y. Here, x and y are the original and reference images, respectively. C1 and C2 are constants.

In this research work, deblurred image selection for identification is based on SSIM index. One image is selected from the five restored images, which has the highest SSIM. SSIM and PSNR both techniques required original image as a reference for computation. In blind image restoration techniques, original image is not available; in that case, for quality assessment of restored images ‘No-reference image quality assessments’ techniques are useful [21–23].

4 Results of Blur Parameters and Deblurring

In Fig. 3, blur angle estimation of motion blurred image (b) is explained. It is motion blurred by θ = 90° and L = 200. According to the proposed angle estimation algorithm first, the Fourier spectrum of the motion blurred image is computed. Figure 3c is the centered Fourier transform spectrum of the image (b). Using Otsu's method, log spectrum of the image is converted into the binary spectra. Spectrum image (d) is the binary log spectra of the centered Fourier transform. In Fig. 3e, x-axis contains the value of angles (0–180°) and y-axis represents the values of Radon transform of binary log spectra. The peak value at 90° of a graph is the blur angle.

Table 1 shows the results of blur angle (θ) and blur length (L) estimation. Blur length estimation algorithm is depending on the previously estimated blur angle value. It is observed that few degrees of error in angle estimation also cause an error in length estimation. PSF cannot be constructed properly using faulty values as it will create the ringing artifacts as well as the worst restoration of palm images. To overcome this problem, five nearby values of angles are recorded. Five blur lengths are calculated using these five blur orientations; similarly, five PSFs are

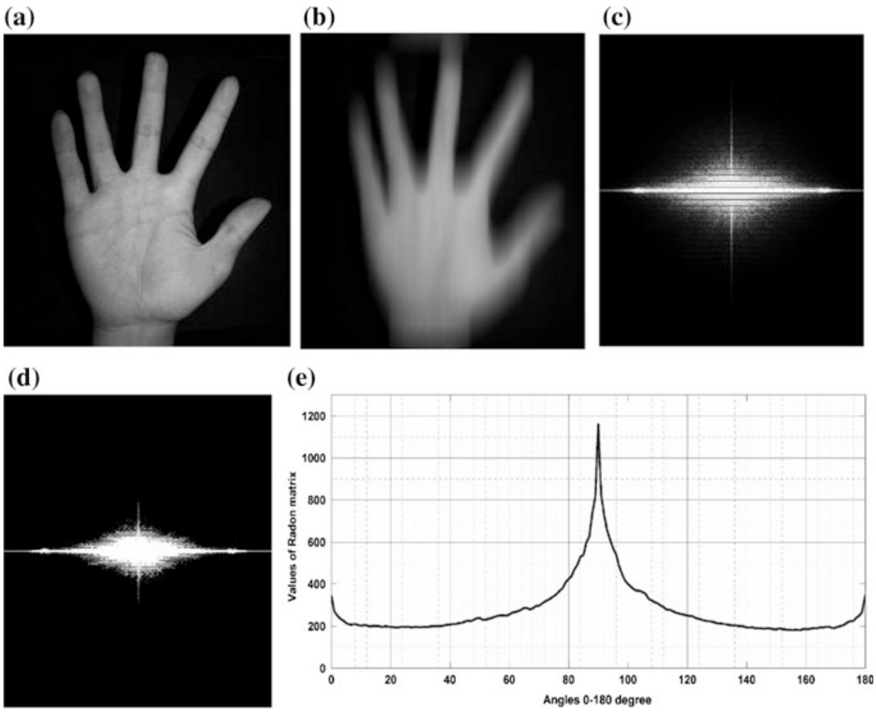


Fig. 3 **a** Original image, **b** motion blurred image by θ = 90° and L = 200, **c** centered Fourier transform spectrum, **d** binary log spectrum, and **e** Radon transform plot

Table 1 Blur parameter estimation results

Actual blur		Estimated blur parameters					
		Person-1		Person-2		Person-3	
L	θ	θ	L	θ	L	θ	L
10	4	0, 180, 1, 179, 2	10, 9, 21, 33, 20	0, 180, 1, 179, 2	10, 9, 21, 33, 30	0, 180, 1, 179, 2	10, 9, 11, 33, 10
15	70	67, 70, 69, 68, 66	16, 15, 16, 17, 14	68, 67, 69, 66, 70	17, 16, 16, 14, 15	68, 67, 66, 69, 70	17, 16, 14, 16, 15
20	17	180, 0, 16, 15, 17	18, 19, 22, 20, 20	180, 0, 16, 17, 15	18, 19, 22, 20, 20	180, 0, 14, 16, 15	18, 19, 20, 22, 20
40	5	180, 0, 1, 4, 5	39, 40, 41, 41, 40	180, 0, 1, 4, 5	39, 40, 41, 41, 40	180, 0, 1, 4, 5	39, 40, 41, 41, 40
50	30	27, 28, 26, 29, 30	51, 51, 50, 50, 51	27, 28, 26, 29, 25	50, 51, 50, 50, 50	28, 27, 26, 29, 25	51, 50, 50, 50, 50
66	10	180, 0, 9, 8, 10	64, 65, 66, 68, 66	9, 0, 180, 8, 10	66, 65, 64, 68, 66	180, 0, 9, 8, 10	64, 65, 66, 68, 66
70	120	123, 122, 124, 125, 121	70, 69, 69, 69, 70	123, 122, 124, 125, 121	70, 69, 68, 69, 70	123, 122, 124, 121, 125	70, 69, 68, 70, 69
90	160	162, 161, 163, 160, 164	89, 89, 89, 91, 91	162, 161, 163, 160, 164	89, 89, 89, 90, 91	162, 161, 163, 160, 164	89, 89, 89, 90, 91
100	0	0, 180, 1, 179, 178	100, 99, 124, 123, 144	0, 180, 179, 1, 2	100, 99, 123, 124, 145	0, 180, 179, 1, 2	100, 99, 123, 124, 145
200	90	90, 91, 89, 88, 92	200, 219, 200, 200, 240	90, 89, 91, 88, 92	200, 200, 219, 200, 240	90, 91, 89, 92, 88	200, 219, 200, 240, 200

constructed. PSFs are used to deblur the input blurred palm image. At the end, five deblurred images are stored with PSNR and SSIM index. The image which has the highest SSIM is chosen for the biometric identification.

As stated before in some cases, exact value of angle may lie under any of nearby peak. Figure 4a shows the five maximum values for the blur angle estimation for Person-2, $\theta = 5°$ and $L = 40$ as shown in Table 1. At the highest peak, estimated

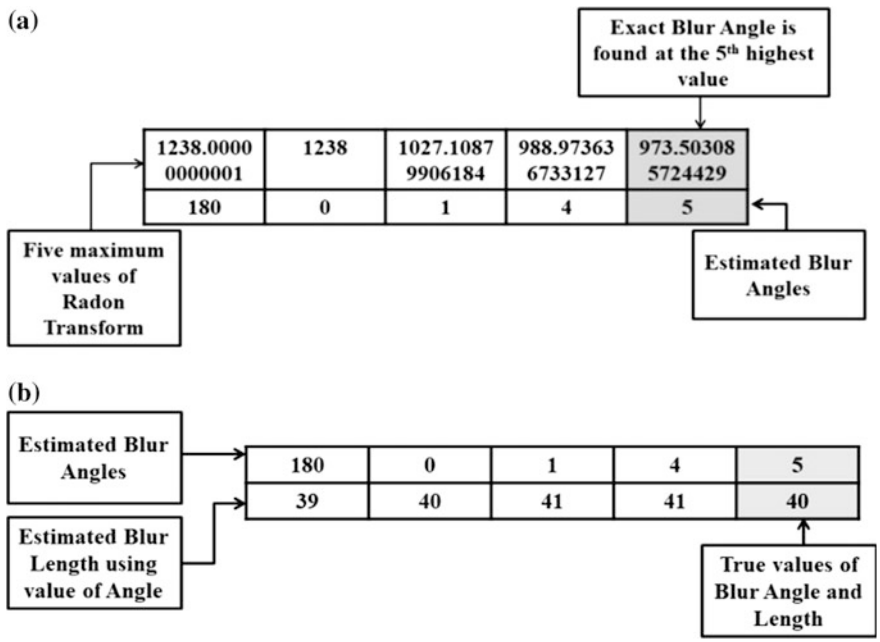


Fig. 4 **a** Blur angle estimation for motion blurred image with θ = 5°, L = 40 and **b** blur length estimation for each value of angle

blur angle is 180°, which is not the correct blur angle. The exact value of blur angle is found at the fifth highest value of the Radon transform. Five estimated angles are: θ = [180, 0, 1, 4, 5]. Using these values estimated blur lengths are shown in Fig. 4b, L = [39, 40, 41, 41, 40]. Truly detected blur parameters are θ = 5° and L = 40. Palmprint image of Person-2 is restored by all the estimated values of angle and length. Five PSFs are constructed using values of blur parameters. Results of restoration are shown in Fig. 5. Figure 5b–f are restored images with faulty PSF. Image (b) is restored with $\theta = 0°$ and L = 40, (c) is restored with $\theta = 1°$ and L = 41, (d) is restored with $\theta = 4°$ and L = 41, and (f) is restored with $\theta = 180°$ and L = 39. All of these images have ringing artifacts, and texture of palm is not properly restored. Figure 5e is restored with accurate blur angle and length, $\theta = 5°$ and L = 40 having no ringing artifacts with PSNR = 40.5628 and SSIM = 0.9466 (as shown in Table 2 for Person-2) which are comparatively higher than other deblurred images.

Table 2 shows the PSNR and SSIM values of deblurred images of respective persons of Table 1. Figure 6 shows the sample deblurred images. The biometric identification system is tested for motion blurred image with blur length = 100 and blur angle = 0° to reduce the total evaluation time. Because for L = 100 and θ = 0°, blur angle is lies under the first highest values of Radon transform for all cases. Thus, there is no need to check the five values of blur angle.

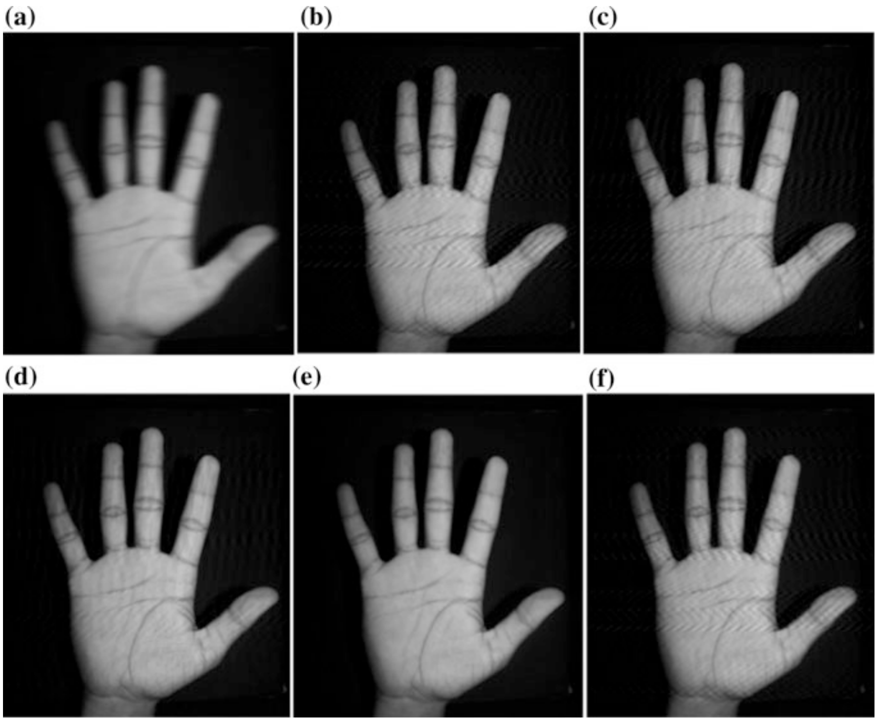


Fig. 5 **a** Motion blurred image by L = 40 and = 5, **b** deblurred image with = 0° and L = 40 (PSNR = 32.0112, SSIM = 0.7816), **c** deblurred image with = 1° and L = 41 (PSNR = 32.0127, SSIM = 0.7646), **d** deblurred image with = 4° and L = 41 (PSNR = 36.8123, SSIM = 0.8790), **e** deblurred image with = 5° and L = 40 (PSNR = 40.5628, SSIM = 0.9466), and **f** Deblurred image with = 180° and L = 39 (PSNR = 32.0244, SSIM = 0.7727)

Table 2 Quality assessment parameters results

L/θ	Quality assessment parameters					
	Person-1		Person-2		Person-3	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
10/4	41.6469	0.9555	41.0933	0.9556	39.4056	0.9484
15/70	42.0897	0.9585	41.4443	0.9578	39.4285	0.9475
20/17	42.0114	0.9573	41.3444	0.9563	38.4009	0.9347
40/5	41.0917	0.9465	40.5628	0.9466	38.4494	0.9326
50/30	39.2686	0.9139	39.0283	0.9188	37.2039	0.9060
66/10	40.6110	0.9401	40.1879	0.9414	38.0738	0.9257
70/120	37.6317	0.8770	37.2028	0.8781	35.8916	0.8674
90/160	38.5749	0.8949	39.7953	0.9361	37.5588	0.9166
100/0	41.5604	0.9517	40.9837	0.9516	38.9148	0.9397
200/90	40.2460	0.9371	39.7645	0.9379	37.6957	0.9199

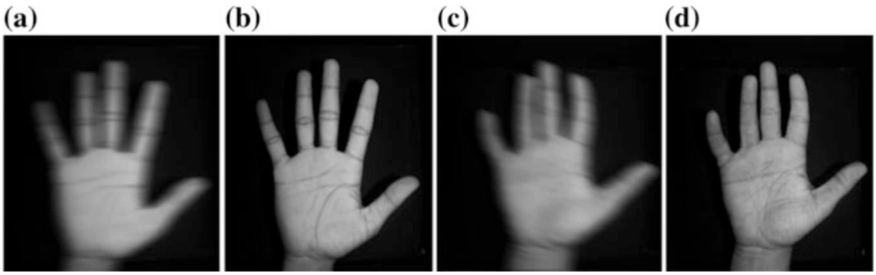


Fig. 6 **a** Motion blurred image with θ = 0° and L = 100, **b** deblurred image from (**a**), **c** motion blurred image with θ = 160° and L = 90, **d** deblurred image from (**c**)

5 Identification Results

We have used the database of [1] 100 persons with the age group of 18–50. A total of eight right-hand images for each person have been collected. Thus, 800 hand images are used to test the identification algorithm. A total of 53 (17 hand geometry + 36 palmprint) features are extracted for identification. The feature vector database is created using 'k'—hand images where 'k' = 2, 4 or 6 (Table 3). The average value of respective standard deviations is taken as a final feature vector, when more than one image is in operation. The reverse is the case for testing sets.

To test the robustness of algorithm, every test image was first blurred using parameters: blur length = 100 and blur angle = 0°. As shown in system block diagram (Fig. 1), the deblurring algorithm is used to deblur the palmprint images. The feature vector is calculated for every deblurred image and compared with all feature vectors stored in the database. Minimum Euclidian distance is calculated from these comparisons for identification of the person.

The % genuine acceptance rate (GAR) is shown in Table 4 for identification only by palmprint features. We achieved 97% GAR using DCT-based palmprint features with more images for training to have more features in the database.

Table 3 Number of images used for training and testing purpose

k	Training set (TR)	Testing set (TS)
2	2 × 100 = 200	6 × 100 = 600
4	4 × 100 = 400	4 × 100 = 400
6	6 × 100 = 600	2 × 100 = 200

Table 4 Comparison of % GAR for only palmprint-based identification

Method	TR = 2 TS = 6	TR = 4 TS = 4	TR = 6 TS = 2
Palmprint (DFT)	90.1667	90.2500	94.0000
Palmprint (DWT)	81.0000	83.5000	88.0000
Palmprint (DCT)	90.1667	91.5000	**97.0000**

To develop the more secure system, fusion at decision level is done in which separate authentication decision is made for each unimodal biometric system. These decisions are finally combined for identification. Table 5 shows the % GAR for person identification by combining hand geometry (HG) and palmprint (Fig. 7).

From the results, it can be observed that the bimodal biometric systems give better recognition rate compared to unimodal systems. Here, recognition rate for low TR is significantly improved from 90.1667% (palmprint only, Table 4) to 95.5% (HG and palmprint) for DCT. It can be observed that the DCT gives more recognition rate compared to DFT and DWT. The maximum recognition rate obtained is 99.5% for DCT-based bimodal biometric system.

For imposter analysis, out of 100 persons, 80 persons are considered as genuine users and other 20 persons are considered as imposter. The feature vector for the database is created by 80 genuine users, and comparison for identification process is done for all 100 users which include 20 imposter users also. Here, threshold is decided based on the minimum error vector which is obtained during the comparison.

Table 5 Comparison of % GAR for hand geometry- and palmprint-based identification

Method	TR = 2 TS = 6	TR = 4 TS = 4	TR = 6 TS = 2
HG & palmprint (DFT)	95.5000	95.0000	99.5000
HG & palmprint (DWT)	92.3333	93.0000	94.5000
HG & palmprint (DCT)	95.5000	96.7500	**99.5000**

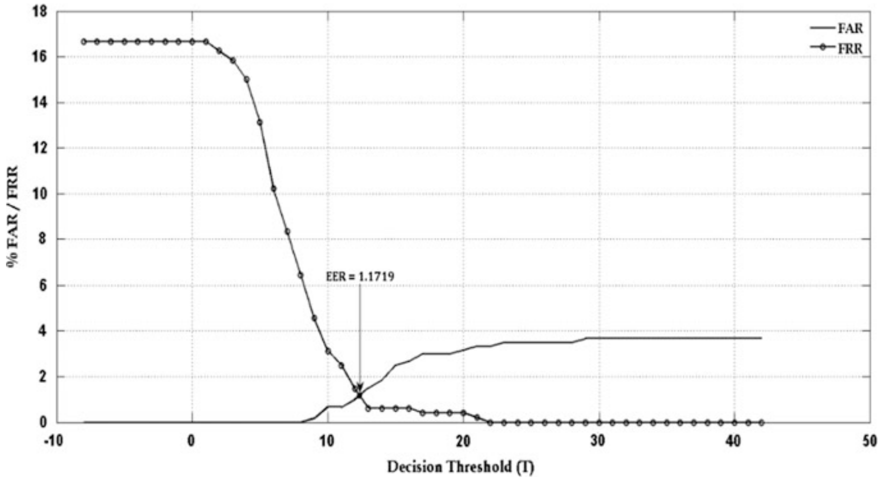


Fig. 7 Threshold versus FAR/FRR for HG and palmprint based identification system

6 Conclusion

In this paper, Radon transform is used for the blur angle estimation with some preprocessing steps to achieve less error in angle estimation. Exact estimation of blur angle improves the result of length estimation using cepstral method. Restored images using accurate blur parameters are free of any ringing artifacts. Palm texture and boundaries of palm images are preserved in the deblurred images. Maximum PSNR achieved is approximately 42 dB, and SSIM = 0.96 is achieved using the proposed algorithm. Previous researchers gave deblurring of palm images, but it was not tested against the identification systems. The proposed system gives 99.5% for recognition rate against blur palm images. Also, the EER for the bimodal identification system achieved is 1.1719.

References

1. Dale, M.P., Joshi, M.A., Galiyawala, H.J.: A single sensor hand geometry and palm texture fusion for person identification. Int. J. Comput. Appl. **42**(7), 11–16 (2012)
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Prentice Hall (2002)
3. Shah, M.J., Dalal, U.D.: Hough transform and cepstrum based estimation of spatial-invariant and variant motion blur parameters. In: 2014 International Conference on Advances in Electronics, Computers and Communications (ICAECC), pp. 1–6. IEEE, Bangalore, India (2014)
4. Heflin, B., Parks, B., Scheirer, W., Boult, T.: Single image deblurring for a real-time face recognition system. In: IECON 2010—36th Annual Conference on IEEE Industrial Electronics Society, pp. 1185–1192. IEEE, Glendale, USA (2010)
5. Tiwari, S., Shukla, V.P., Singh, A.: Certain investigations on motion blur detection and estimation. In: Proceedings of International Conference on Signal, Image and Video Processing, pp. 108–114, Patna (2012)
6. Shen; Y., Dang, J., Lei, T., Luo, W.: Motion blur parameters estimation based on frequency and spatial domain analysis. In: 2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 387–390. IEEE, Yichang, China (2012)
7. Deshpande, A.M., Patnaik, S.: Radon transform based uniform and non-uniform motion blur parameter estimation. In: 2012 International Conference on Communication, Information & Computing Technology (ICCICT), pp. 1–6. IEEE, Mumbai, India (2012)
8. Shakeel, M.S., Kang, W.: Efficient blind image deblurring method for palm print images. In: 2015 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), pp. 1–7. IEEE, Hong Kong, China (2015)
9. Mughaddam, M.E., Jamzad, M.: Finding point spread function of motion blur using radon transform and modeling the motion length. In: 2004 Proceedings of the Fourth IEEE International Symposium on 4th Signal Processing and Information Technology, pp. 314–317. IEEE, Rome, Italy (2004)
10. Cho, T.S., Pari, S., Horn, B.K.P., Freeman, W.T.: Blur kernel estimation using the radon transform. In: 2011 IEEE Conference on 4th Computer Vision and Pattern Recognition (CVPR), pp. 241–248. IEEE, Colorado Springs, CO, USA (2011)
11. Cai, C., Liu, A., Zhang, B.: Motion deblurring from a single image. In: 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), pp. 406–410. IEEE, Nanchang, China (2016)

12. Wu, W.W., Zhong, J.H., Wang, Z.Y.: A new method for restoration of defocused image. In: 2010 International Conference on 9th Machine Learning and Cybernetics (ICMLC), pp. 2402–2405. IEEE, Qingdao, China (2010)
13. Mirel, I.: Non-iterative deblurring for image restoration. In: 2012 10th International Symposium on Electronics and Telecommunications (ISETC), pp. 319–322. IEEE, Timisoara, Romania (2012)
14. Mesarovic, V., Galatsanos, N., Katsaggelos, A.: Regularized constrained total least-squares image restoration. IEEE Trans. Image Process. **4**(8), 1096–1108 (1995)
15. Chan, S.H., Khoshabeh, R., Gibson, K., Gill, P., Nguyen, T.: An augmented lagrangian method for video restoration. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 941–944. IEEE, Prague, Czech Republic (2011)
16. Stanley, H.C., Khoshabeh, R., Gibson, K.B., Gill, P.E., Nguyen, T.Q.: An augmented lagrangian method for total variation video restoration. IEEE Trans. Image Process. **20**(11), 3097–3111 (2011)
17. Anitha, M.L., Rao, K.A.R.: A novel bimodal biometric identification system based on finger geometry and palm print. In: 2014 19th International Conference on Digital Signal Processing (DSP), pp. 574–579. IEEE, Hong Kong, China (2014)
18. Goh, M.K.O., Connie, T., Teoh, A.B.J., Ngo, D.C.L.: A fast palm print verification system. In: 2006 International Conference on Computer Graphics, Imaging and Visualisation, pp. 168–172. IEEE, Sydney, Australia (2006)
19. Mansoor, A.B., Mumtaz, M., Masood, H.: Directional energy based feature level multimodal system using palm and fingerprints. In: 2010 International Workshop on Emerging Techniques and Challenges for Hand-Based Biometrics (ETCHB), pp. 1–4. IEEE, Istanbul, Turkey (2010)
20. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**(4), 600–612 (2004)
21. Fu, Y., Wang, S.: No-reference image quality assessment based on HSV. In: International Symposium on Computer, Consumer and Control (IS3C), pp. 1093–1096, IEEE (2016)
22. Kamble, V., Bhurchandi, K.M.: No-reference image quality assessment algorithms: a survey. Int. J. Light and Electron Opt. **126**(11), 1090–1097. Elsevier (2015)
23. Zhang, C., Shi, Z., Xu, B., Feng, B.: No-reference image quality assessment for defocus restoration. In: 2016 First International Conference on Multimedia and Image Processing (ICMIP), pp. 51–56. IEEE, Bandar Seri Begawan, Brunei (2016)

A Low Power, High Swing and Robust Folded Cascode Amplifier at Deep Submicron Technology

Varsha Bendre, Abdul Kadir Kureshi and Saurabh Waykole

Abstract Single-pole operational amplifiers are majorly characterized as constant gain and stable operational amplifiers. Folded cascode amplifier falls in such category of single-pole operational amplifier with high gain and improved output swing, maintaining constant phase margin for stable operation. Stability is even better in this case, compared to multistage amplifiers. The focus of this paper is to achieve deep negative feedback, to increase the small-signal gain and the output voltage swing as much as possible along with less power dissipation. The designed folded cascode amplifier is capable of providing 67.44 dB gain, 1.77 V output swing while dissipating 0.13 mW power and at the same time ensuring the higher margin of stability to be 62.58° at typical voltage for 180 nm technology. This amplifier is also tested for variability analysis for temperature and voltage range, and it is robust since it exhibits desired response for 0–125 °C temperature range and can work up to 0.9 V supply voltage satisfactorily which exhibit reduction in power dissipation.

Keywords Operational amplifier · Gain · Stability · Output swing Robust · Variability · Transconductance · Scaling · Monte Carlo simulations Analysis

V. Bendre (✉)
Rajarshi Shahu College of Engineering, Savitribai Phule Pune University, Pune, India
e-mail: varshabendre22@gmail.com

A. K. Kureshi
Vishwabharati Academy's College of Engineering, Ahmednagar, India
e-mail: akkureshi@rediffmail.com

S. Waykole
Pimpri Chinchwad College of Engineering, Pune, India
e-mail: saurabh2749@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_59

1 Introduction

In this era of high performance and microdevices, CMOS transistor scaling has almost reached its limits. There has been tremendous work carried by researchers in digital domain as well as analog domain. Still, there is a challenge as far as mixed-signal design is considered in today's competitive world. There is an exemplary need of analog front end since all the real-time signals are analog in nature, and hence, researchers are probing in finding ways to develop high-performance portable devices with compact size and capable of operating at low power.

The operational amplifier (op amp) is the most versatile block used in analog and mixed-signal integrated circuits applications. Power dissipation is a major challenge in the deep submicron technology and is becoming more important in portable applications to improve battery life span. At the same time, it is also a challenge to maintain the speed and area requirements. So it is a challenge to design low-power op amps with significant speed and other performance parameters like high output swing, phase margin, and sufficient gain–bandwidth product. There are basic three kinds of OTA topologies, classical two-stage, folded-cascode OTAs, and telescopic OTAs.

In this paper, the design of folded-cascode op amp with wide-swing current mirror is presented in Sect. 2. In Sect. 3, variability issues are discussed, and Sect. 4 presents experimental setups and corresponding simulation results. Finally, Sect. 5 provides the conclusion and future scope of the presented work.

2 Folded-Cascode Op Amp

Due to shrinking in channel length of the transistors in nanometer, it becomes tougher to attain sensible op amp gains due to various non-ideal effects and parasitics [1]. Hence, in analog and mixed-signal designs, wide-swing cascode current mirrors are often used. Buffered op amps are used to drive resistive loads, but many contemporary integrated CMOS op amps need to drive only capacitive loads. Unbuffered op amps or OTAs are preferred in such modern high-speed CMOS op amps which are used in low-power electronics applications. Consequently, it becomes possible to implement faster op amps with superior signal swings compared to those driving resistive loads.

This can be achieved by using self-compensating op amps driving only capacitive loads. The principle of this op amp is to have only a single high-impedance node at the output of an op amp. The impedance observed at all other nodes is relatively low [2]. Due to this, the speed of the op amp is enhanced. As a consequence, voltage swing increases at the output node than voltage signal at all nodes. These types of op amps are usually more stable but also slower due to increase in self-compensating capacitance value. These types of op amps are usually referred as folded-cascode op amp, as shown in Fig. 1. These structures are also preferred over telescopic topologies for larger output swing, reasonable gain, and stable phase margins.

The circuit, shown in Fig. 1, is a folded-cascode op amp with wide-swing cascode current mirrors instead of conventional current mirrors [3]. The purpose of using such mirrors is to increase output impedance, for boosting the dc gain of the op amp. In folded-cascode op amp, input differential transistors and cascode transistors are of different types. Due to this, its gain can be quite reasonable, though it is basically a single gain stage. Such folded structures are easy to design and help in getting wider output swing with improving common mode voltage range [4].

This input common mode range of the op amp designed in this paper is about 1 V, and output swing is up to 1.77 V for typical TSMC 180 nm technology. Wide-swing current mirror is used for differential-to-single-ended conversion and is realized by the transistors M7, M8, M9, and M10. The dominant-pole compensation is achieved by the load capacitor.

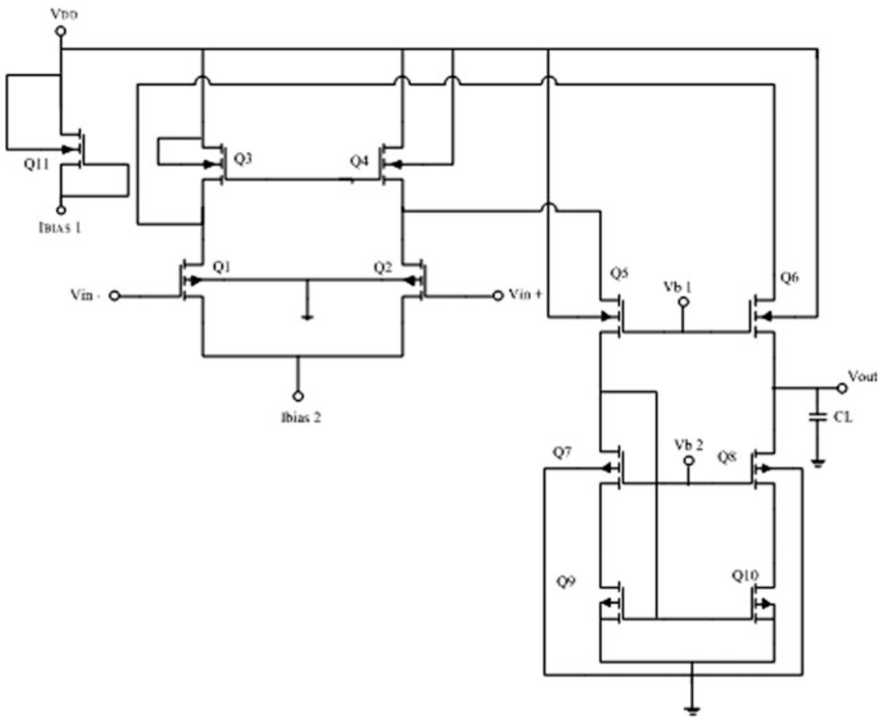


Fig. 1 Folded-cascode op amp

2.1 Small-Signal Analysis

In this analysis unlike conventional op amp, it is presumed that the current from M1 directly flows through the drain of M6 and thus to the load capacitance and the current from M2 goes indirectly through M5 and the current mirror consisting of M7–M10. Also, it is assumed that maximum amount of current flows through M1, and hence, these two paths have slightly different transfer functions. The high-frequency poles and zeros are non-dominant and can be ignored, because they are located at high frequency compared to unity gain frequency. An approximate small-signal transfer function for the folded-cascode op amp is given by:

$$\mathrm{Av} = \frac{\mathrm{Vout(s)}}{\mathrm{Vin(s)}} = \mathrm{gm}_1 \cdot \mathrm{Z_L(s)} \tag{1}$$

where gm_1 is the amplifier's transconductance gain and $\mathrm{Z_L(s)}$ is the output impedance.

The open-loop gain of op amp is further calculated as:

$$\mathrm{Av} = \frac{\mathrm{gm}_1 \cdot \mathrm{rout}}{1 + \mathrm{s} \cdot \mathrm{rout} \cdot \mathrm{CL}} \tag{2}$$

where rout is the output impedance of the op amp and CL is the load capacitance.

For high frequencies, the load capacitance dominates, and hence

$$\mathrm{Av} = \frac{\mathrm{gm}_1}{\mathrm{sCL}} \tag{3}$$

The gain–bandwidth product of the op amp is given as

$$\mathrm{UGB} = \frac{\mathrm{gm}_1}{\mathrm{CL}} \tag{4}$$

The input transconductance can be increased by using long-channel transistors and ensuring that the input transistor pair's bias current is significantly larger than the cascode transistors bias current. This will also result in improvement of bandwidth. To maximize the dc gain of the designed op amp, it is considered that the current flowing through all the transistors connected to output node is at small levels. This will not only maximize the input transconductance but also maximize output impedance. Maximum amount of bias current through the input differential pair results in large transconductance of the input devices and hence in the improvement of thermal noise performance of op amp.

Slew rate of the op amp is given by

Table 1 Aspect ratios of transistors

Transistors	Aspect ratios	Transistors	Aspect ratios
M1	250	M7	5
M2	250	M8	5
M3	100	M9	5
M4	100	M10	5
M5	10	M11	10
M6	10		

$$SR = \frac{ID4}{CL} \tag{5}$$

Power dissipation in the op amp is given by

$$\mathrm{Pd} = \mathrm{Vdd} \cdot 2(\mathrm{ID1} + \mathrm{ID6}) \tag{6}$$

Using standard technology equations and corresponding EDA tool specifications, circuit theory equations, aspect ratios of the transistors are calculated and listed in Table 1.

3 Variability-Aware Design

Variability-aware designs have always been an issue in analog integrated circuit design. With continuous scaling, short-channel effects start to have a significant influence on the transistor characteristics. Not only this, but many manufacturing variations and process variations can create a fluctuation in physical properties of fabricated devices. In addition to this, transistor performance gets pretentious by voltage and temperature variations during chip functioning. Process, voltage, and temperature (PVT) variations set up among the immense challenges in path of transistor scaling [5].

With technology moving toward the nanometer regime, process variability due to the fabrication steps or random variations from dopant fluctuations becomes an increasingly critical concern. The electrical parameter variations of a device are mainly contributed by various fabrication steps such as oxidation, ion implantation, lithography. With various high-speed technological inventions, random variations in circuit such as the temperature and the power supply voltage (Vdd) increase significantly. This results in considerable variations in the circuit performance and degradation in yields which increases manufacturing expenses [1, 6].

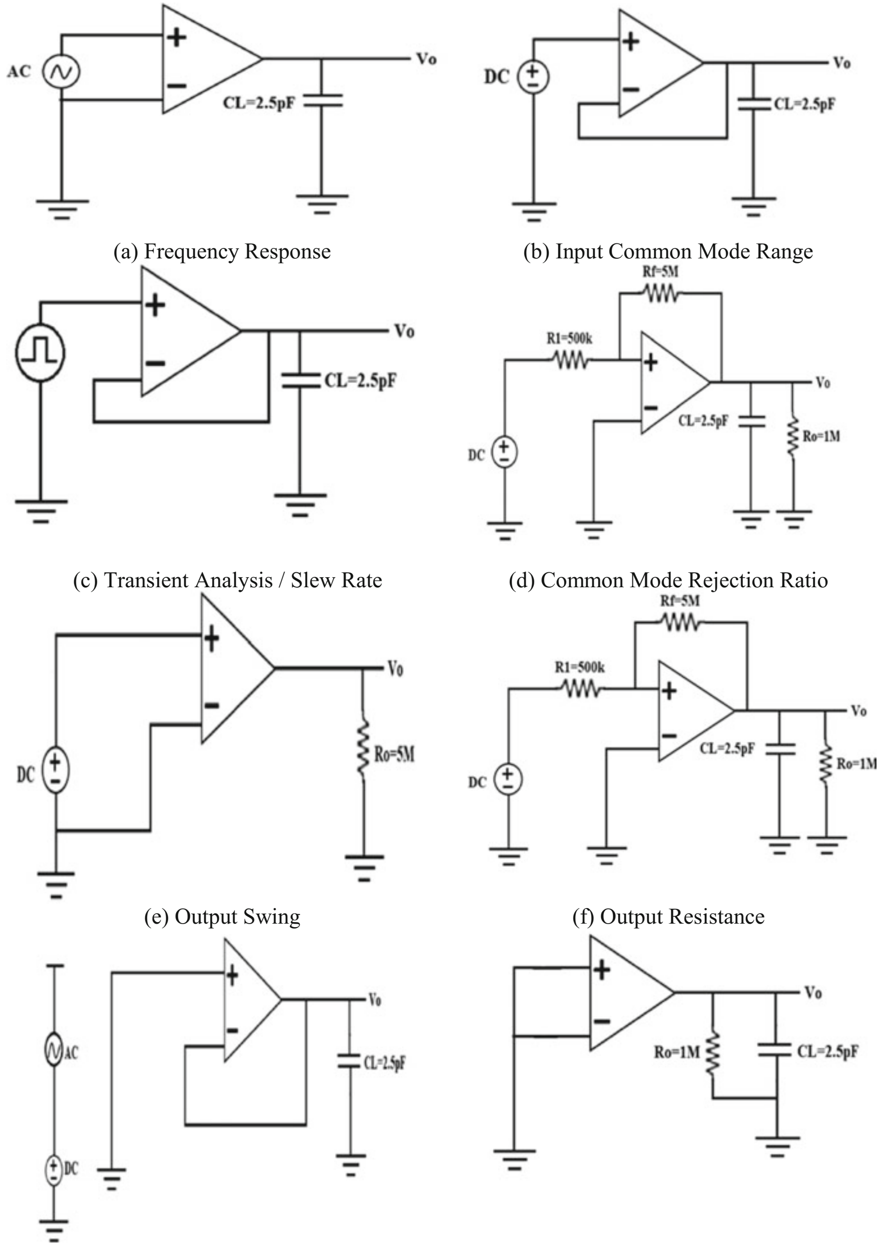
There may be minor deviations in supply voltage from the ideal value of design during routine operation due to environmental conditions on chip. The power supply may not be constant throughout the chip, and hence, there may be significant variation in propagation delay over the entire chip. There may also be voltage plunge due to nonzero resistance in the supply chains. The self-inductance of a supply line contributes also to a voltage drop. These fluctuations in supply voltage are mostly due to signal integrity issue resulting from wire resistance, called as IR drop and inductive noise, termed as di/dt noise.

Due to the power dissipation of transistors, there is variation in temperature which is unavoidable. The mobility of carriers decreases with increased temperature, depending on the doping concentration. For higher doping concentrations, the starting temperature from which mobility starts decreasing is higher. The propagation delay increases due to reduction in mobility of electrons and holes. The threshold voltage of a transistor is inversely proportional to the temperature. Hence with an increase in temperature, threshold voltage reduces and current drive increases which results in a better delay performance. Hence, the propagation delay is majorly influenced by the variation in temperature.

4 Simulation Results

The folded-cascode OTA with wide-swing cascode current mirror is designed using conventional methodology and simulated in Mentor Graphics Eldo with BSIM3v3.3 level 53 models at 180 nm CMOS technology. The op amp operates with the 1.8 V power supply and dissipates only 131 μW power. Simulations results for DC, AC, and transient analyses are carried out, and this section explains test benches for calculation of various electrical characteristics of op amp as shown in Fig. 2.

Due to miniaturization, the effects of variations in process, supply voltage, and temperature (PVT) are becoming a most important hindrance for optimum performance [5]. Therefore, the probable performance deprivation due to PVT variations has become a major decisive factor in evaluating the performance of a new technology [7]. Designed OTA is tested for random variations like supply voltage and temperature variations of chip. Fig. 3b, d, f shows the effect of supply voltage variation on DC, transient and AC response of op amp respectively. Supply voltage is scaled step by step up to 1 V, and results of variations are observed on DC analysis and transient analysis using 100 iterations of Monte Carlo simulations. The result shows op amp can work satisfactorily up to supply voltage of 1 V, without degrading other parameters much. Similarly, temperature is varied over the positive range from 0 to 125 °C for 100 iterations using Monte Carlo simulations, and results show negligible effect on DC, transient and AC analysis of op amp as shown in Fig. 3a, c, e respectively.



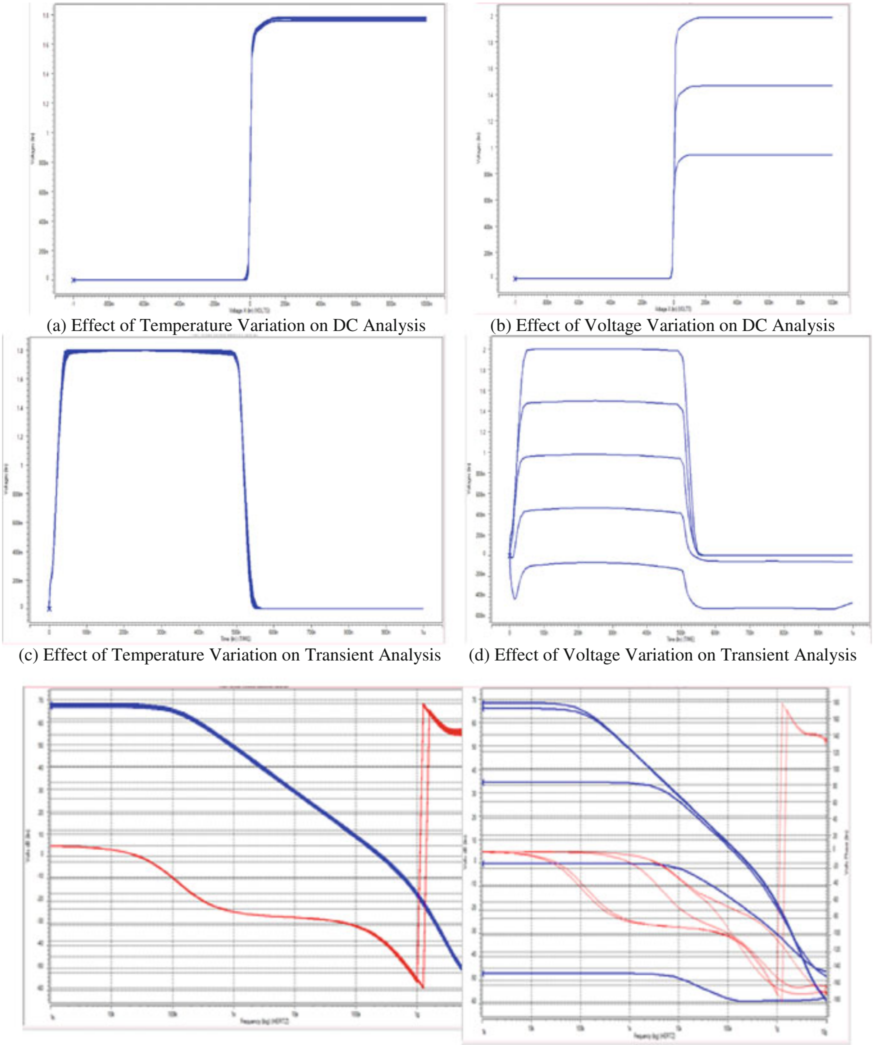
(a) Frequency Response (b) Input Common Mode Range

(c) Transient Analysis / Slew Rate (d) Common Mode Rejection Ratio

(e) Output Swing (f) Output Resistance

(g) Power Supply Rejection Ratio (h)Offset Voltage

Fig. 2 Test benches for various performance parameters of op amp



(a) Effect of Temperature Variation on DC Analysis

(b) Effect of Voltage Variation on DC Analysis

(c) Effect of Temperature Variation on Transient Analysis

(d) Effect of Voltage Variation on Transient Analysis

(e) Effect of Temperature Variation on AC Analysis

(f) Effect of Voltage Variation on AC Analysis

Fig. 3 Effects of temperature and voltage variations on DC, transient and AC analysis

Simulation results are carried out at 180 nm technology node and summarized in Table 2. Table 2 also shows comparison of classical OTA [8] with respect to different performance parameters.

Table 2 Results of folded-cascode op amp at 180 nm technology

Performance parameter	Simulation results			
	[This work] 1.8 V	[This work] 1 V	[8] 1.8 V	[8] 1.2 V
Gain (dB)	67.44	52.2	75.1	65.27
Unity gain–bandwidth (MHz)	255	217	30.5	29.24
Phase margin (°)	62.58	71	53.8	54.17
Slew rate (V/μs)	69.4	20	0.37	0.2
ICMR (V)	0–1	0–0.9	0–1.05	0–0.8
CMRR (dB)	88	88	77.8	69.5
Power dissipation (μW)	131	–	536.5	352.7
PSRR (dB)	60	40	–	–
Output swing (V)	1.77	0.94	1.75	0.8
Output resistance (KΩ)	2	2	–	–

5 Conclusion

In this paper, folded-cascode CMOS OTA is implemented at 180 nm using Mentor Graphics. Significant amount of stability and significant reduction in power is observed compared to conventional OTA implemented using the same technology. With high-swing cascode current mirror in the second stage, gain, phase margin, slew rate, PSRR, unity gain–bandwidth, output swing, and output resistance are at acceptable level even though the system is single pole. Also the effect of temperature variations and supply voltage variations is compared for DC analysis and transient analysis of folded-cascode OTA, and it is found that the designed OTA is capable of giving high swing and is robust.

References

1. Haron, N.Z., Hamdioui, S.: Why is cmos scaling coming to an end? In: 3rd International Design and Test Workshop, pp. 98–103, Dec 2008
2. Allen, P.E., Holberg, D.R.: CMOS Analog Circuit Design, pp. 310–333. Oxford University Press, New York (2002)
3. Johns, D.A., Martin, K.: Analog Integrated Circuit Design, pp. 137–140. Wiley (2008)
4. Zhang, K., Wu, D.: A high-performance folded cascode amplifier. Energy Procedia (2011)
5. Wirnshofer, M.: Variation-Aware Adaptive Voltage Scaling for Digital CMOS Circuits, pp. 5–14. Springer, Dordrecht, The Netherlands (2013)
6. Shah, P., Neema, V., Daulatabad, S.: Effect of process, voltage and temperature (PVT) variations in LECTOR-B technique at 70 nm technology node. In: IEEE International Conference on Computer, Communication and Control, Indore, pp. 1–6, Sept 2015
7. Cho, G.: Assessment of CNTFET based circuit performance and robustness to PVT variations. In: 52nd IEEE International Midwest Symposium on Circuits and Systems, Aug 2009

8. Bendre, V., Kureshi, A.K.: Performance analysis of operational transconductance amplifier at 180 nm technology. In: Second International Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH), pp. 271–276, Ghaziabad (2016)
9. Al Tarawneh, Z., Russell, G., Yakovlev, A.: An analysis and optimization of the robustness of c-element structures to the effects of process variations. In: Proceeding of 2nd European Workshop on CMOS Variability, Grenoble, France (2011)

Syllable-Based Concatenative Speech Synthesis for Marathi Language

Pravin M. Ghate and Suresh D. Shirbahadurkar

Abstract The objective of speech synthesis is to convert text into speech, using the syllable got best the results in term quality speech. Speech processing plays important role in human–machine interaction. Syllable segmentation is done using the different parameters like cutoff frequency, amplitude, the magnitude of speech input signal. The proposed paper suggests a method of syllable segmentation; it converts the syllable representation to modified syllable waveform clips that can be combined together to produce as sound [1]. Syllables are a better choice for speech synthesis in Marathi language. We do manually identify, segment, and label syllable units from speech data for experiment purpose [2]. The concatenative speech synthesis methods provide highly understandable speech utterance.

Keywords Syllable segmentation · Speech synthesis · Text normalization

1 Introduction

Computer produces the speech when the input is text; this process of generating speech is called speech synthesis [1]. We have taken the database from IIIT Hyderabad for the Marathi Language. A total of 500 sentences are available for speech signal processing. After segmentation, syllable may differ in their size. In this methodology, the system selects the best sequence of basic units to synthesize an intelligent natural sounding speech. In this work, syllables were considered because of quality of speech improved.

P. M. Ghate (✉)
JSPM's Rajarshi Shahu College of Engineering, Pune 411033, Maharashtra, India
e-mail: pmghate@gmail.com

S. D. Shirbahadurkar
Zeal College of Engineering & Research, Nahre, Pune 411041, Maharashra, India
e-mail: shirsd112@yahoo.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_60

2 Syllable-Based Concatenative Speech Synthesis

Concatenative speech synthesis for the Marathi language utilizes syllable, a word as basic unit. By using syllable segmentation algorithm, break whole speech signal into small; the basic unit is called syllable. As we mentioned in previous paragraph, text-to-speech system converts normal language of text into speech sound file [3]. For that, speech signal is broken into the pieces of sound and stored in the database used for speech synthesis. For the conversion of text into speech, the system must generalize all the speech units uniquely. In Indian languages for speech processing, general form of syllable is C*VC*, where C is consonant, V is vowel, and C* indicates the presence of zero or more consonants. There are 36 consonants and 12 vowels in Marathi language. The researchers have formed some syllabification rules to generate computationally reasonable syllables. The written text is converted into syllabic representation which later represents in waveform. These waveforms are stored in database called speech corpus of syllable and then subsequently combined to generate sound.

2.1 *Text Normalization of Speech Signal*

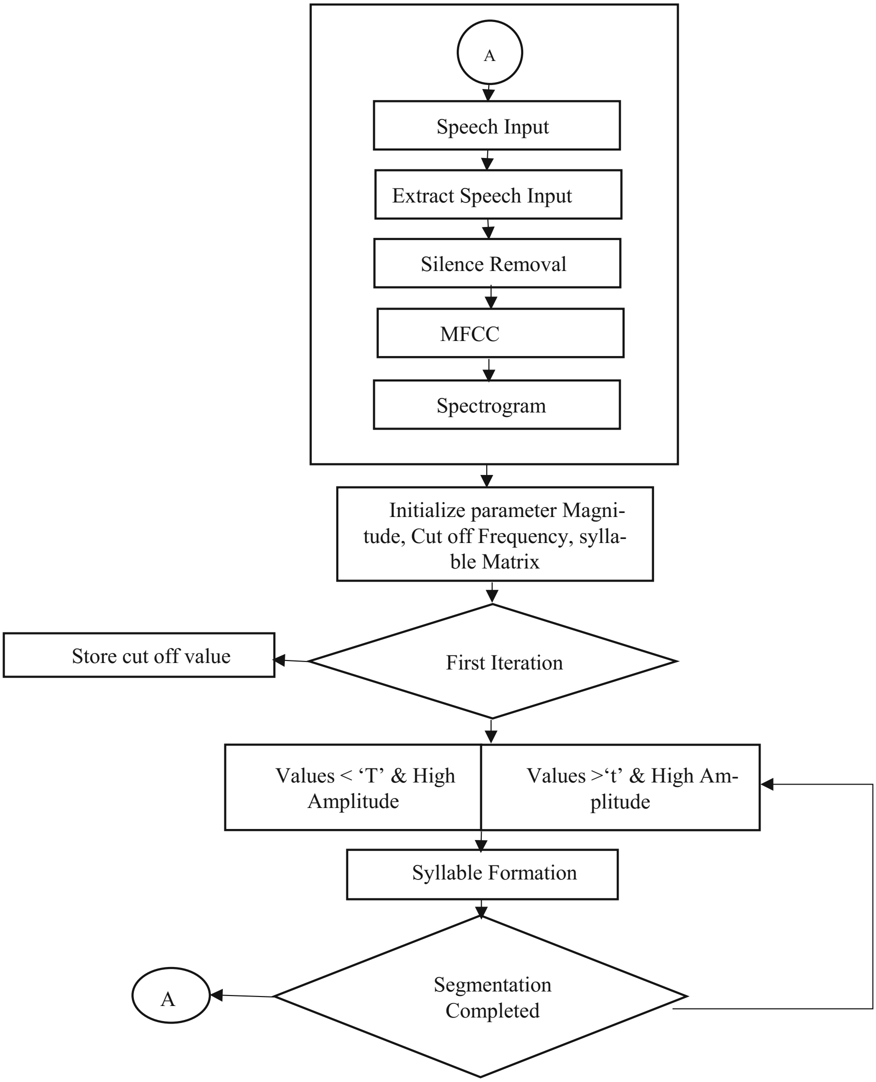
Before actual processing of speech signal, this process of text normalization plays a vital role; input text documents consist of words with the collection of written elements such as statistics, date, time, abbreviations, numbers, symbols. In written language, the text consists of abbreviations and numbers; then, the system must determine how these abbreviations should be read out [4]. A special pronunciation like abbreviations, acronyms, special symbols should be stored in the lexicon. In text normalization, the first stage is the removal of punctuations like double quotes, full stop, and comma.

The text includes many diverse units like syllable, word; all these diverse units firstly converted into actual utterances. After that, system can synthesize these units as speech. Open and close brackets, equals, greater-than, dollar, percent, pound sign, comma etc., are termed as literal symbols. In text normalization, if these symbol appear in the sentence then it should be deleted form sentence. The explanation for all the literals and numerical representation has to be declared in speech corpus [4].

2.2 *Syllable for Indian Language (Marathi Language)*

A syllable is small unit of written language called as sound of vowel that is generated while pronouncing a word. A linguistic unit could be a little unit of communication that consists of uninterrupted sound. In Marathi, अ, आ, ई, [1] a total of 12 vowels are the waves generated without any constriction by other vocal organs. Syllable is a phonetic-phonological basic unit of a word. It consists of syllable start, syllable core, and syllable end. Syllables are not influenced by neighboring sound elements. The segmentation of syllables is relatively easy. A disadvantage for synthesis purposes is that if only syllables are used, a huge number of syllables are required [5]. Hence in present Marathi TTS system, hybrid (most common words and syllables) approach is used.

2.3 Syllable Segmentation Algorithm Flow



2.4 Syllable Segmentation for Marathi Language Algorithm

There are a few steps for dividing sentence into syllables.

Step 1: Speech signal input
Step 2: Calculate MFCC and plot spectrogram
Step 3: For syllable formation, calculate cutoff values and magnitude
Step 4: If first iteration, calculate cutoff values and store, else continues
Step 5: Syllable formation

In step 1, the speech signal of 16 kHz with duration 8–10 s has been taken for syllable segmentation.

For example: In Fig. 1 as sentence is splinting into the words that are W1, W2, W3, W4, W5, using segmentation algorithm form the syllables S1, S2, S3, S4, S5, S6, S7.

As shown in Fig. 2, the input speech signal in the Marathi language "कारण आपल्याकडे ती पद्धत नाही" was applied to the segmentation algorithm.

The first syllable is separated depending upon the cutoff frequency and amplitude parameter, the first syllable which is as shown in Fig. 3 "का" "रण".

Also we find the amplitude graph as shown in Fig. 4 and frequency in Fig. 5.

In step 2, calculate MFCC of above signal, that is,

1. Frame the signal into short frames.
2. For each frame, calculate the periodogram estimate of the power spectrum.
3. Apply the Mel filter bank to the power spectra, and sum the energy in each filter.
4. Take logarithm of all filter bank energies.
5. Take DCT of the log filter bank energies.
6. Keep DCT coefficients 2–12 and discard the rest.

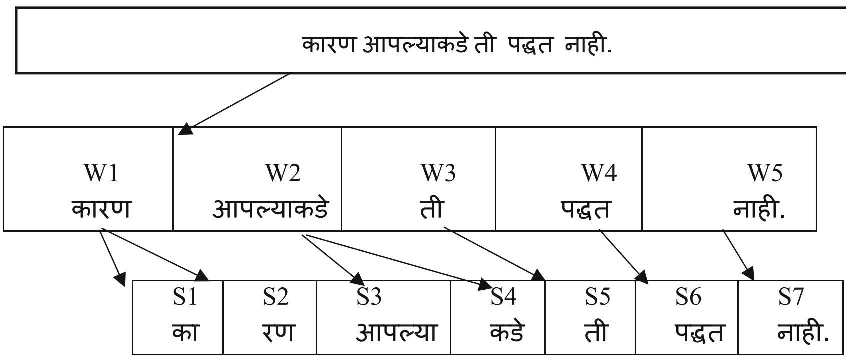


Fig. 1 Sentence segmentation for input speech signal

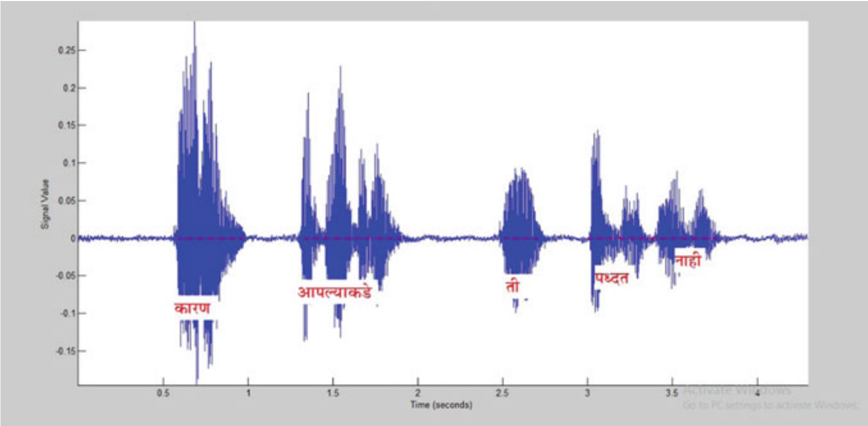


Fig. 2 Input speech signal (full sentence)

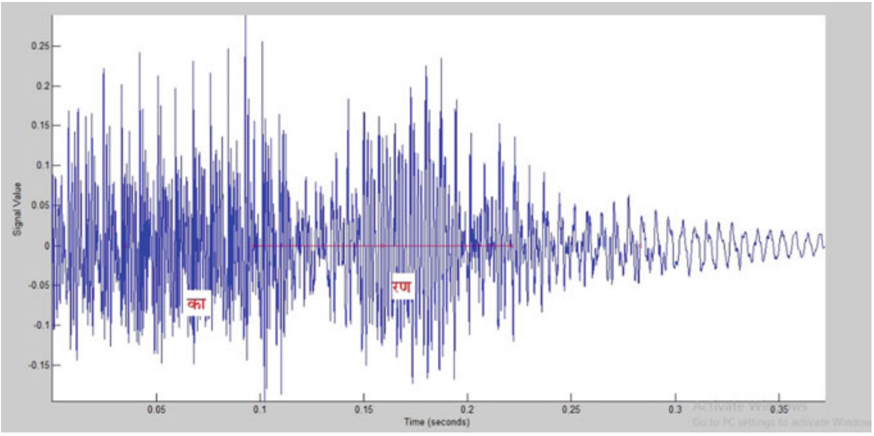


Fig. 3 First syllable of the sentence

The formula for converting from frequency to Mel scale is:

$$M(f) = 1125In(1 + f/700) \tag{1}$$

To go from Mels back to frequency:

$$M^{-1}(m) = 700\left(\exp\frac{m}{1125}\right) - 1 \tag{2}$$

Equation 1 represents the MFCC, and Eq. 2 is used to go back to frequency which is used to find the coefficients. For the segmentation, use the 2–12 coefficients

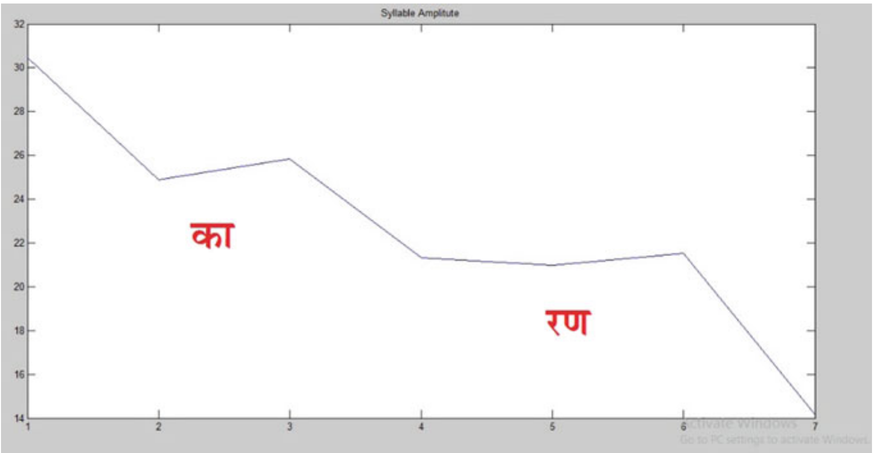


Fig. 4 Amplitude of first syllable

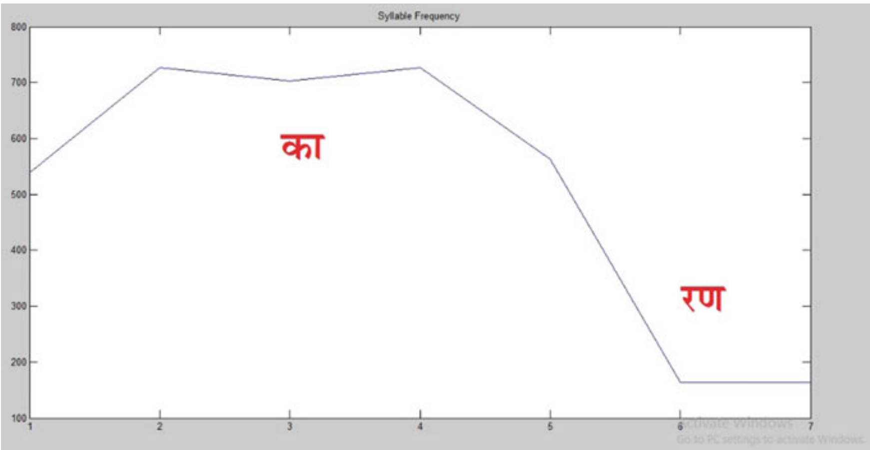


Fig. 5 Frequency of first syllable

In step 3, after calculating the cutoff frequency and finding the amplitude and frequency for speech signal, the whole signal values are stored in one variable for comparing with t and T for setting boundary of the signal. The signal is segmented into the syllable and stored in the syllable corpus.

2.5 Prosody of Speech Signal

Prosody of the speech signal means speed and volume of the word are pronunciation.

The word from sentence needs to be split into the small pieces of sound, and decide which word needs to be emphasized [6]. The prosody is nothing but pitch and placement of pauses in the word so that speech is natural sound [7]. This task is easy for human but difficult for the machine. Prosody plays vital role to understand attitude from the message. The concatenation of speech signal from syllable along with fabrication of pauses in between the word that all pauses depend on the pitch, loudness of signal [8]. Pitch refers the rate of vocal cord vibration. So for regeneration of speech prosody is very important.

3 Waveform Concatenation of Signal

All the segmented data is stored in speech corpus in the form of syllable; for the better quality of speech, concatenate the basic unit of sound with an appropriate place. The quality of speech improves by smoothing the signal. The synthesized speech is produced by combining pieces of prerecorded speech [6]. Overall observation suggests the concatenated synthesis gives better quality speech output. The following steps were followed by the synthesis machine to convert text into speech.

Step 1: The input speech signal that is the sentence was split into words.

Step 2: From word, generate syllable by using the segmentation algorithm.

Step 3: Each syllable unit is chosen to form speech database by identifying its position (start, middle, and end).

Step 4: The frequency and amplitude of syllable clips are changed according to their respective position which is shown in Table 1.

Table 1 Amplitude and frequency values based on position of vowels

Words\position	Class	Start		Middle		End	
		Amp (dB)	Freq (Hz)	Amp (dB)	Freq (Hz)	Amp (dB)	Freq (Hz)
अ, आ, ची,	Class-1	30.2	731.32	19.5	594	11.2	362.7
ती, जी	Class-2	26	697	17.8	653.4	13.2	354.1
हो, वा	Class-3	27.8	692.5	17.3	593	12	280.9
ही, हे, व	Class-4	28	704.1	18	592	11.5	271
ला, ते, ही	Class-5	24.4	661.8	16	506.1	9.4	233.8
क, का, प्र, के	Class-6	37	912	21.7	871.2	13.2	425.3
म, कु, ई	Class-7	26.4	687.2	18.2	481.6	12.9	310.2

Step 5: The syllable-form database is concatenated using waveform concatenation method with respective pause within each word [9]. To avoid discontinuities between syllables, spectral smoothing is applied.

4 Results

The speech signal is divided into the syllable and word. This syllable is divided into the three parts—start, middle, and end parts of signal. By dividing the syllable into the different class, there are different categories from class 1 to class 7. All classes are divided as per the amplitude and frequency which are shown in Table 1. From Figs. 6 and 7, we observe the variation in start, middle, and end. Figure 6 represent the changes in amplitude of start, middle, end word and frequency changes in for the same parameter. A set of 500 sentences were selected for the experiment.

If we observe Table 1, in that there are different values of vowels and consonants of Marathi Language. In class 1 if अ, आ, ची, which are words, are present in the beginning, the frequency is maximum that is 731.32 Hz, if same vowels are present in the middle, then the frequency is 594 Hz, and if these are present in end, then the frequency is 362 Hz. Form this table, it is clear that if vowels are available at the beginning, then more pressure is used by vocal cord. From that frequency, we identify the starting word of sentence. This same method is used for other classes also [10].

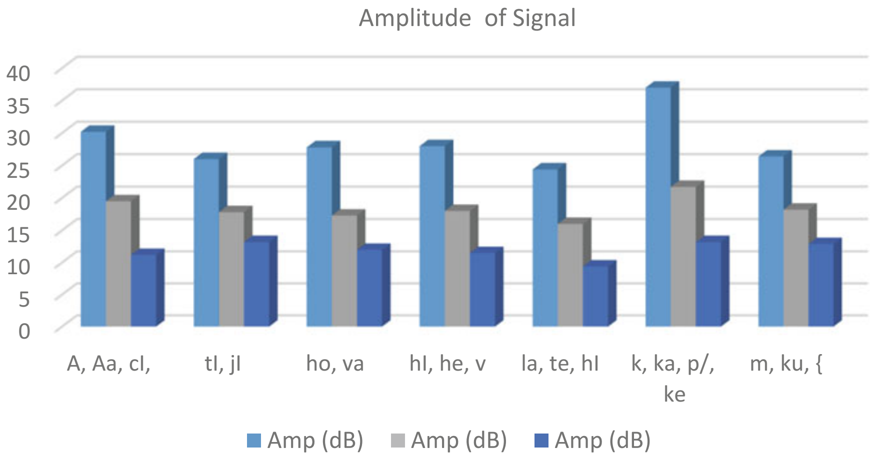


Fig. 6 Amplitude of signal

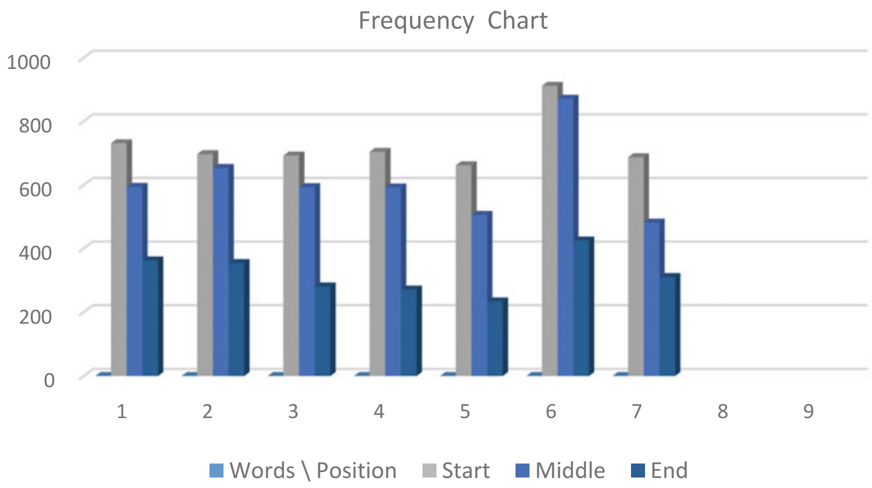


Fig. 7 Frequency of the signal

5 Conclusion

For regional language, syllable-based concatenative speech synthesis for Marathi language produces quality output. Large collection of syllables were collected and stored in a database for further synthesizing speech. That speech can be extracted by concatenating pieces of prerecorded speech database. As compared to any other variables, the syllable-based output is natural and of high quality. Still there is the scope to improve the quality of speech.

References

1. Ananthi, S., Dhanalakshmi, P.: Syllable based concatenative synthesis for text to speech conversion. In: Computational Intelligence in Data Mining—Volume 3, Smart Innovation, Systems and Technologies 33. https://doi.org/10.1007/978-81-322-2202-6_6
2. Kawachale, S.P., Chitode, J.S.: Comparable analysis of neural and non-neural approach in syllable segmentation in Marathi TTS. Electrical Engineering and Control. LNEE, vol. 98, pp. 451–460
3. Schuller, B., Zhang, Z., Weninger, F., Burkhardt, F.: Synthesized speech for model training in cross-corpus recognition of human emotion. Int. J. Speech Technol. **15**, 313–323 (2012)
4. Campbell, N.: Conversational speech synthesis and the need for some laughter. IEEE Trans. Audio Speech Lang. Process. **17**(4), 1171–1179 (2006)
5. Campbell, N.: Developments in corpus-based speech synthesis: approaching natural conversational speech. IEICE Trans. **87**, 497–500 (2004)
6. Badin, P., Fant, G.: Notes on vocal tract computation. Techical Report, STL-QPSR (1984)
7. Banks, G.F., Hoaglin, L.W.: An experimental study of duration characteristics of voice during the expression of emotion. Speech Monogr. **8**, 85–90 (1941)

8. Atal, B.S., Hanauer, S.L.: Speech analysis and synthesis by linear prediction of the speech wave. J. Acoust. Soc. Am. **50**, 637–655 (1971)
9. Carlson, R., Sigvardson, T., Sjolander, A.: Data-driven formant synthesis. Technical Report, TMH-QPSR (2008)
10. Campell, N., Hamza, W., Hog, H., Tao, J.: Editorial special section on expressive speech synthesis. IEEE Trans. Audio Speech Lang. Process. **14**, 1097–1098 (2006)
11. Clark, R.A.J., Richmond, K., King, S.: Multisyn: opendomain unit selection for the festival speech synthesis system. Speech Commun. **49**, 317–330 (2007)
12. Courbon, J.L., Emerald, F.: A text to speech machine by synthesis from diphones. In: Proceeding of ICASSP. PTR, Upper Saddle River (2002)
13. Kim, J.K., Hahn, H.S., Bae, M.J.: On a speech multiple system implementation for speech synthesis. Wirel. Pers. Commun. **49**, 533–543 (2009)
14. Saraswathi, S., Vishalakshy, R.: Design of multilingual speech synthesis system. Intell. Inf. Manag. **2**, 58–64 (2010)
15. Ahmed, M., Nisar, S.: Text-to-speech synthesis using phoneme concatenation. Int. J. Sci. Eng. Technol. **3**(2), 193–197 (2014)

Inverted U-shape Monopole Antenna with Notch-Band Characteristics for UWB Communication Applications

Priyanka Gawhale, Bharat Patil, Archana Wasule and Vijay K. Sambhe

Abstract Here, the notch-band antenna is developed and designed for ultra-wideband (UWB) applications. Designed antenna is fabricated on FR4 substrate with permittivity 4.4. Furthermore, the designed antenna is comparable with two antennas given in the literature with similar dimension. Also, the important parameters affecting bandwidth are analysed with rigorous simulation. The results are comparable, and it is discussed with the other candidates given in the literature. The central arm is designed, and its dimension is optimized to get the particular band at 3.5 GHz. The designed antenna supports from 3 to 12.5 GHz, i.e., 9.5 GHz bandwidth. Further, the antenna is simple and it is compatible with the other electronic circuitry.

Keywords Impedance bandwidth · Radiating patch · Radiation pattern Ultra-wideband antenna

1 Introduction

Currently, UWB communication has been much focused by the researchers as 3.1–10.6 GHz frequency band are released [1–4]. Antennas with small and compact size with multi-band operation have become more popular due to integration of multiple communication system for a particular system. These antennas are portable and

P. Gawhale · B. Patil
Usha Mittal Institute of Technology, Mumbai, India
e-mail: gawhalepriyanka@gmail.com

B. Patil
e-mail: bharat.patil@umit.sndt.ac.in

A. Wasule
K. J. Somaiya Polytechnic, Mumbai, India
e-mail: ar.wasule@gmail.com

V. K. Sambhe (✉)
Veermata Jijabai Technological Institute, Mumbai, India
e-mail: vksambhe@vjti.org.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_61

adjust with the modern communication systems. To occupy more number of operation, basically two techniques are given in the literature. One is to create multiple patches for generating different harmonics and second to integrate multiple radiators in single patch to resonate at different frequencies.

To drive short-range wireless communication application and ultra-wideband communication particularly include (a) increasing demand for low-cost wireless handy devices, (b) accommodate assigned frequency bands and other available frequency band and (d) decreasing semiconductor cost and power consumption for communication applications.

Frequency interference are handled by some of the antennas and reported in [5–12], whereas extended rectangular slot with inverted L-patch for the band-notch characteristics is introduced in [5]. Further, circular ring with rectangular ground, for UWB and band up to 21 GHz is cited in [6].

Furthermore, symmetrical U-shape radiating structure with rectangular ground plane is introduced in [7]. In [8], thin slot on the radiating structure, such as L-shaped slot is etched. U-shaped on the radiating surface in [9–12] is introduced for the band stop characteristics. In [11], notch in the ground is introduced as a novel antenna. By adding split ring resonator, bandwidth is enhanced in [13]. In [14, 15], additional load is introduced for the rejection of unwanted bands. Open slot antenna with variable ground plane in [16], and Ohm slot in [17]. CPW-fed antenna with reconfigurable rejection bands is designed in [18]. MEMS switch and PIN diodes are also used for the rejection unwanted bands in [19, 20]. After all, the above designs are more complex for the generation of stop-band. Also, the fabrication cost is increased for practical operation due to complex structure. Therefore, simple and compact antenna structure is essential for current communication application. Here, simple and compact antenna structure is designed and described in the next section.

2 Antenna Geometry

Here, compact wideband slot antenna with notch band is designed. The antenna gives band-notch characteristics without re-tuning the radiating patch and also compatible to the other microwave devices. To design antenna structure, transmission line model is used for calculations [1]. For simulation, commercial HL3D software is used from Mentor Graphics.

Initially, rectangular patch of 12 (LR) × 12 (WR) mm^2 and approximately same ground plane is considered to cover ultra-wideband frequency rang. Antenna and ground plane are separated by 1.5 mm gap. The antenna is connected by 50 Ω and ended with SMA connector. The antenna is further optimized to cover UWB bands. The antenna structure we called here antenna 1 is designed and has effective control the bandwidth from 3.7 to 12.86 GHz. Thereafter, central portion of the antenna 1 is removed as the maximum current is at the outside edges and has negligible effect on the impedance bandwidth or radiation patterns. Now, the central arm of antenna 2 is divided into two parts with gap ‘g1’ = 0.2 mm, and the gap is optimized to

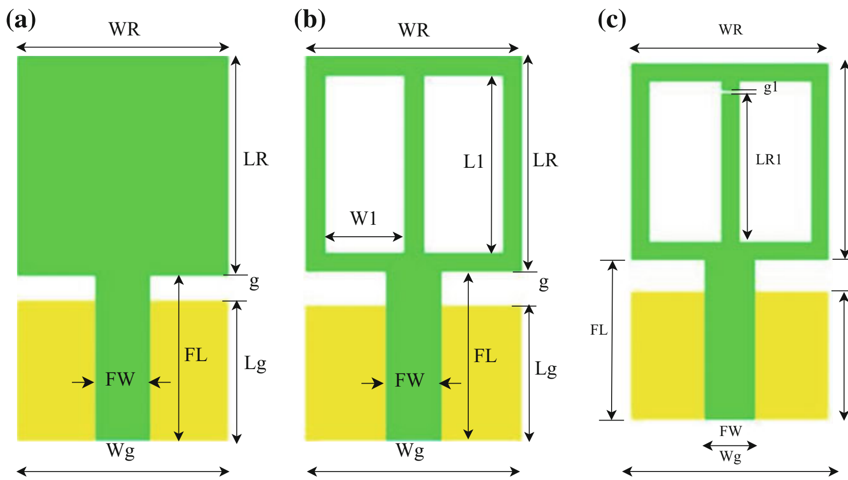


Fig. 1 Geometry of the antenna **a** Antenna 1 with LR = 12 mm, WR = 12 mm, Lg = 8 mm, Wg = 12 mm, and gap 'g' = 1.5 mm. **b** Antenna 2 with LR = 12 mm, WR = 12 mm, Lg = 8 mm, Wg = 12 mm, and gap 'g' = 1.5 mm. **c** Antenna 3 with LR = 12 mm, WR = 12 mm, Lg = 8 mm, Wg = 12 mm, LR1 = 10.2 mm, g1 = 0.2 mm, and gap 'g' = 1.5 mm

reject 3.5 GHz frequency bands. The geometry and detailed configuration antennas is shown in Fig. 1. Micro-strip line with 10.5 (FL) × 3 (FW) mm^2 is used for impedance matching. The centre arm of antenna 3 is divided into two parts with gap 'g1' = 0.2 mm. The length 'LR1' of the centre arm decides the central resonating frequency of the notch band. Therefore, centre strip is very crucial to design since the stop-band resonance and the impedance bandwidth is changed as the central strip changes. For comparison, return loss versus frequency plot of the three antennas is shown in Fig. 4a. The plot indicates antenna 3 covers UWB bands with notch band at 3.5 GHz.

3 Results and Discussion

The antenna surface current of the designed antenna is shown in Fig. 2. The central arm has maximum current density at 3.5 GHz, whereas less current is observed at higher frequencies in the central arm. The vertical current component is more in the central arm, and it is maximum at 3.5 GHz. However, the horizontal current component is less in the central arm. Further, antenna radiation pattern is shown in Fig. 3.

The radiation plot indicates, the cross-polarization occurs at higher frequencies, but it is ≤ -20 dB. Also, the pattern is not symmetrical for $\phi = 0°$ and $\phi = 90°$. The radiation pattern shows approximately omnidirectional pattern which is a unique feature of the proposed antenna. Also, the cross-polar component increases at higher

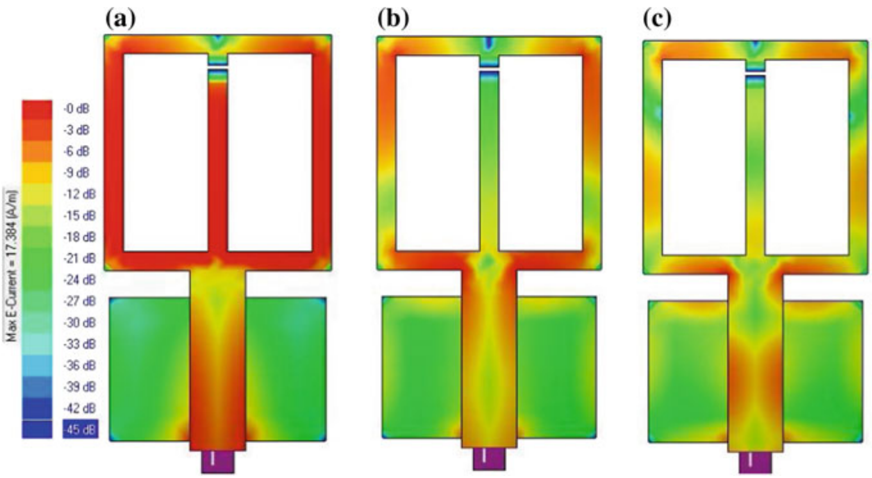


Fig. 2 Surface current distribution of the proposed antenna **a** 3.5 GHz. **b** 7.5 GHz. **c** 11.5 GHz

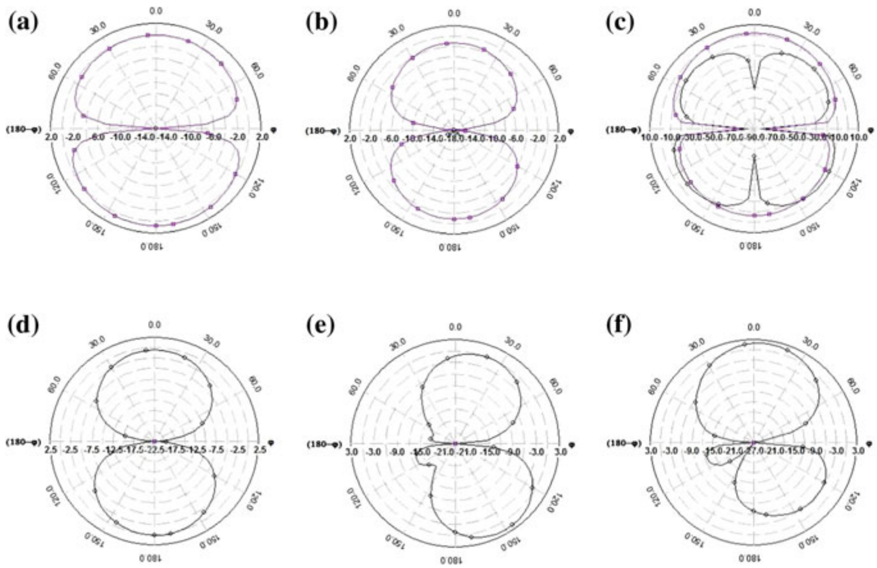


Fig. 3 The radiation pattern of the proposed antenna **a** 3.1 GHz $\phi = 0°$. **b** 7.0 GHz $\phi = 0°$. **c** 11.0 GHz $\phi = 0°$. **d** 3.1 GHz $\phi = 90°$. **e** 7.0 GHz $\phi = 90°$. **f** 11.0 GHz $\phi = 90°$

frequencies, but it is limited through overall impedance bandwidth. The measured and simulated return loss versus frequency plot of the antenna is shown in Fig. 4b. It indicates there is similarity between the measured and simulated results. The mismatch occurs due to fabrication and soldering inefficiency. In addition, the effect of different parameters which affects on impedance bandwidth is analysed. For clarity

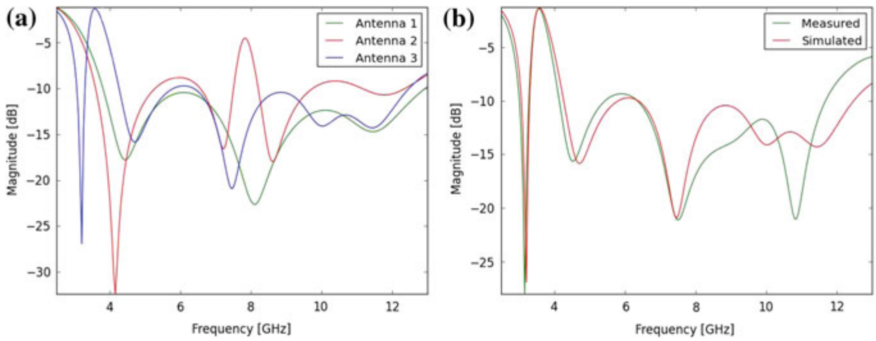


Fig. 4 **a** Return loss versus frequency plot of three antennas and **b** measured return loss versus frequency plot of the antennas

purpose, one parameter is varied at a time and other parameters kept constant. Initially, radiating patch with length 'LR' is increased from its optimum length 12.5–13 mm and then 13.5 mm. The effect of 'LR' on bandwidth is shown in Fig. 5a. It is seen that increased 'LR' shifts the first resonance at lower ends and there is less effect on the entire impedance bandwidth. Also, the entire impedance bandwidth remains constant.

Further, the central arm and its gap 'g1' is very crucial parameters as far as proposed notch band is concerned. Therefore, 'LR1' is varied from its optimum length 10.5–9.5 mm and then 8.5 mm and is shown in Fig. 5a. It is noted that decreased 'LR1' shifts first resonance towards higher frequency ends and also decreases the bandwidth, and it is due to the vertical current component. Next, gap 'g1' is very crucial as far as designed antenna is concerned. Gap is increased from 0.2 to 0.4 mm and then 0.6 mm, and results is shown in Fig. 5b. It shows the shift in first resonant frequency bands towards higher frequency ends and has less effect on the UWB frequency band, and the impedance bandwidth remains constant. Also, the gap 'g' is a very crucial parameter for the above-designed antenna. The gap 'g' for the designed antenna is 1.5 mm, and it is increased to 2 mm and then 2.5 mm. The result of different gap 'g' is shown in Fig. 5c. The plot shows the shifts in lower frequency band as well as UWB band towards lower frequency ends and has less effect on the total bandwidth.

Finally, ground plane parameter affects impedance bandwidths in monopole antenna. The optimum length of the ground plane is 8 mm. It is increased to 9 mm, and then 10 mm, and the result is shown in Fig. 5d. The plot indicates that bandwidth is shifted towards lower ends. Also, the effect of increased width of the ground plane is noted and has less effect on impedance bandwidth as shown in Fig. 5e. The effect of length 'Lg' and width 'Wg' and the patch length and width has similar effect on the impedance bandwidth as far as monopole antenna is concerned. At the end, the photograph of the proposed antenna with front and back view is shown Fig. 6.

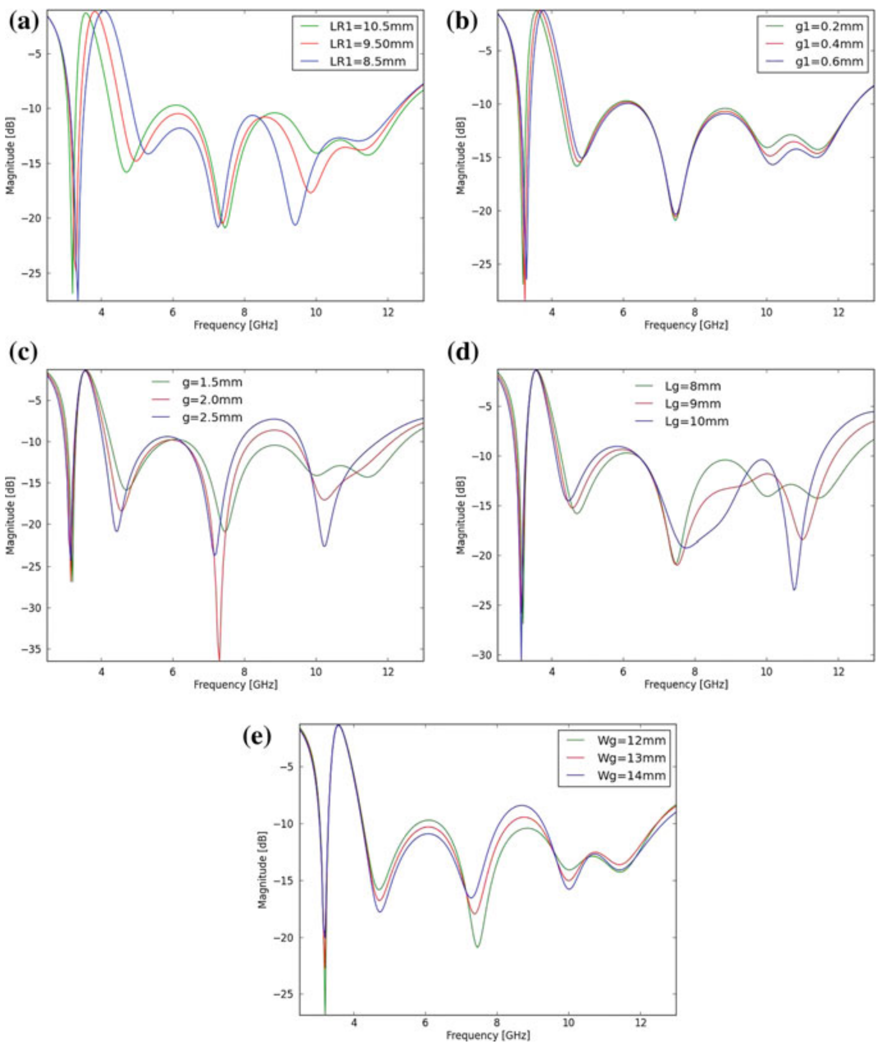


Fig. 5 Return loss versus frequency plot of the antenna for different **a** 'LR1', **b** 'g1', **c** 'g', **d** 'Lg', and **e** 'Wg'

4 Conclusion

Here, the antenna is designed for UWB applications with band-notch characteristics. The impedance bandwidth is less than −10 dB return loss for the entire frequency range from 3 to 12.5 GHz with stop band at 3.5 GHz. The different parameters which affect the impedance bandwidth are discussed with simulated results. This antenna

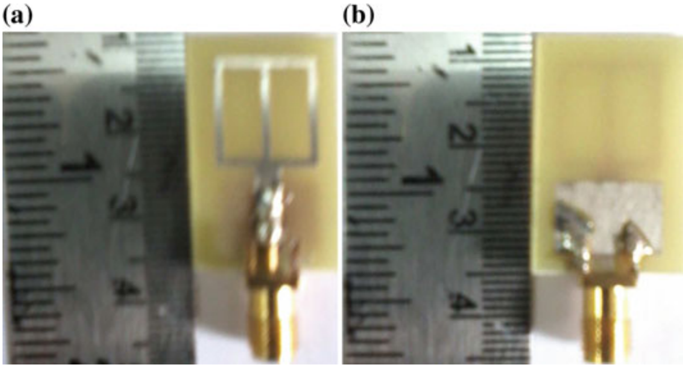


Fig. 6 Photograph of the proposed antenna **a** Front view. **b** Back view

has stable radiation patterns within the impedance bandwidth. Antenna have small in size and simple structure. Also, antenna provides omnidirectional radiation patterns for particular frequency band.

References

1. Chahine, S.A., Addam, M., Rahim, H.A., Itani, A., Jomaa, H.: A modified circular disc monopole ultra wide band antenna. In: 2009 International Conference on Advances in Computational Tools for Engineering Applications, pp. 71–75. July 2009
2. Dong, Y.D., Hong, W., Kuai, Z.Q., Chen, J.X.: Analysis of planar ultrawideband antennas with on-ground slot band-notched structures. IEEE Trans. Antennas Propag. **57**(7), 1886–1893 (2009). July
3. Shokri, M., Shirzad, H., Movagharnia, S., Virdee, B., Amiri, Z., Asiaban, S.: Planar monopole antenna with dual interference suppression functionality. IEEE Antennas Wirel. Propag. Lett. **12**, 1554–1557 (2013)
4. Ma, J., Lu, C., Peng, C., Jia, J.: Design of compact band-notched ultra-wideband antenna. In: 2010 IEEE International Conference on Ultra-Wideband, vol. 1, pp. 1–3. Sept 2010
5. Simeoni, M., Cicchetti, R., Yarovoy, A., Caratelli, D.: Plastic-based supershaped dielectric resonator antennas for wide-band applications. IEEE Trans. Antennas Propag. **59**(12), 4820–4825 (2011). Dec
6. Naqvi, I.H., Khaleghi, A., Elzein, G.: Performance enhancement of multiuser time reversal uwb communication system. In: 2007 4th International Symposium on Wireless Communication Systems, pp. 567–571. Oct 2007
7. Tsitouri, C.I., Panagiotou, S.C., Dimousios, T.D., Capsalis, C.N.: A circular switched parasitic array of log-periodic antennas with enhanced directivity and beam steering capability for ultra wideband communications applications. In: 2008 Loughborough Antennas and Propagation Conference, pp. 281–284. March 2008
8. Chawanonphithak, K., Phongcharoenpanich, C., Kosulvit, S., Krairiksh, M.: 5.8 ghz notched uwb bidirectional elliptical ring antenna excited by circular monopole with curved slot. In: 2007 Asia-Pacific Microwave Conference, pp. 1–4. Dec 2007
9. Golovin, V.V., Lukyanchikov, A.V.: The broadband compact antenna with the radiation field of circular polarization. In: 2008 4th International Conference on Ultrawideband and Ultrashort Impulse Signals, pp. 95–96. Sept 2008

10. Qiu, J., Lin, S., Zhong, L., Lu, G.: Study of planar ultrawide-band antennas. In: 2008 4th International Conference on Ultrawideband and Ultrashort Impulse Signals, pp. 60–62. Sept 2008
11. Russo, I., Freidl, P.F., Leitgeb, E.: Slot-coupled UWB circularly polarized patch antenna with matching reflector. In: Proceedings of the 12th International Conference on Telecommunications, pp. 179–184. June 2013
12. Lu, K.H., Lin, K.C., Lin, S.K., Lin, Y.C.: Wideband circularly-polarized aperture antenna arrays utilizing uwb directional coupler. In: Proceedings of the 2012 IEEE International Symposium on Antennas and Propagation, pp. 1–2. July 2012
13. Chung, C.C., Kamarudin, M.R.: Novel design of circular uwb antenna. In: 2009 Asia Pacific Microwave Conference, pp. 1977–1979 Dec 2009
14. Sharma, S., Savita, M.K.: Multi band circular antenna for ultrawideband wireless systems. In: 2012 Nirma University International Conference on Engineering (NUiCONE), pp. 1–3. Dec 2012
15. Su, S.W., Wong, K.L., Chang, F.S.: Compact printed band-notched ultra-wideband slot antenna. In: 2005 IEEE Antennas and Propagation Society International Symposium, vol. 2B, pp. 572–575. July 2005
16. Zhao, F., Tang, H., Zhao, C., Gao, X., Zhuo, P., Zhang, F.: Design of novel dual band-notched disk monopole antennas. In: The 2012 International Workshop on Microwave and Millimeter Wave Circuits and System Technology, pp. 1–4. April 2012
17. Jung, J.: Compact band-notched ultra-wideband antenna with parasitic elements. Electron. Lett. **44**(2), 1104–1106 (2008)
18. Sharma, M.M., Kumar, A., Ranga, Y., Bhatnagar, D.: An ultra-wideband antenna with axis symmetrical elliptical slots for tunable band-notched characteristics. In: 2011 Asia-Pacific Microwave Conference, pp. 725–728. Dec 2011
19. Ojaroudi, N., Ojaroudi, M., Ghadimi, N.: UWB omnidirectional square monopole antenna for use in circular cylindrical microwave imaging systems. IEEE Antennas Wirel. Propag. Lett. **11**, 1350–1353 (2012)
20. Bao, X.L., Ammann, M.J.: Wideband dual-frequency dual-polarized dipole-like antenna. IEEE Antennas Wirel. Propag. Lett. **10**, 831–834 (2011)

A Distance-Based Outlier Detection Using Particle Swarm Optimization Technique

Abdul Wahid and Annavarapu Chandra Sekhara Rao

Abstract Due to rapid evolution in technologies, data acquisition techniques generate a huge amount of data in real-world scenario and capture information about marketing, business, banking, government, science and technology, etc. Regardless of huge amount of data, some data point shows exceptional or abnormal behavior in a dataset which are called as outliers. Outlier detection plays a significant role in several fields like: fraud detection, medical diagnosis, intrusion detection, financial application. To detect outliers in a given dataset, firstly outlier detection algorithms build a normal pattern in a dataset and then allocate an outlying degree to each point of a dataset based on deviation from normal pattern. This paper presents a distance-based outlier detection algorithm using particle swarm optimization (PSO) technique. In this algorithm, it assigns an outlying degree to each data point using the sum of distances between points and its k-nearest neighbor set, and PSO is used to detect outlying subspaces where outliers may exist. Finally, experimental studies show that our proposed algorithm is more effective in terms of efficiency and accuracy by using UCI dataset.

Keywords Outlier detection · High-dimensional data · Distance-based
Particle swarm optimization · Outlying degree

A. Wahid (✉) · A. C. S. Rao
Department of Computer Science and Engineering,
Indian Institute of Technology (Indian School of Mines), Dhanbad,
Dhanbad 826004, Jharkhand, India
e-mail: abdul.cspg14@nitp.ac.in

A. C. S. Rao
e-mail: acrao232@yahoo.co.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_62

1 Introduction

An outlier is a pattern or data point that differs significantly, and also inconsistent with other data points in a given dataset [1], and may contain some useful or significant information about the data that cannot be easily discoverable. In data mining and statistics literature, outliers are also known as anomalies, defects, discordant, faults, abnormalities, deviants, and so on. Mostly in real-world application, data are produced by one or more different generating processes, which may reveal either some activity which is going on in the system or observations about the collected data. At the point when data generating process acts abnormally, it brings about the initiation of outliers.

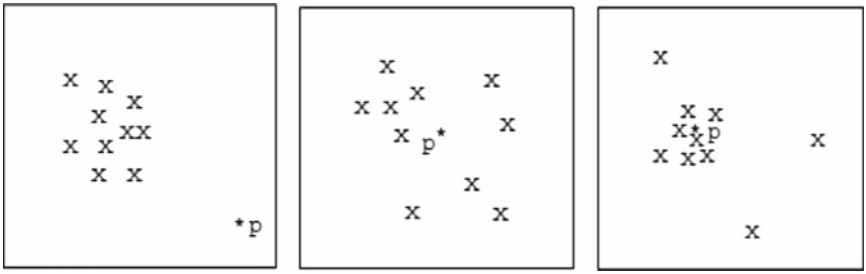
Outlier detection is a challenging task in data mining to detect or identify data points that differs very greatly from rest of the data in a dataset. The importance of outlier detection is because of the fact that outliers may additionally suggest a new trend within the technique that generates the data or sign fraudulent activity inside the data or alter data into significant information, which is used in wide variety of applications, like medical diagnosis, fraud detection, intrusion detection, marketing systems, astronomical spectroscopy.

Nowadays, many outlier detection algorithms have been proposed, for example, statistical-based, clustering-based, distance-based, and subspace-based techniques [2, 3]. In statistical-based methods, it assumes that data follow a standard distribution, and it detects outliers by identifying objects that deviate from the distribution. The drawback of this method is that we will not always have a priori understanding regarding the underlying distribution of the datasets, especially for high-dimensional datasets.

In clustering-based technique, it forms cluster model to describe underlying behavior of dataset and generally some data points in dataset considered as an outlier if those are far away from the center of clusters. For example, cluster histograms have been used for modeling and guiding outlier detection. However, clustering-based approaches must build a clustering model, which limits the outlier detection performance. In distance-based techniques, it finds out outliers through computing distances between an object and other objects in a dataset and subspace-based methods were designed to locate outliers by searching subspace.

Aiming for these issues and challenges, we have proposed an outlier mining algorithm based on distance to handle massive and high-dimensional dataset efficiently. In this method, it detects outliers by calculating the distance of point to its k-nearest neighbor set and then applies particle swarm optimization (PSO) technique to find out outlying subspaces where outliers may exist and then finally compares outlying degree of data point to a threshold value to detect outliers.

To understand what are outlying subspaces, we have taken an example which is shown in Fig. 1, where it shows two different dimensional views of high-dimensional data. From Fig. 1a, it is clear that point p has some different characteristics with respect to others in same data space that means it behaves like outliers in this 2-D view. But in Fig. 1b, c, it is clear that point p does not exhibit



1(a) 1(b) 1(c)

Fig. 1 Two different dimensional views of high-dimensional data

same characteristics as in Fig. 1a. So it is a very important task to find correct outlying subspaces or set of attribute where outlier may exist.

The main contributions of this paper are as follows:

- To compute the outlying degree of each data point in high-dimensional massive datasets, we used an outlying degree measurement technique which is built on the sum of the distances between point and its k-nearest neighbor set.
- One heuristic algorithm has been used to search outlying subspaces where outliers may exist.
- To detect outliers, we compared the value of outlying degree (OD) to a threshold value **Th,** which is calculated using Eq. 5.

The rest of the paper is structured as follows. Section 2 surveys the existing work related to outlier mining techniques. And in Sect. 3, an outlying degree measurement technique is presented, where it measures the OD of each data point in a subspace. Section 4 discusses the outlier detection technique using PSO technique. Section 5 presents experimental studies that describe the performance of the results on various datasets taken from UCI repository. And the conclusion is given in Sect. 6.

2 Related Works

In the present scenario, so many outlier detection algorithms have been proposed by researchers. With the techniques adopted, existing outlier detection algorithms can be categorized into following techniques: density-based, distance-based, distribution-based, and clustering-based outlier detection techniques [2, 3]. In density-based outlier detection technique, it detects an outlier if its local density differs from its neighborhood. Different density estimation methods can be applied to measure the density. As an example, local outlier factor (LOF) [4] score shows how outlierness score of an object differs from its neighborhood. Studies in [5] have demonstrated that it is more reliable to take those points who have the larger LOF scores as outliers, rather than comparing the LOF score with a threshold value. After that, so many variations of LOF algorithm have been projected. For example, in local distance-based outlier factor (LDOF) [6] algorithm, it discovers outliers in scattered datasets by using the relative distance of an object to its neighborhood, and in improving influenced outlierness (INFLO) [7] score, it considers both neighbors and reverse neighbors of an object while calculating its relative density distribution.

Müller et al. [8] projected one new algorithm, a connectivity-based outlier factor (COF), with considering underlying patterns of the data, but all the methods based on LOF and COF, discussed above, detect outlier using the relative distribution through distance. Numerous density-based approaches have been proposed based on kernel density estimation in [9–11].

Tang et al. [12] proposed a distance-based approach for outlier detection which is based on the well-known nearest neighbor algorithm; it detects outliers by calculating distance of an object with its neighbors; i.e., the greater is the distance of an object to its neighbors, the more likely it is an outlier. The distances among objects need to be calculated in either raw data space or feature subspace.

In distribution-based approach, it considers an object as an outlier if deviates too much from a standard distribution (e.g., Poisson distribution, normal distribution) but the drawback of the distribution-based method is that we do not know the underlying distribution of dataset; i.e., its underlying distribution is unknown, and data may not follow any standard distribution for many real-world problems. And in cluster-based approach [13–15], it considers an object as an outlier if it does not belong to any cluster. Clustering techniques in [16] detect outliers in large attributed graphs. However, clustering-based approaches must build a clustering model, which limits the outlier detection performance.

3 Outlying Degree Measurement

Before discussing the algorithm to detect outliers for high-dimensional data, first we discuss outlying degree (OD) measurement techniques.

3.1 Outlying Degree (OD) Measure

Let us describe how to measure outlying degree of a data point in a subspace $\boldsymbol{s}$. Let us assume that given a dataset $\boldsymbol{D}$ that includes $\boldsymbol{M}$ objects and $\boldsymbol{d}$ number of attributes, where $A = \{A_1, A_2, A_3 \ldots A_{d-2}, A_{d-1}, A_d\}$ be the set of attributes, and $P = \{P_1, P_2, P_3 \ldots P_{M-2}, P_{M-1}, P_M\}$ be the set of objects/points in a dataset $\boldsymbol{D}$. For the calculation of an outlying degree of each data point in a dataset, we define a degree to which the point diverges from the rest of the data point in the same subspace and that degree are called as outlying degree (OD) of a point in subspace s. OD is described as the sum of the distances between a data point and its k-nearest neighbor set in a data subspace [17]. Mathematically, the outlying degree of a point p in subspace s is calculated as:

$$\mathrm{ODs}(\mathrm{p}) = \sum_{i=1}^{k} Dist(p, pi) \tag{1}$$

where pi is k-nearest neighbor set of point p in subspace s, i.e., $p_i \in$ KNN Set(p, s), and Dist(p, p_i) is the Euclidean distance between two numeric points p and p_i. And distance between any two points can be calculated as:

$$Dist(p_x, p_y) = [\sum(((p_{xi} - p_{yi})/(Max_i - Min_i))^2]^{1/2} \tag{2}$$

In Eq. (2), $\mathrm{Max_i}$ and $\mathrm{Min_i}$ denote maximum and minimum values of data point in *i*th dimension (Fig. 2).

The algorithm for calculating an outlying degree of data point in a subspace, which is based on distance, is described in Algorithm 1. Firstly, it sorts the value of each attribute in ascending order and then stores in another dataset Di. After sorting, it calculates the k-nearest neighbor of each data point from Di and creates a D[i][j] matrix where i refers to rows and j refers to column of a dataset, which is given at line 6. And then, it calculates the outlying degree of each point as sum of the distance between data point and its k-nearest neighbor set. After getting the outlying degree of each point, it stores in OD_Mat, which is M x d matrix.

Algorithm 1: Computing Outlying Degree (OD)
Input:
Original dataset ***D,*** dimension number ***d***;
Output:
Outlying degree matrix ***OD_Mat***;
1. *M*=|*D*|; // Number of objects in a dataset D.
2. K=√*M*; // Value of k.
3. **for** (i=1;i≤d;i++) //Loop from i=1 to d(i.e no of dimensions)
4. $D_i \leftarrow$ Sort(D); //sort the value of *i*-th attribute of dataset ***D***, based on ascending order and then store in $D_{i.}$
5. **for** (j=1;j≤M;j++)
6. $p_i^j \leftarrow$ Compute (D[i][j]); //KNN set of each 1-D point and D[i][j] is a matrix that formed after sorting the value of D. where p_i^j of each 1-D point is : $p_i^j(p_{ij}) = \{nn^j_k(p_{ij}) \cup p_{ij}\}$.
7. $OD(p_{ij}) = \sum_{i=1}^{k} Dist(p, pi)|$ pi ∈ KNN Set (p_{ij}).// calculate outlying degree of each 1-D point.
8. OD_Mat ← OD(p_{ij}); // store outlying degree of each 1-D point in OD_Mat
9. **end for**
10. **end for.**

Fig. 2 Algorithm for outlying degree calculation

4 Outlier Detection Using PSO

PSO is an evolutionary algorithm, inspired from bird flocking [18, 19]. This evolutionary becomes popular because PSO is simple and effective for solving wide range of searching problems and maintains a swarm of individuals as particle, where each particle represents a solution. A particle has position and velocity which is updated at each iteration like this:

$$position\,[i+1] = position\,[i] \,+\, v[i+1] \tag{3}$$

where velocity component $v[i + 1]$ represents step size.

The velocity is calculated as:

$$v[i+i] = wt * v[i] + a_1 * rand() * (P_{best} - position[i]) + a_2 * rand() * (G_{\mathrm{best}} - position[i]) \tag{4}$$

where *wt* is the inertia weight, a_1 and a_2 are acceleration coefficients, and *rand()* is random number in the range of 0 and 1. P_{best} is the personal best, and G_{best} is the global best position of the particles. In the beginning, each particle has a random position and velocity. Particle evaluates the fitness function and updates its velocity in such a manner that the particle drawn toward P_{best} and $G_{best.}$

The algorithm for outlier detection using PSO is illustrated in Algorithm 2, where it takes two inputs—first is OD-Mat generated by Algorithm 1, and second one is threshold value **Th**, which is calculated as:

$$\mathrm{Th} = \alpha * \sqrt{\sum (\overline{ODei})^2} \tag{5}$$

where dim(ei) = 1 and ei $\subseteq s$ and

$$\overline{ODei} = 1/M * \sum_{j=1}^{M} ODei(ps(j)) \tag{6}$$

where $ps(j)$ is jth sampling point in a subspace s, M is the number of objects, and α is a constant factor that should be greater than 1 (i.e., $\alpha > 1$).

The particle in PSO is initialized by *initialize*() function that includes position of the particle, velocity of the particle, fitness function: given in Eq. 1 of this paper, individual best position P_{best}, and globally best position G_{best}. And then, PSO searches some better particles that can be considered as outliers (Fig. 3).

```
Algorithm 2: Outlier Detection using PSO:
Input:
   Outlying Degree Matrix OD_Mat, Threshold value Th;
Output:
    Outlier Set OD;
  1.  for (i=1;i<g; i++) do // g is experiment generation number.
  2.  Initialize (); // initialize the particle swarm, particle P position is position, speed is
      speed, Adaptive fitness value, P_best, and G_best.
  3.     for (j=1;j<n;j++)  do // n is the experiment number.
  4.          If   (G_best ≤ Th )// distance threshold
  5.               Break;
  6.         else
  7.               for all (p:P) do
  8.                The position and speed of the particle p are calculated as:
      Position[i+1]=position[i] + v[i+1]
      S[i+1]=wt * v[i] +a1*rand() * (P_best –position[i]) +a2*rand() * (G_best-position[i]).
  9.                    Calculatation();// calculating fitness value from Equation 1.
  10.                  LocalBest(); // optimal value for local search.
  11.                 end for
  12.               GlobalBest()// global optimal value.
  13.          end if
  14.    end for
  15.    if  (G_best ≤ Th )
  16.    Output OD;
  17.   End if
  18. End for
```

Fig. 3 Algorithm for outlier detection using PSO

5 Experimental Studies

The aim of this experimental study is to find out the accuracy and effectiveness of the proposed algorithm for outlier mining using PSO technique which is presented in this paper.

5.1 Experimental Setup

To demonstrate the accuracy and effectiveness of the proposed algorithm, we use Python programming to implement both the proposed algorithm and HiCS [20]. We have conducted these experiments on a system that contains Intel(R) Core(TM) i3-2350 M CPU, and 4 GB RAM, with Windows 10 operating system. To evaluate the accuracy and effectiveness of the algorithm, we processed on datasets taken from UCI repository.

5.2 UCI Dataset

We used four real datasets from UCI machine learning repository for evaluation. The dataset was chosen in such a way that results in considerable variability in terms of number of objects and the number of attributes. These four datasets are: Yeast, Ionosphere, Multiple Features (mfeat), and Isolet. Table summarizes the characteristics of these datasets.

In our experiment, PSO parameters chosen after several experiments are as: particle size 50, acceleration coefficients a1 and a2 both are 0.5, inertia weight (wt) is set to 0.8, test number is 10, and the number of iteration is set to 1000.

5.3 Comparison with HiCS Algorithm

This section discusses the effectiveness of our proposed algorithm with well-known HiCS outlier detection algorithm. The evaluation is based on four real-life datasets from UCI machine learning repository (for reference, see Table 1).

5.3.1 Accuracy Analysis

We performed analysis on a given UCI dataset using our proposed algorithm, where threshold value Th is chosen according to Eq. 5 and α is set to 3. Figure 4 shows that our proposed algorithm's accuracy is higher than HiCS.

5.3.2 Efficiency Analysis

From Fig. 5a, when a number of data points are same, the time taken by our proposed algorithm with respect to HiCS is less while we increase the number of

Table 1 Description of UCI dataset

UCI datasets	Number of dimensions	Number of data points	Outlier points and its number	Normal points and its number	Number of outliers
Yeast	8	1105	ERL 5	NUC 100, MIT 200, CYT 400, and ME3 400	05
Ionosphere	34	205	Bad 5	Good 200	5
mfeat	649	1210	8–9, 5 from each category	0–7, 150 of each category	10
Isolet	617	4048	21–26, 8 of each category	1–20, 200 of each category	48

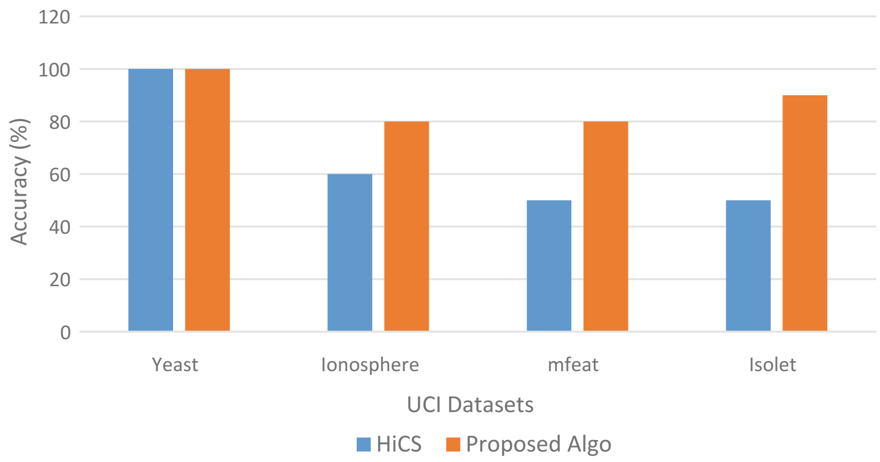


Fig. 4 Proposed algorithm with HiCS on accuracy

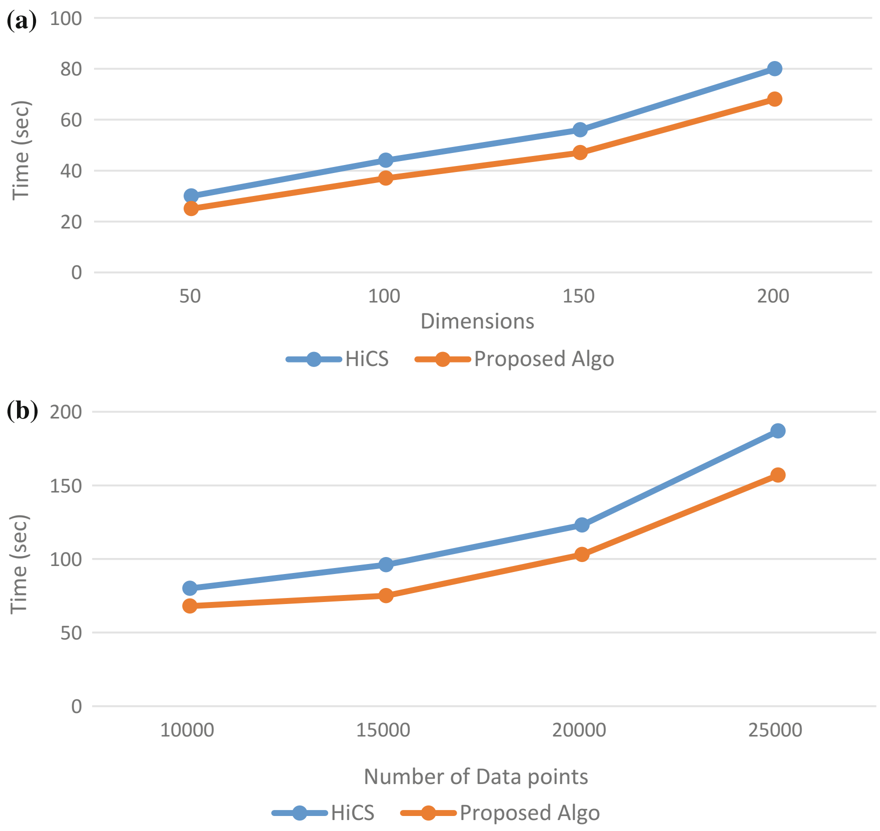


Fig. 5 **a** Efficiency on attribute dimension (number of data points = 10,000). **b** Efficiency on different datasets (number of dimensions = 200)

dimensions. The tilt degree of our proposed algorithm is also lower than the tilt degree of HiCS algorithm.

Similarly in Fig. 5b, the impact of dimensions is still high and the time complexity of those two algorithms is also linear with dimensions. The tilt degree of HiCS algorithm is greater than the proposed algorithm, and its efficiency is also low.

6 Conclusions

This paper presents an algorithm for outlier detection for massive and high-dimensional dataset using particle swarm optimization (PSO) technique. In this proposed algorithm, the outlying degree calculation is based on the distance between a data point and its k-nearest neighbor set and then we apply PSO algorithm on an outlying degree to detect outliers by comparing with the threshold value given by Eq. 5.

Further to validate the effectiveness of proposed algorithm, we performed experiments on four different datasets taken from UCI repository and experimental study shows proposed algorithm has better performance in terms of accuracy and efficiency as compared to HiCS algorithm.

References

1. Knox, E.M., Ng, R.T.: Algorithms for mining distancebased outliers in large datasets. In: Proceedings of the International Conference on Very Large Data Bases, pp. 392–403, Aug 1998
2. Baklouti, R., Mansouri, M., Nounou, M., Nounou, H., Hamida, A.B.: Iterated robust kernel fuzzy principal component analysis and application to fault detection. J. Comput. Sci. **15**, 34–49 (2016)
3. Fisher, W.D., Camp, T.K., Krzhizhanovskaya, V.V.: Anomaly detection in earth dam and levee passive seismic data using support vector machines and automatic feature selection. J. Comput. Sci. **20**, 143–153 (2017)
4. Hubert, M., Rousseeuw, P.J., Segaert, P.: Multivariate functional outlier detection. Stat. Methods Appl. **24**(2), 177–202 (2015)
5. Aggarwal, C.C.: A segment-based framework for modeling and mining data streams. Knowl. Inf. Syst. **30**(1), 1–29 (2012)
6. Perozzi, B., Akoglu, L., Iglesias Sánchez, P., Müller, E.: Focused clustering and outlier detection in large attributed graphs. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining, pp. 1346–1355, Aug 2014
7. Kriegel, H.P., Kroger, P., Schubert, E., Zimek, A.: Outlier detection in arbitrarily oriented subspaces. In: 2012 IEEE 12th International Conference on Data Mining (ICDM), pp. 379–388, Dec 2012
8. Müller, E., Schiffer, M., Seidl, T.: Statistical selection of relevant subspace projections for outlier ranking. In: 2011 IEEE 27th International Conference on Data Engineering (ICDE), pp. 434–445, April 2011

9. Jin, W., Tung, A.K. Han, J.: Mining top-n local outliers in large databases. In: Proceedings of the seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 293–298), Aug 2001
10. Hautamaki, V., Karkkainen, I., Franti, P.: Outlier detection using k-nearest neighbour graph. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR, vol. 3, pp. 430–433, Aug 2004
11. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: ACM Sigmod Record. vol. 29, pp. 93–104, May 2000
12. Tang, J., Chen, Z., Fu, A., Cheung, D.:. Enhancing effectiveness of outlier detections for low density patterns. In: Advances in Knowledge Discovery and Data Mining, pp. 535–548 (2002)
13. Zhang, K., Hutter, M., Jin, H.: A new local distance-based outlier detection approach for scattered real-world data. In: Advances in Knowledge Discovery and Data Mining, pp. 813–822 (2009)
14. Jin, W., Tung, A.K., Han, J., Wang, W.: Ranking outliers using symmetric neighborhood relationship. In: PAKDD, vol. 6, pp. 577–593, April 2006
15. Tang, J., Chen, Z., Fu, A.W.C., Cheung, D.: A robust outlier detection scheme for large data sets. In: 6th Pacific-Asia Conferences on Knowledge Discovery and Data Mining (2001)
16. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. KDD **96**(34), 26–231 (1996)
17. Aggarwal, C.C., Yu, P.S.: Outlier detection for high dimensional data. In: ACM Sigmod Record, vol. 30, pp. 37–46. ACM (2001)
18. Hu, W., Yen, G.G.: Adaptive multi objective particle swarm optimization based on parallel cell coordinate system. IEEE Trans. Evol. Comput. **19**(1), 1–18 (2015)
19. Kennedy, J.: Particle swarm optimization. In: Encyclopedia of Machine Learning, pp. 760–766 (2011)
20. Keller, F., Muller, E., Bohm, K.: HiCS: high contrast subspaces for density-based outlier ranking. In: 2012 IEEE 28th International Conference on Data Engineering (ICDE), pp. 1037–1048, April 2012

Fast and Memory Leak-Free Verification for Bridge Transitions in VLSI Designs

Gaurav Sharma and Lava Bhargava

Abstract The bridge devices are highly useful in joining two separate functionalities to establish an information link between them. Traditional verification approaches are unable to verify both bridges DUTs and bridge interface transition at the same time. The work proposes a Universal Verification Methodology (UVM) framework for bridge devices along with the conjunction of advanced verification environment to verify bridge devices. This proposed testbench architecture is very consistent in verification of all bridge devices. The testbench performs significantly better as compared to SystemVerilog testbench. The case study in this paper is the most used bridge for data transmission: Advanced High-Performance Bus (AHB) to Advanced eXtensible Interface (AXI4) Bridge v3.0 to validate the results on proposed architecture. The testbench architecture elaborates the memory leak, memory allocation, and functional coverage to verify the bridge functionalities. The assertion-based verification (ABV) utilizes the self-checking capabilities of testbench environment.

Keywords UVM · UVC · Testbench · Sequencer · Driver

1 Introduction

Verification is an important stage in product development. Verification expends the majority of time in the aggregate product development cycle. Innovation requests a quick and vigorous verification method to satisfy the hole between item request and supply. Innovative prerequisites drive us to assemble bigger, capable, and highly error-free designs. Complex designs are exceedingly mistakes inclined. Traditional

G. Sharma (✉) · L. Bhargava
Department of Electronics and Communication, Malaviya National Institute of Technology, Jaipur 302017, Rajasthan, India
e-mail: 2015rec9014@mnit.ac.in

L. Bhargava
e-mail: lavab@mnit.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_63

test and verification methodologies do not fit with them. UVM is perfectly suitable for verification of such complex and big VLSI designs [1]. UVM methodology carries an object-oriented base class library with the provision of transaction-level modeling (TLM). Parallel verification in UVM makes the structure faster and efficient as compared to other traditional methodologies. The low-speed AHB to high-speed AXI transactions is the case study in the work to propagate the testbench to fast and memory leak-free objectivity. The environment is universally reusable for all bridge protocols. Assertions are attentively adhering in profile to enact the necessary conditions of design [2]. To improve the pace, two different interfaces are present in architecture. The interfaces work parallel to avoid queues and buffers. Synchronization plays an important role in parallel transitions for both interfaces. Different clocks with same frequency are assigned to improvise the parallel working interfaces. The bridge is a connection between two designs. Both the bridge designs are in parallel, so that input of one is connected to the output of other [3]. The case study depicts that the AHB is passing output to AXI and output of AXI is driving the AHB input.

2 Objectives of Work

The objectives of the proposed UVM environment are as follows:

(1) Implementation of covergroups input signals on corresponding interfaces for bridge connection including verification of complete bridge design.
(2) Multiple test case run to ensure the working of bridge device.
(3) Joining the environment with an assertion-based verification (ABV) topology to make the debugging easy, if any necessary condition does not meet. Self-checking mechanism makes a smooth testbench flow.
(4) Incorporation performance metrics to ensure the functional and code coverage of bridge devices. These matrices identify the uncovered input/ output attributes of DUV.
(5) Designing random input transactions in proper flow to extract the design bugs in faster and structural manner.
(6) Incorporation memory leakage encounter algorithm for fast executing of testbench hierarchy.

3 Testbench Framework

3.1 UVM Environment

UVM environment is made reusable to facilitate its further usage when DUV is replaced also called Universal Verification Component (UVC) [4, 5]. An UVM environment is a logical group of the agent, sequencer, driver, monitor, scoreboard,

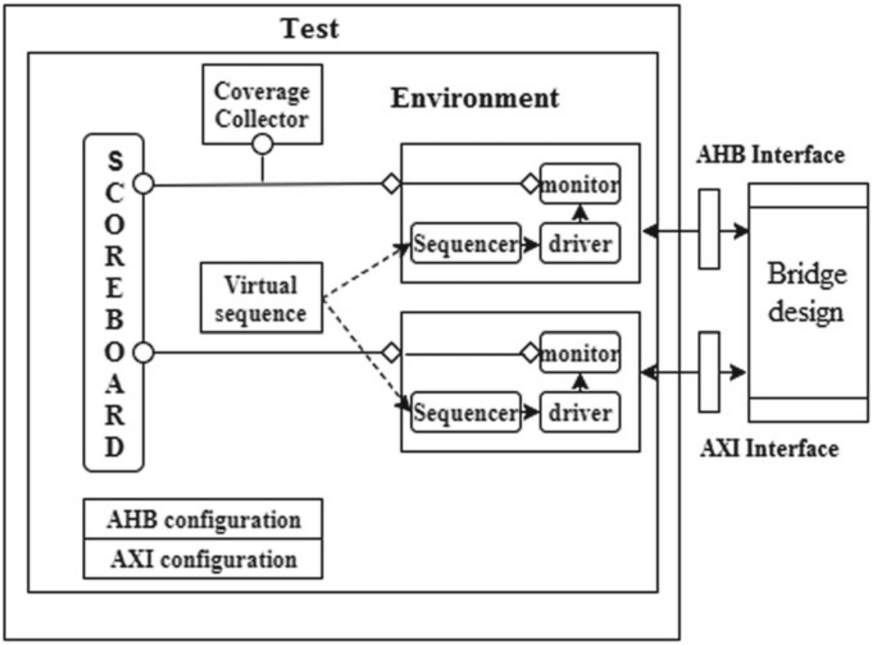


Fig. 1 Proposed UVM testbench framework

coverage collector, etc. as shown in Fig. 1. All these components extend from UVM base classes to take the parent class functionality.

3.2 Protocol Universal Verification Components

Following are the Universal Verification Components for the case study of AHB-Lite to AXI bridge:

Agent: Agent is a part of UVC to drive the sequences to the DUT from the interface.

Sequencer: AHB sequencer generates the sequences to pass it to the driver for driving the DUT. These sequences properly propagate to test the desired functions and trades of bridge design [6]. A bridge design needs two different sequencers separately for the continuous working of testbench at the required pace.

Driver: DUT accepts the pin-level data by default. So driver does the task to perform this conversion. The driver converts the sequence data to pin-level for DUT. It sends this data to interface to further pass it to DUT [7].

Monitor: Monitor recollects the pin-level data to convert it back to sequence data. The monitor is responsible for coverage and scoreboarding. The coverage is collected as functional and code coverage to ensure the desired functions of the bridge.

Interface: Interface is core part attached to the top of a testbench. All common mechanisms like clock and default settings are placed here to govern the data and response.

Coverage collector: Coverage collector is connected from analysis port of monitor to extract the functional and code coverage [8]. The monitor connectivity is necessary because all the input covergroups are defined here to authenticate the results with rational and logical stipulations.

Performance metrics: Performance metrics decides the acceptance criterion for a testbench. The functional and code coverage metrics manipulates the acceptable working of a testbench environment. The performance metrics are captured after simulation. The criterion is to decide whether the design possesses the validation functionalities or not. It controls and observes the testbench on functional and performance constraints.

Scoreboard: The scoreboard is a cross-checking class in UVM environment. The write/read transitions are ensured and authenticated using this class in UVM testbench environment. It provides an authenticity to testbench. It verifies that the read data and write data are same and assure that there is no interference in between them.

Virtual sequence: The virtual sequences are the random sequences. Transaction-level modeling (TLM) produces these random stimuli. The random sequences hit the DUT inputs for every possible test case. The virtual sequences vary according to test cases. These sequences decide the success or failure of a test case. The virtual sequences pass to separate sequencers to proceed for next transition.

4 Design Under Test

The AHB to AXI bridge is the case study in the proposed verification work. AHB slave and AXI master perceptually response with each other. The bridge is a connection between two designs. The motive is to verify those models which are connected to the bridge. The protocol is followed to keep the proper verification of bridge data. Both the designs are connected in such a way that output of one design is input to other design via bridge as shown in Fig. 2.

The testbench interface provides clock and reset signal in separate manner to attain synchronization. s_ahb is AHB slave AHB signals, similarly m_axi is AXI signals. s_ahb_haddr, s_ahb_hwdata, s_ahb_hrdata and m_axi_awaddr, m_axi_wdata, m_axi_r are address, write data, and read data signals for AHB and AXI, respectively. The write and read must be performed on same address.

The three write exchange AXI write channels are write data channel, write address channel, and write response channel. The two read exchange AXI read channels are

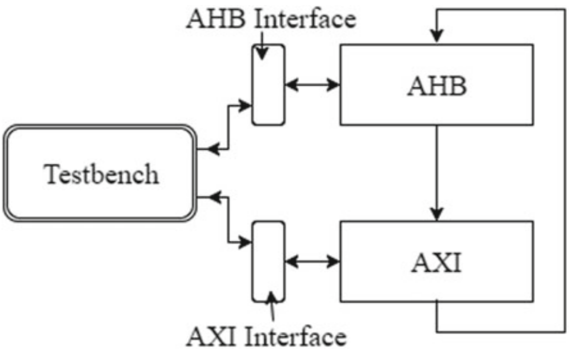


Fig. 2 Block diagram of proposed testbench

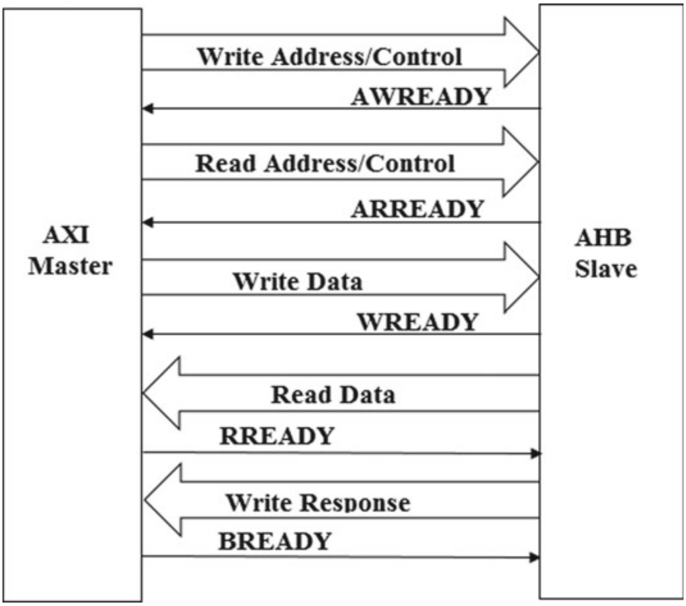


Fig. 3 AHB to AXI bridge channels

read data channel and read address channel shown in Fig. 3. The bridge controller guides the selection of these channels. The control channel activates write/read through write/address and read/address controller signals shown in Fig. 3. Parallel enactment of the control channels reduces the time elapsed between write and read. The condition when AXI does not respond to AHB input signals activates timeout signaling flag [3, 9].

5 Results and Discussions

Referring to Table 1, the proposed testbench framework justifies a 100% functional coverage. The testbench environment creates the random stimulus for AHB slave address (s_ahb_haddr) and write data (s_ahb_hwdata) signals. The coverage stipulations clear the success of testbench for sequences, environment packages, and all necessary coverage parameters. Seventy-four covergroups and 31 assertions are successfully implemented in the environment to take care of functional coverage and necessary design conditions.

The controller constitutes of control signals to decide the write or read operation. The control signal (hwrite) value 1 instructs to write and value 0 instructs to read the data generated by random stimulus input supply. The signals for write transmission on single burst are shown in Fig. 4. Referring to Fig. 4, AHB write data (wdata) corresponds to address (haddr) for write transmission. Similarly, Fig. 5 illustrates the transfer of read data from AHB (hrdata) to AXI (rdata) on same address (ahb_haddr). Figure 6 describes the increment of address by four write data transmission on corresponding particular addresses. Increase 4 write data transfer logically increments the next address by 4 to instantiate the next data transfer. In Fig. 5, ahb_hwdata appears at axi_wdata along with an address increment of 4 every time. The signal waveform justifies the transfer of AHB data to AXI for bridge transition verification.

Memory leakage is detected by recursive profiling of memory to trace the allocated memory. If allocated memory is decreasing, it identifies the deficit structure of memory. The problem is encountered by taking a trace and track of memory

Table 1 Performance metrics of proposed testbench (%)

Design scope	Coverage (%)
AXI_agent_package	100
AXI_coverage_Monitor	100
AXI_sequence	100
AXI_base_sequence	100
AXI_read_sequence	100
AHB_agent_package	100
AHB_coverage_Monitor	100
AHB_sequence	100
AHB_idle_mseq	100
AHB_wrapx_mseq	100
AHB_incr4_seq	100
Covergroups (74)	100
Assertion Success(31)	100
Sequence package	100

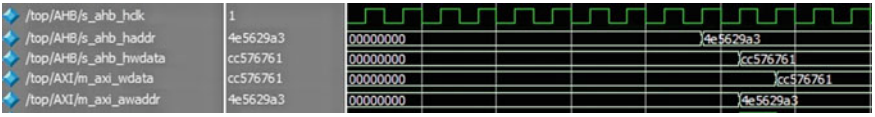


Fig. 4 Waveform for write transfer in single burst termination

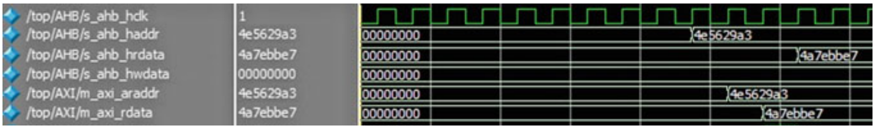


Fig. 5 Waveform for read transfer in single burst termination

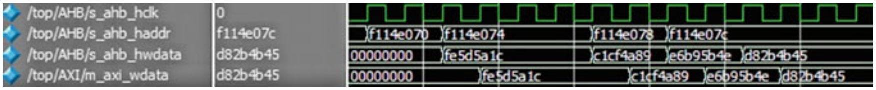


Fig. 6 Waveform for write transfer in single burst termination at address increment of 4

Table 2 Testbench refinement

Testbench	C.P.U time (s)	Verification time (s)
System verilog	78.2	45644.52
UVM	56.9	34845.29

allocated and memory in use. The memory in use may vary with the concerning type the of simulation and random stimulus, but allocated memory is fixed.

Table 2 judges the testbench refinement of proposed framework as compared to SystemVerilog. The process time is improved by 27.23% including write/read latency. The acting simulation refines by 23.65% with scoreboarding and coverage collection mechanism.

The curve in Fig. 7 intensely shows the percentage refinement in verification time with runs. The curve does not saturate because of randomization. The random TLM sequence flow in testbench makes things unpredictable with the surety of verification aspects including all test criterion. The comparative study of UVM and SystemVerilog makes UVM a unanimous winner. The probability of accurate verification is high in UVM proposed framework.

In the case study of bridge devices, it is 23,553 iterations to achieve maximum refined results of 28.98% with expectable coverage as shown in Fig. 7. After the curve's peak in the graph, the profiling depicts small variations in simulation verification timings. These fluctuations are unpredictable due to randomized transactions from transaction-level modeling (TLM). Applications are also responsible for opting cache as a memory for information storage. Increasing cache memory exposes the design and programming errors. The proposed results are memory leak-free. The allocated memory is constant throughout the experiment as 25 bytes.

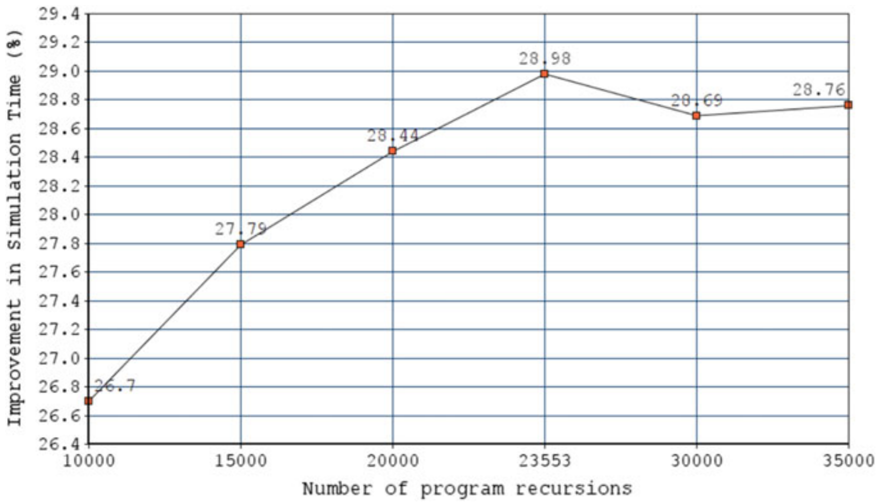


Fig. 7 Percentage refinement in simulation verification time with respect to program recursions

6 Conclusion

The work contributes a fast and memory leak-free UVM testbench architecture. The coverage gain of performance metrics shows 100% in results. The testbench is implemented with the highly uses bridge protocol design AHB to AXI DUT to authorize the test outputs. The framework correctly verifies the write/read on same addresses along with increment four address. Simulation verification time shows a drastic refinement compared to SystemVerilog. Memory leakage detection strategy is associated with UVM framework to identify design memory bugs. The memory allocation deterministic method shows a constant allocated memory to nullify the argument of memory leakage.

Acknowledgements This work is supported by annual contingency grant from Ministry of Electronics and Information Technology (Meity), Government of India.

References

1. Salah, K.: UVM-Based Smart Functional Verification Platform: Concepts, Pros, Cons, and Opportunities IDTS (2014)
2. Madan, R., Kumar, N., Deb, S.: Pragmatic approaches to implement self-checking mechanism in UVM based testbench. In: 2015 International Conference on Advances in Computer Engineering and Applications (ICACEA), pp. 632–636, 19–20 March 2015
3. XILINX—All Programmable. www.xilinx.com/support/documentation/ip_documentation/ahblite_axi_bridge/v3_0/pg176-ahblite-axi-bridge.pdf

4. Verification Academy Advance UVM understanding by Tom Fitzpatrick mentor graphics verification engineer. https://verificationacademy.com
5. UVM 1.2 (Universal Verification Methodology). [Online]. Available http://www.accellera.org/downloads/standards/uvm
6. Moretti, G.: Accellera's support for ESL verification and stimulus reused. In: IEEE Design & Test (2016)
7. Juan Francesconi, J., Agustin, R., Pedro M.J.: UVM based testbench architecture for unit verification. In: 2014 Argentine School of Micro-Nanoelectronics, Technology and Applications
8. Viswanathan, V., Runhaar, J., Reed, D., Zhao, J., Zhao, J.: Tough bugs vs smart tools—L2/L3 cache verification using system verilog, UVM and verdi transaction debugging. In: 2016 17th International Workshop on Microprocessor and SOC Test and Verification
9. Gayathri, M., Sebastian, R., Mary, S.R., Thomas, A.: A SV-UVM framework for verification of SGMII IP core with reusable AXI to WB bridge UVC. In: 3rd IEEE International Conference on Advanced Computing and Communication Systems (ICACCS) (2016)

Privacy-Preserving K-NN Classification Using Vector Operations

Hanumantharao Jalla and P. N. Girija

Abstract This paper presents a privacy-preserving K-nearest neighbor (PPKNN) classification algorithm in privacy-preserving data mining (PPDM) domain to preserve privacy of customers in business organization. This paper is about modification of K-nearest neighbor (K-NN) classification algorithm using vector operations. It modifies each cell of the original data record by dividing into three sub-components as a single unit row vector. Similarly, the test data record is converted into cells of column unit vectors. Finally, the dot product is applied between row and column vectors to preserve the distance between the data records as original data records. In modified dataset when distances between the records are preserved, then PPKNN works similar to K-NN algorithm. In this work, PPKNN is applied on real datasets referred from UCI machine learning repository and compares classification accuracies with K-NN algorithm.

Keywords K-NN · Vector operations · Privacy-preserving data mining

1 Introduction

A recent development in data mining raises an issue of privacy of individual customers in the business enterprise while sharing their original data with third party or other clients for data analysis. The sensitive data of customer's data is

H. Jalla (✉)
CBIT, Hyderabad, TS, India
e-mail: hanucs2000@gmail.com

P. N. Girija
SCIS, UOH, Hyderabad, TS, India
e-mail: pn_girija@yahoo.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_64

exposed to real world which includes name, address, and living status. If it happens, the reputation of business organization gradually decreases since the original data of individual customer's data revealed to the society. But, in some situations the business enterprises feel mandatory to share their customer's data with real world to run their business smoothly in competitive world or by obeying some government policies. Preserving such data leads to a new domain called privacy-preserving data mining (PPDM). Till now, many researchers addressed the issues related to privacy by applying perturbation, anonymity, and cryptography techniques. In perturbation techniques, a random value with mean value zero and variance value one is chosen in the distribution added to the original data and then modified data is shared with real world [1]. Perturbation techniques produce same information as original data, but it has some drawbacks when attacker applies statistical-based analysis on modified data. If he/she succeeds to identify random value, then the entire original data can be reconstructed. The K-anonymity technique is used to replace a group of K data values which has some common properties by a data value including a special character like *. The cryptography techniques are used in distributed environment. Secure multi-party computation (SMC) is mainly used in PPDM as cryptography technique [2–4]. It requires lot of computations to provide privacy to customers.

In this work, data modification and data analysis are combined into single method known as PPKNN. When all PPDM methods are applied to original data, then the resultant modified data is used for data analysis. In addition to the above PPDM methods, it modifies the original K-NN classifier using vector operations. The working principle of K-NN and PPKNN is same, but PPKNN contains data modification technique in addition to data analysis as K-NN to preserve the privacy of customers. The novelty of proposed method transforms low-dimensional data to high-dimensional data like nonlinear SVM classifier and also preserves the distances between records to extract knowledge like original data.

The rest of the paper is organized as follows: Related work is discussed in Sect. 2. The proposed method PPKNN classifier and experimental work are discussed in Sects. 3 and 4, respectively. Finally, concludes the work in Sect. 5.

2 Related Work

In the field of PPDM, lot of research is addressed to pasteurize data through linear transformation techniques. The proposed work is based on transformation technique to preserve the distance between the data records after modification. Few of the researchers used transformation technique as modification technique in the field of PPDM such as Fourier transforms, wavelet transforms, and linear transforms. On the modified data, these techniques are preserving distances between the records [5–8]. Lu [9] proposed distance preserving method for PPDM by using diffusion map. Generally, diffusion maps are used for data reduction but he used encryption technique to preserve statistical information such as distance between the data records. Khaled et al. [10] proposed a method for reducing original dimensional space to low-dimensional space by using multi-dimensional scaling (MDS). Wang and Zhang [11] addressed a framework based on matrix factorization in the context of PPDM; they have used singular value decomposition (SVD) and nonnegative matrix factorization (NMF) methods. The framework focuses the accuracy and privacy issues in classification. Agarwal and Srikanth [12] built classifier from the perturbed training data; later in 2001, a distortion-based approach for preserving the privacy was introduced by Agrawal and Aggarwal [13].

Few researches proposed privacy-preserving classifiers in distributed environment. Zhan et al. [14] proposed secure computing protocol by using homomorphic cryptography for a privacy-preserving K-NN classification. Artak and Vladimir [15] proposed a privacy-preserving K-NN by extended Yao's Millionaire problem sharing the data between more than two clients.

This paper proposes a PPKNN classifier using vector operations [16]. The proposed method converts original dimensional space to high-dimensional space by converting each element of object into row unit vector with size 3.

3 Proposed Model

In this paper, we proposed a method to PPDM domain that is PPKNN using vector operations shown in Fig. 1. This method is applied to original customer's data when the business organization or data owner decided to publish their customer's data into public domain by obeying government policies. The proposed method combines both data modification to protect privacy of individual customers and data analysis for competing with other organizations those are in the same domain. PPKNN is discussed in Sect. 5. In this work, we are evaluating proposed method with K-NN classifier accuracy.

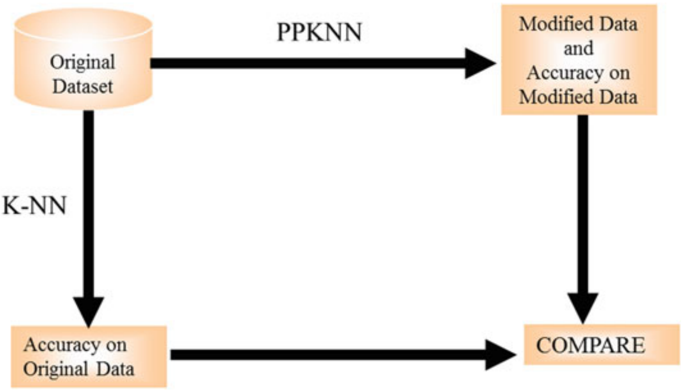


Fig. 1 Proposed model for PPKNN

4 Privacy-Preserving K-NN

This paper presents PPKNN classification algorithm for PPDM. It is similar to standard K-NN algorithm with minor changes. The PPKNN modifies original data to publish in the real world, but it preserves the distances between the records as original data.

Suppose $D = \{(d_1, c_1), (d_2, c_2), \ldots, (d_n, c_k)\}$ is a dataset, where $d_i = \{d_{i1}, d_{i2}, \ldots, d_{im}\}$ and $d_j = \{d_{j1}, d_{j2}, \ldots, d_{jm}\}$ is m-dimensional data. d_i and d_j are training and testing data, respectively, and $c_1, c_2, \ldots, c_k$ are class labels. The transformed data is $D^* = \{(d_1^*, c_1), (d_2^*, c_2), \ldots, (d_n^*, c_k)\}$ where $d_i^* = \{U_1, U_2, \ldots, U_m\}$—here U_t is row unit vector $1 \leq t \leq m$—and $d_j^* = \{V_1, V_2, \ldots, V_m\}$—here V_t is column unit vector $1 \leq t \leq m$.

The algorithm takes three parameters as input such as *D*, *K*, and K-fold are original data, number of nearest neighbors, and number of folds, respectively. In this algorithm, training data T_r and test data T_e are considered as input and the modified data T_r' and accuracy as output. Assume that dataset is represented as a matrix format. A row indicates an object, and a column indicates an attribute. Initially, D is divided into K folds. Among *K* folds, K – 1 folds are taken as training and remaining fold is test data. For each iteration, a single fold is taken as test data that is not taken in previous iteration.

Consider each object of training data as vector U_t and each object of testing data as vector V_p. Each element t_i in vector U_t is transformed into three components like t_i^2, $-2t_i$, and 1 as row vector, and each element p_i in vector V_p is converted into 1, pi, and pi^2 as a column vector. The dot product between each element of training and testing vector will be $t_i^2 - 2t_i p_i + p_i^2$. The square root of dot product results in equal to the Euclidean distance between the training and test records. Finally, compute the nearest neighbors of test data in training data by using the input parameter K to decide the class label of test data. Algorithm is shown below.

Algorithm: PPKNN

Input: D is a dataset, D= {(d_1,c_1), (d_2,c_2),..... ,(d_n,c_k)} be the set of objects where d_i is the m- dimensional feature vector and c_j is class that d_i belongs to,
K is the desired number of nearest neighbors,
K-fold is number of cross validations.

Output: T_r' (modified data) and Its Accuracy

Begin
For each fold
Sep1: Divide D into Training Data T_r and Test Data T_e
Step2: For each object of T_r
For each element of T_{ri}
$T_{ri}' \leftarrow [T_{ri}^2 \; -2T_{ri} \; 1]$
End
End
Step3: For each object of T_e
For each element of T_{ei}
$T_{ei}' \leftarrow [1 \; T_{ei} \; T_{ei}^2]^t$
End
dist (T_r, T_e) $\leftarrow T_r' \cdot T_e'$
Sort the T_r in ascending order based on distance and select first K objects.
Assign a class to T_e based on majority Vote:
$c' \leftarrow \mathrm{argmax}_y \Sigma_{(Tri\,,\,ci)}$ belonging to D
End
Step4: Measure the Accuracy of a classifier
End
Measure the Average of k accuracies.
End

5 Experimental Work

This work considers seven real datasets which are liver disorder and breast cancer Wisconsin (diagnostic) (WDBC) from UCI machine learning repository. Dataset details are shown in Table 1. We consider only numerical attributes.

Experiments are conducted on above datasets by using K-NN and PPKNN classifier on original data. Data mining tool Weka is used for K-NN classifier, and PPKNN is implemented in MATLAB. Choose three K (nearest neighbors) values around square root of the total number of rows in a dataset. Original data is converted into high-dimensional space when run PPKNN classifier and produce same accuracy as K-NN classifier because it maintains the same distance between test and training records. Assuming that each row in Table 2 is an object or a

Table 1 Dataset description

Dataset name	Number of attributes	Number of rows	Number of class labels
Liver Disorder	7	345	2
WDBC	30	569	2
Iris	4	150	3
Blood transfusion	4	748	2
Appendicitis	7	106	2
Australia credit	14	690	2
Haberman	3	306	2

Table 2 Original data

A_1	A_2
85	92
87	70
86	54
91	78

Table 3 Modified data

A_{11}	A_{12}	A_{13}	A_{21}	A_{22}	A_{23}
7225	–170	1	8464	–184	1
7569	–174	1	4900	–140	1
7396	–172	1	2916	–108	1
8281	–182	1	6084	–156	1

customer record and it is taken as training data. According to the proposed method, each cell of a row is converted into three components. For example, a cell value 85 is converted into three components such as 7225, −170, and 1 form as a unit row vector. Same procedure is applied to all cells in the training data, and the resultant modified data is shown in Table 3. When modified data is published into real world, an attacker or unauthorized person finds it difficult to reconstruct the original data and it creates the confusion to attacker about total number of attributes in the published data. When a new test data record is needed to classify, then convert each cell of test record into column unit vector and then apply dot product between converted training data and test data. Proposed PPKNN classifier and K-NN classifier are applied on different datasets described in Table 1. The same input parameter K (nearest neighbors) is used for both algorithms. We choose three different K values those are nearer to the square root of the total number of records in a dataset, and another parameter K-fold value is fixed to 5. K-fold cross-validation is one way to improve over the holdout method. The dataset is divided into K subsets, and the holdout method is repeated *K* times. Each time, one of the K subsets is used as the test set and the other K – 1 subsets are put together to form a training set. The classifier accuracy is computed as average accuracy across all *K* trials. The advantage of this method is in test data every data point is used only once and gets to be in training set K − 1 times. The variance of the resulting estimate is reduced as K value is increased. After each run of K-NN and PPKNN classifier, confusion matrix is generated for each classifier.

Comparison of PPKNN and K-NN classifier accuracies on liver and WDBC datasets is shown in Fig. 2. The deviation is very less for K = 15 on liver dataset for both classifiers accuracy. Because of our proposed method maintains same Euclidean distance in modified dataset as original dataset. It is observed that there is trade-off between accuracy and privacy; i.e., high accuracy gives low privacy, and high privacy gives low accuracy. Accuracy difference between K-NN and PPKNN classifiers on WDBC is also less. When we apply PPKNN on original datasets, some noise is added through vector operations like data perturbation techniques such as SAN and MN. The modified data will preserve the privacy of data objects. Accuracies of PPKNN and K-NN on iris and blood transfusion datasets are shown in Fig. 3. The accuracy of PPKNN on iris dataset is 91.3%, and most classifiers accuracy on iris including K-NN gives 95.33%. So, PPKNN algorithm maintains same statistical property like Euclidean distance as original data. Figure 4 and Fig. 5. show the accuracies of appendicitis, Australia, and Haberman datasets, respectively. The PPKNN algorithm removes noise in Haberman dataset. So, it works better than K-NN for k = 12 and 17. The proposed PPKNN classifier gives acceptable accuracy when compared to K-NN classifier because the square root of sum of dot product between modified training and test records is equal to Euclidean distance between training and test of original data. So, PPKNN classifier maintains neighborhood preservation and class compactness [14] in modified data.

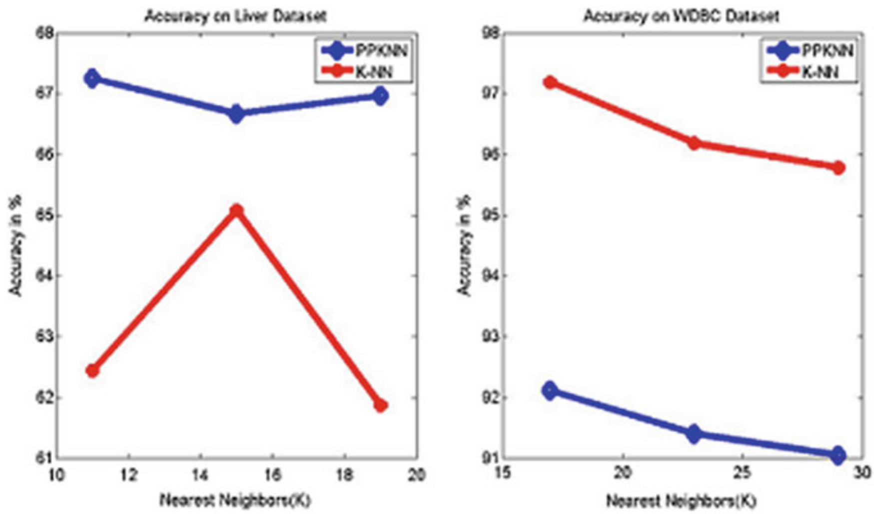


Fig. 2 Accuracies of K-NN and PPKNN on liver and WDBC dataset

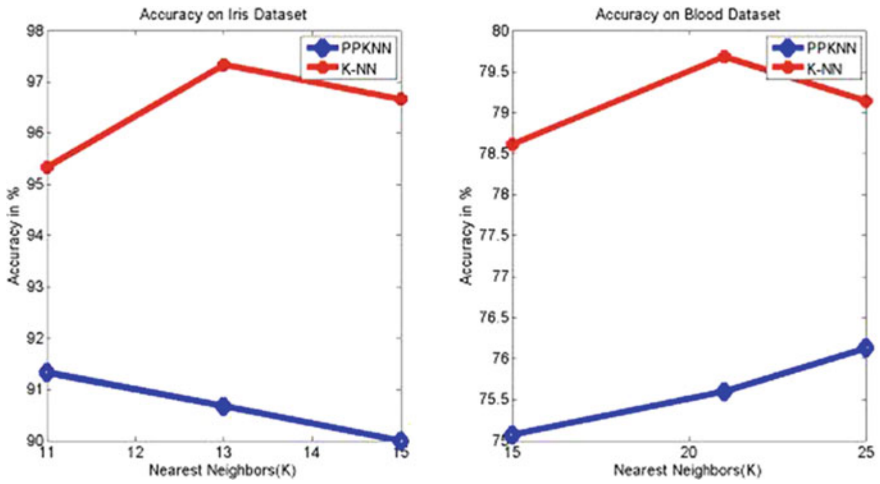


Fig. 3 Accuracies of K-NN and PPKNN on iris and blood transfusion dataset

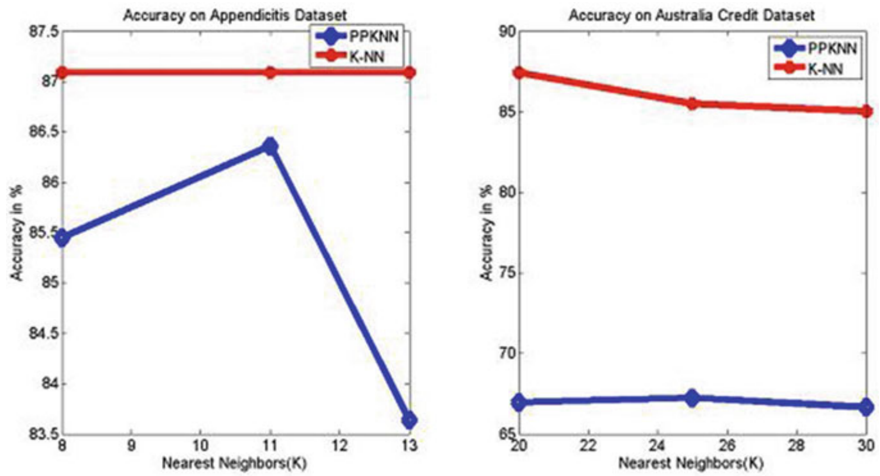


Fig. 4 Accuracy between K-NN and PPKNN on appendicitis and Australia credit datasets

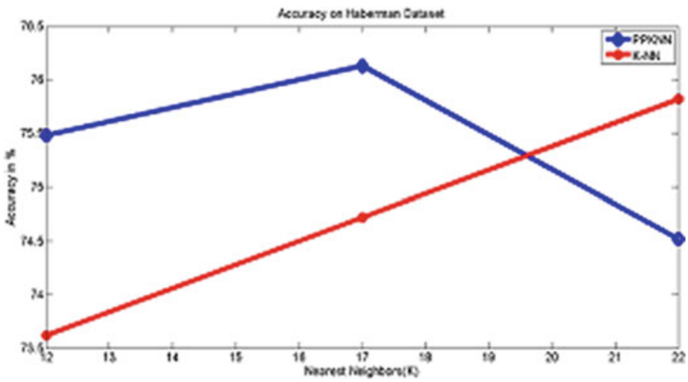


Fig. 5 Accuracy between K-NN and PPKNN on Haberman dataset

6 Conclusion and Future Work

In this work, we combined data modification and analysis into one method that is PPKNN for PPDM using simple vector operations. The proposed method works similar to original K-NN classifier, and it also preserves distance between training and test records of a dataset like other distance preserving methods. PPKNN is protecting privacy of individual customers by modifying the original data, and it gives same information by preserving the distance between the records in modified data as original data. As a future work, distributed scenario such as horizontal and vertical data can be addressed if more than one client collaborates to perform data analysis on their original data without revealing sensitive information of customers.

References

1. Traub, J.F., Yemini, Y., Woznia kowski, H.: The statistical security of a statistical database. ACM Trans. Database Syst. **9**(4), 672–679 (1984)
2. Yao, A.C.: How to generate and exchange secrets. In: Proceedings 27th IEEE Symposium on Foundations of Computer Science, pp. 162–167 (1986)
3. Vaidya, J., Clifton, C.: Privacy preserving association rule mining in vertically partitioned data. In: Proceedings of the eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 639–644
4. Vaidya, J., Clifton, C.: Privacy-preserving k-Means Clustering over vertically partitioned data. In: Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 206–215 (2003)
5. Mukharjee, S., Chen, Z., Gangopadhyay, A.: A Privacy-preserving technique for euclidean distance-based mining algorithms using fourier-related transforms. VLDB J. **15**(4) 293–315
6. Patel, V., Jain, Y.K.: Wavelet transform based data perturbation method for privacy protection. In: IEEE computer society, ICETET-2009, pp. 294–297 (2009)
7. Liu, L., Wang, J., Zhang, J.: Wavelet-based data perturbation for simultaneous privacy-preserving and statistics-preserving. In: IEEE ICDM workshop-2008, pp. 27–35 (2008)

8. Huang, J.-W., Su, J.-W., Chen, M.-S.: FISIP: a distance and correlation preserving transformation for privacy preserving data mining. In: 2011 IEEE TAAI, pp. 101–106 (2001)
9. Lu W.: Privacy preserving classification algorithm based random diffusion In: IEEE Map, SKG-2009, pp. 318–321 (2009)
10. Khaled, A., Rayward-Smith, V.J., Wang, W., de la Iglesia, B.: Non-linear dimensionality reduction for privacy-preserving data classification. In: IEEE PASSAT-2012 pp. 694–701 (2012)
11. Wang, J., Zhang, J.: Addressing accuracy issues in privacy preserving data mining through matrix factorization. In: Proceedings of IEEE ICISI, pp. 217–220 (2007)
12. Agrawal, R., Srikant, R.: Privacy preserving data mining. In: Proceeding of SIGMOD Conference on Management of Data, pp. 439–450 (2000)
13. Agrawal, D., Aggarwal, C.C.: On the design and quantification of privacy preserving data mining algorithm. In: Proceeding of ACM SIGMOD, pp. 247–255 (2001)
14. Zhan, J., Chang, L., Matwin, S.: Privacy preserving K-nearest Neighborclassification. IJNS **1**(1), 46–51 (2005)
15. Artak, A., Vladimir E.-C.: Privacy-preserving K-NN for small and large data sets. In: Proceedings of the Seventh IEEE International Conference on Data Mining Workshops (ICDMW), pp. 699–704 (2007)
16. Ravikumar, P., Cohen, W.W., Fienberg, S.E.: A secure protocol for computing string distance metrics. PSDM, ICDM 40–46 (2004)

A Comparative Study on Segmentation Techniques for Brain Tumor MRI

Imayanmosha Wahlang, Pallabi Sharma, Syeda Musfia Nasreen, Arnab Kumar Maji and Goutam Saha

Abstract Segmentation of brain tumor from medical images is an interesting topic which is investigated by many researchers. It is important to locate tumor at an early stage so that it can be easily healed and can be used for further diagnosis. There are different imaging techniques which are used in segmentation of brain tumor. Among them, Magnetic Resonance Imaging (MRI) is most widely used radiological tool as it is radiation free in nature. For detecting the size, shape, and location of the tumor many segmentation algorithms were used. In this paper, an exhaustive study of brain tumor from MRI images has made by using different techniques. A comparative study and performance evaluation of different techniques based on certain performance metrics are also discussed in this paper.

Keywords Tumor · MRI · Segmentation · Performance metrics

1 Introduction

Brain tumor occurs in many forms and can be threatening to most people, especially during the last stage. Proper diagnosis is required so that remedy can be provided at an early stage. For diagnosis of tumor, many steps are involved like preprocessing, segmentation, feature extraction, and classification. One of the most challenging

I. Wahlang (✉) · P. Sharma · S. M. Nasreen · A. K. Maji · G. Saha
North-Eastern Hill University, Mawlai, Umshing, Shillong, Meghalaya, India
e-mail: imayanwahlang@gmail.com

P. Sharma
e-mail: pallabisharma93@gmail.com

S. M. Nasreen
e-mail: musfianasreen@gmail.com

A. K. Maji
e-mail: arnab.maji@gmail.com

G. Saha
e-mail: dr.goutamsaha@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_65

tasks in brain diagnosis is segmentation. Segmentation is usually done after preprocessing, where an image will be free of noise. It is also used to divide an image into a region or objects where information is found. Segmentation is an important task where pixels will be grouped into equivalence classes. There are various methodologies that have been used in the past for segmenting an image to locate tumor. In [1], a comparative study based on segmentation techniques was done where correlation and structural similarity index were used to compare different methods like watershed algorithm, Fuzzy C-Means, region growing algorithm, morphological, deformable method, and K-Means algorithm. In [2], study on segmentation methods in brain MRI was done. Although they discussed several techniques that can be used for segmentation but no experimental analysis was provided in those papers. In our existing work, we have discussed several algorithms along with their advantages and disadvantages, supported by experimental results. Segmentation techniques that existed and which can be used in brain MRI can be broadly classified into the following:

i. **Thresholding Method** [3]: Thresholding is a method where a local or global threshold of pixels is fixed. It basically divides an image into foreground and background. There are techniques like Otsu method, histogram analysis [4] that can be used to find threshold in order to detect a region.
ii. **Region-Based Method** [3]: This method finds a region where abnormalities occur. A region will be located based on seed pixels. Existing techniques are region growing and split and merge algorithm [5].
iii. **Edge-Based Method** [3]: In this method, the edge pixels are found. It can be used for finding partition between different objects. Gray histogram technique and gradient-based method are the two main classifications where edge pixels can be found out. There are many existing techniques for this purpose like Canny edge detection, Sobel edge detection, etc.
iv. **Clustering-Based Method** [3]: In this method, clusters will be obtained where a group of similar objects is grouped together. Existing techniques used for this purpose are k-means [5, 6], fuzzy c-means (FCM) [6, 7], etc.
v. **Hybrid Method** [3]: Many algorithms have been implemented by using two or more methods for obtaining the desired result in segmentation of tumor. Some of which can be found in [5–9].
The methods which are discussed in this section is implemented using algorithms like watershed, K-Means, etc. These algorithms are discussed in brief in the next section. The paper is organized as follows:
Section 1 gives an introduction of segmentation and its different classification. Section 2 describes the study and algorithms that can be used in brain MRI. In Sect. 3, experimental result is given, and in Sect. 4, conclusion and future work are discussed.

2 An Exhaustive Study on Segmentation Techniques of Brain MRI

As discussed in the previous section, there are several existing algorithms for segmentation of brain MRI images. They are mainly classified as watershed algorithm [9], K-Means [10, 11], Fuzzy C-Means [12, 13], Otsu thresholding [14, 15], adaptive thresholding [16], Sobel operator [17], Canny operator [18], and Prewitt edge detection [19]. In this section, we are going to discuss each algorithm briefly. The algorithms are further implemented, and comparative results are shown in the next section.

2.1 Watershed Algorithm [20]

It is an algorithm that works on grayscale images. Many algorithms exist for computing watersheds one of which is Meyer's flooding algorithm. The algorithm starts with finding gradient of an image where a set of markers are chosen [9]. Each neighboring pixels are inserted into a priority queue depending on the gradient magnitude. Each and every pixel is assigned as marked when inserted into a queue. The process will continue till all the unmarked pixels are marked and inserted into a queue.

2.2 K-Means

K-Means is a clustering-based and unsupervised learning algorithm where the number of clusters is predefined [10]. The number of cluster centers is chosen randomly. Using Euclidean distance, each point is allocated to a center depending on the minimum distance of that point to the center. Once a point is allocated to a center, it will be counted in the next iteration. The steps are repeated till all the pixels are allocated to different centers.

2.3 Fuzzy C-Means (FCM)

The fuzzy clustering method can be used in situations that may arise due to partial spatial resolution, intensity of overlapping, poor contrast, noise, and intensity inhomogeneities [12]. The algorithm starts with the random selection of clusters' center which is followed by calculation of fuzzy membership using Euclidean distance. The fuzzy centers are computed, and the process is repeated until the minimum objective function is achieved.

2.4 Otsu Thresholding [14]

It is used to automatically clusters image based on threshold. The image is divided into foreground pixel and background pixel by assuming that pixels can be of two classes which followed by calculation of optimum threshold in which within class variance is high and between class variance is low.

2.5 Adaptive Thresholding [16]

It is a type of thresholding method that converts grayscale images into binary images. The threshold value at each pixel depends on the neighboring pixel intensities. Firstly, the region is selected and the weighted average is calculated. Next, threshold value is found by subtracting with a constant parameter. Based on the threshold obtained, it classifies images into classes.

2.6 Sobel Operator [17]

It performs a measurement based on the gradient where it finds the absolute gradient in x- and y-directions. It usually consists of two 3×3 kernels that are used to find approximate magnitude of changes in x- and y-direction. It can be found using Eq. (1).

$$G = \sqrt{G_x^2 + G_y^2} \tag{1}$$

where G_x is magnitude in x-direction and G_y is magnitude in y-direction.

2.7 Canny Operator [18]

It is considered as the most powerful edge detector. It is an edge detection algorithm which starts with smoothing operation using a Gaussian filter for noise removal, followed by finding gradient of an image. Next, non-maximum suppression is followed by thresholding and hysteresis.

Table 1 Advantages and disadvantages of different methods

Sl. no.	Methodology	Advantages	Disadvantages
1	Watershed algorithm	Provides closed contour	Over segmentation, sensitive to noise
2	K-means	Detection of tumor region and fast computation	Sensitive to outliers
3	Fuzzy c-means	Could detect tumor region	No. of iterations is more
4	Otsu thresholding	Simple and fast computation	Constant threshold leading to inaccurate result (binary)
5	Adaptive thresholding	Simple, thresholding is not applied over the whole image at a time but a block of pixels are taken	Wrong detection of edges, outliers cannot be detected and inaccurate
6	Sobel operator	Fast computation and detects edges where magnitude is high	Low accuracy and sensitive to noise
7	Canny operator	Simple and fast computation	Could not detect edges completely and sensitive to noise
9	Prewitt edge detection	Simple	Only works for vertical and horizontal edges and sensitive to noise

2.8 Prewitt Edge Detection Algorithm [19]

It is similar to Sobel operator that starts with finding gradient of the image. It is used for edge detection where it helps to find edges in horizontal or vertical direction based on gradient pixel.

Based on the above study, we have found out several advantages and disadvantages of the segmentation techniques which are listed in Table 1.

3 Experimental Result

Performance metrics that are used for comparing the different segmentation techniques are correlation, entropy, Peak Signal-to-Noise Ratio (PSNR), computation time, and Mean Absolute Error (MAE) [4, 5].

i. **Peak Signal-to-Noise Ratio (PSNR)**: It is a measurement that gives the ratio between power of a signal and power of corrupting noise. It can be obtained using the formula given in Eq. (2) [21, 22].

$$PSNR = 10log_{10}\frac{255^2}{\sum_{i,j} X(i,j) - \bar{X}(i,j)} \tag{2}$$

where X(i, j) denotes the original image pixel and $\bar{X}(i,j)$ denotes the restored image pixel.

ii. **Entropy**: It is a measure of randomness that describes texture and is found using Eq. (3) [5].

$$Entropy = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j) \times logp(i,j) \tag{3}$$

where P(i, j) is the probability function defined over pixel (i, j).

iii. **Correlation**: It is defined as measure of the degree of correlation a pixel has to its neighbor over the whole image. Correlation is specified using the formula given in Eq. (4) [1, 5].

$$Correlation = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{i \times P(i,j) - \mu_x \times \mu_y}{\sigma_x - \sigma_y} \tag{4}$$

iv. **Computation Time**: It is the time taken to complete a computational process. It is calculated from the time that the process starts till the process finishes.

v. **Mean Absolute Error (MAE)**: It is used to measure accuracy. It measures how much-predicted value differs from the actual value. Higher its value poorer is the image quality [4].

100 MRI brain images* were used for comparison of segmentation methods, and the following results were obtained by calculating the average performance of the different methods. Table 2 gives the overall performance of the implemented techniques. Some of the result obtained after application of segmentation methodologies

Table 2 Performance comparison table of different segmentation techniques for brain MRI

Method	PSNR	Entropy	Correlation	MAE	Time
Watershed	12.7407	1.4029	0.3846	76.4092	0.0415
FCM	9.7254	0.5770	0.3680	48.0294	0.6536
Adaptive	6.8753	0.8737	0.1535	81.3109	0.0403
Thresholding	9.8574	0.5688	0.9296	81.8784	0.0013
K-means	14.9201	3.4801	0.9722	29.8304	0.7650
Sobel operator	13.5818	0.2736	0.2518	81.9824	0.0047
Canny operator	11.2948	0.4105	0.2858	81.9820	0.0077
Prewitt	13.5269	0.2754	0.2500	81.9824	0.0070

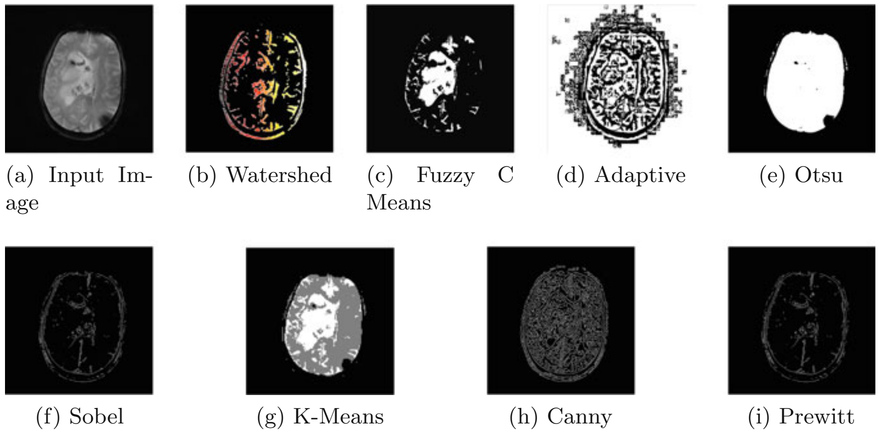


Fig. 1 Segmentation output of different methods

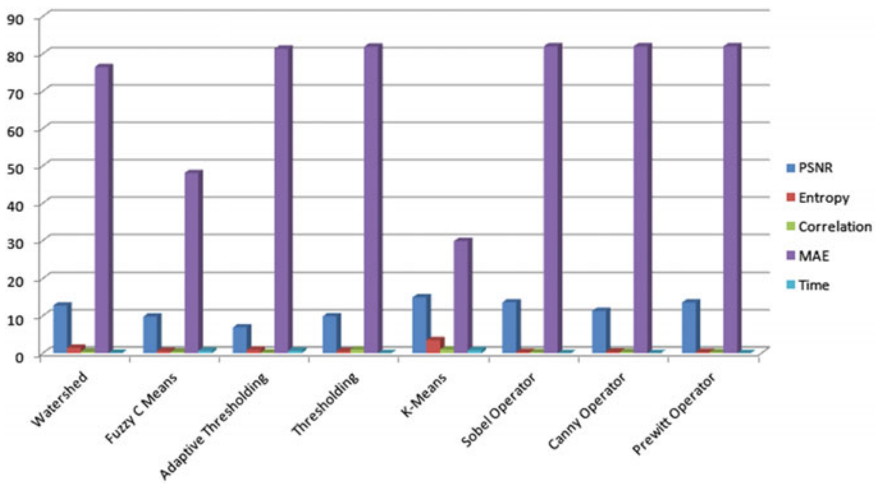


Fig. 2 Graphical representation of comparison of different methods

along with the graphical representation of comparison can be seen in Figs. 1 and 2, respectively.

After successful implementation, it was found that Sobel operator gives the best entropy, adaptive thresholding the best correlation, and K-Means the best MAE. Overall, Fuzzy C-Means (FCM) gives the best output when entropy, correlation, PSNR, and MAE are taken into account.

4 Conclusion and Future Work

A brief description of various techniques that can be used in segmentation of brain MRI was given in this paper. The different methods were implemented and compared using MATLAB and were found that there is still scope for improvement of segmentation techniques, which can be used in the classification of tumor. An automated and accurate technique is needed so that it would be helpful not only to people but to a medical practitioner as well.

Despite many existing works, brain tumor segmentation is still needed to be implemented for automated segmentation that can be further used for classification purpose to find location, size, and type of tumor. Works based on advanced machine learning like deep learning is implemented to obtain an efficient, accurate, and reliable result but still lack in many aspects like grading or staging of tumor, heterogeneity, and classification which needs to be further investigated.

Acknowledgements *The authors would like to thank Dr. Pranjal Phukan, Department of Radiology, North Eastern Indira Gandhi Regional Institute of Health & Medical Sciences (NEIGR IHMS), who provided MRI brain images and assistance for our work.

References

1. Mustafa, I., Hassan, D., Mawia, A.: A comparison between different segmentation techniques used in medical imaging. Am. J. Biomed. Eng. **2**(6), 59–69 (2016)
2. Deshmukh, R.D., Jadhav, C.: Study of different brain tumor MRI image segmentation techniques. Int. J. Sci. Eng. Comput. Technol. **4**(4), 133 (2014)
3. Harsimranjot, K., Sharma, R.: A survey on techniques for brain tumor segmentation from MRI, classification techniques for medical images. IOSR J. Electron. Commun. Eng. **5**(11), 1–5 (2016)
4. Patel, N., Kumar, R.: Image segmentation & performance evaluation. Int. J. Res. Appl. Sci. Eng. Technol. **9**(2), 46–51 (2014)
5. Balasubramanian, C., Sudha, B.: Comparative study of de-noising, segmentation, feature extraction, classification techniques for medical images. Int. J. Innov. Res. Sci. Eng. Technol. **3**(3), 1194–1200 (2014)
6. Jose, A., Ravi, A., Sambath, M.: A comparison between different segmentation techniques used in medical imaging. Int. J. Innov. Res. Comput. Commun. Eng. **3**(2), 3496–3501 (2014)
7. Kaur, H., Reecha, S.: Segmentation of tumor region from brain MRI images using fuzzy C-means clustering and seeded region growing. Int. Organ. Sci. Res. J. Comput. Eng. **5**(18), 20–24 (2016)
8. Anithadevi, D., Perumal, K.: A hybrid approach based segmentation technique for brain tumor in MRI Images, pp. 1–11 (2016). arXiv:1603.02447
9. Mustaqeem, A., Javed, A., Fatima, T.: An efficient brain tumor detection algorithm using watershed & thresholding based segmentation. Int. J. Image Gr. Signal Process. **10**(4), 34 (2012)
10. Selvakumar, J., Lakshmi, A., Arivoli, T.: Brain tumor segmentation and its area calculation in brain MR images using K-means clustering and fuzzy C-mean algorithm. In: International Conference on Advances in Engineering, Science and Management, pp. 186–190. IEEE (2012)
11. Joseph, R.P., Singh, C.S., Manikandan, M.: Brain tumor MRI image segmentation and detection in image processing. Int. J. Res. Eng. Technol. **1**(3), 1–5 (2014)

12. Monteiro, F.C., Campilho, A.C.: Performance evaluation of image segmentation. In: International Conference Image Analysis and Recognition, pp. 248–259. Springer (2006)
13. Yambal, M., Gupta, H.: Image segmentation using fuzzy C means clustering: a survey. Int. J. Adv. Res. Comput. Commun. Eng. **7**(2), 2927–2929 (2013)
14. Wikipedia. https://en.wikipedia.org/wiki/Otsu%27s_method
15. Tirpude, N., Welekar, R.: Automated detection and extraction of brain tumor from MRI images. Int. J. Comput. Appl. **4**(77), 26–30 (2013)
16. Wikipedia. http://hanzratech.in/2015/01/21/adaptive-thresholding.html
17. Wikipedia. https://en.wikipedia.org/wiki/Sobel_operator
18. Wikipedia. https://en.wikipedia.org/wiki/Canny_edge_detector
19. Wikipedia. https://en.wikipedia.org/wiki/Prewitt_operator
20. Wikipedia. https://en.wikipedia.org/wiki/Watershed_(image_processing)
21. Sumithra, M.G., Deepa, B.: Performance analysis of various segmentation techniques for detection of brain abnormality. In: Region 10 Conference (TENCON), pp. 2056–2061. IEEE (2016)
22. Singh, H., Verma, S., Marwah, G.K.: The new approach for medical enhancement in texture classification and feature extraction of lung MRI images by using gabor filter with wavelet transform. Indian J. Sci. Technol. **35**(8), 1–7 (2015)

Integrating Concentrating Solar Plant-Based System in Multi-area AGC Using LabVIEW

Dillip K. Mishra, Tapas Kumar Panigrahi, Asit Mohanty and Prakash K. Ray

Abstract An attempt has been made to analyze the model of automatic generation control (AGC) of a **two** unequal interconnected deregulated power systems using LabVIEW. In this study, teaching–learning process-inspired algorithm, i.e., TLBO is employed for tuning the gains of PID controller. This study portrays the investigation of an AGC for two-area power system with PID, DE-based PID, and TLBO-enhanced PID. The execution of TLBO-PID is discovered and enhanced in contrast with PID, DE-PID controller as far as stability, overshoot, and damping. The inclusion of generation rate constraint (GRC) leads to difficult task for the design of successful controller. The difficulty is improved by step and random load variation. Examination demonstrates that TLBO improves and optimizes the control parameters at different loading conditions and wide change in sustainable power source, area capacity ratio. Consequently, incorporation of thermal power–concentrating solar power (CSP) for AGC has improved the interconnected power system.

Keywords AGC · CSP · ED · LabVIEW · LFC · PID
TLBO

D. K. Mishra (✉) · T. K. Panigrahi
International Institute of Information Technology, Bhubaneswar, India
e-mail: chandanmishra4@gmail.com

T. K. Panigrahi
e-mail: tapas@iiit-bh.ac.in

A. Mohanty
College of Engineering & Technology, Bhubaneswar, India
e-mail: asithimanshu@gmail.com

P. K. Ray
Naynag Technical University, Singapore, Singapore
e-mail: pkrayiiit@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_66

1 Introduction

The idea behind the AGC is to maintain the system frequency, and tie-line power is remaining constant [1]. Since a grid power requires that load and generation power almost conform step by step, progressive changes as per the output of generators are fundamental. The adjustment can be judged by measuring the system frequency; if it is extending, more power is being made than used, and each one of the machines in the system is stimulating. If the system repeat is decreasing, a more number of loads are on the system than the quick generation can give, and all generators are down. In the present power scenario, AGC deals with load frequency control, economic dispatch, and interchange scheduling. When the sudden load changes, then frequency change and governing system respond quickly. If frequency deviation is little more, then governor may not be able to restore the frequency. At that instant of time, one supplementary control loop, i.e., load frequency control (LFC) has come into play. LFC controls the area frequency and inter-area tie-line power. Economic dispatch (ED) is the capability of sharing the load in optimum fashion in a short duration of time where the number of generating units is available in single area at the most reduced cost subject to the system constraint. The economic dispatch problem is comprehended by specific PC programming which ought to respect the operational and framework imperatives of the accessible assets and comparing transmission abilities. In an interconnected power system, there are different areas related by tie lines with neighbor area. Both LFC and ED require too successful of these exercises. This exercise is executed by interchange scheduling (IS). Each area of the network is responsible for generating sufficient amount of power to meet the load demand for their customer. So that the generation and load can be balanced keeping the frequency maintained constant at 50 Hz. It is not that the area should confirm the power generation to supply the customer demand, but they should take care of any reserve tie-line exchange. It is possible by observing the change in frequency and tie-line power characteristics to choose the most possible generative action either it is to raise or lower. Therefore, at every minute AGC will monitor the tie-line power and frequency of each area for utilization of electrical utilities.

Nowadays, power system is getting to be plainly unpredictable time by time, because of its expansion in size and reconciliation of sustainable power sources [2]. The entrance of sustainable power sources is because of the depletion of conventional power sources and the bad impacts of the power resources on the environment [3]. Then again, sustainable power sources are ample in nature, makes less damage to the earth, and additionally, its conversion efficiency is more. Out of all sustainable power sources, solar is more common [4] under consideration. The overall efficiency of the interconnected power system is measured through simple controlling action. The system controllability and stability area are taken care by the integration of renewable energy sources. For stable AGC work, numerous researchers have revealed utilizing traditional controllers because of the simplicity and controllability of the power system, and different controlling activities are made subsequently [5]. The traditional method and various evolutionary computational techniques, such as differential evolution (DE) [6], genetic algorithms (GAs) [7], bacterial foraging optimization (BFO) [8], have been available in order to solve the AGC problems. Traditional methods such as trial-and-error method always require more time and give poor outcome [9].

2 Proposed Hybrid Electrical Power System

In this study, two unequal area thermal–CSP plants are considered here. In area-I solar plus thermal with reheat and area-II thermal with reheat plant is considered. The capacity of plant is 2000 MW in each area. GRC is also taken for thermal units as 3% per minute. TLBO optimization technique has been applied for tuning the gains of the PID controllers. In the present study, three-area restructure power system comprises solar thermal with reheat, and hydro and solar thermal with non-reheat is considered. Each area has a governing system, turbine, GRC, and PID controller is employed. TLBO algorithm has been applied for tuning the gains of the PID controllers. The dynamic performance is measured with frequency deviation, and tie-line power is measured.

The area control error can be written as [10].

$$\mathrm{ACE} = \beta\Delta\mathrm{F} + \Delta\mathrm{P}_{\mathrm{Tie}} \tag{1}$$

where β can be represented as frequency bias parameter.

The turbine of thermal plant can be represented as [5]:

$$\mathrm{G_T(s)} = \frac{\Delta\mathrm{P_T(s)}}{\Delta\mathrm{P_V(s)}} = \frac{1}{1 + \mathrm{sT_T}} \tag{2}$$

The governor of thermal plant can be represented as

$$\mathrm{G_G(s)} = \frac{\Delta\mathrm{P_V(s)}}{\Delta\mathrm{P_G(s)}} = \frac{1}{1 + \mathrm{sT_G}} \tag{3}$$

Two inputs of speed governing system, i.e., $\Delta\mathrm{P_{ref}}$ and $\Delta\mathrm{f}$ with one output $\Delta P_G(s)$

$$\Delta\mathrm{P_G(s)} = \Delta\mathrm{P_{ref}(s)} - \frac{1}{\mathrm{R}}\Delta\mathrm{f(s)} \tag{4}$$

The transfer function of generator and load can be represented by [4]:

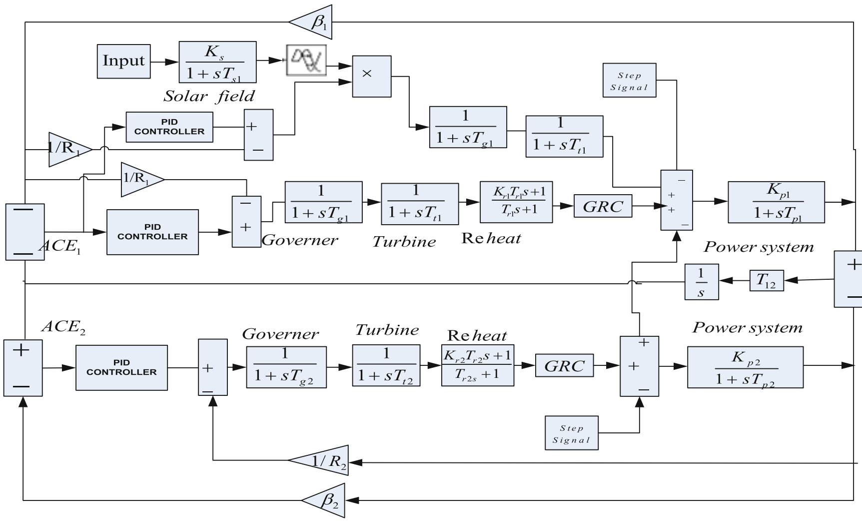


Fig. 1 Transfer function model of two-area hybrid power system

$$G_P(s) = \frac{K_P}{1 + sT_P} \tag{5}$$

where $K_P = 1/D$ and $T_P = 2H/fD$.

The two inputs of the generator load are $\Delta P_T(s)$ and $\Delta P_D(s)$ with single output $\Delta F(s)$ which can be written as [11, 12] (Fig. 1):

$$\Delta F(s) = G_P(s)[\Delta P_T(s) - \Delta P_D(s)] \tag{6}$$

3 Solar Thermal Power Plant

Solar thermal technologies, otherwise called concentrating solar power (CSP), use warm vitality from the sun to produce power. CSP technologies track the sun on either one or two axes, and mirrors are arranged to focus the sunlight in either line-focus concentrators or point-focus concentrators. The high temperatures achieved at the focal point of the concentrators are used to heat an intermediate heat-exchanging fluid, which can either be used for thermal energy storage, boiling water (for steam turbine), or powering thermal engines. Presently, four types of CSP employed in different applications such as: parabolic trough, power tower, linear Fresnel reflector, and Dish Sterling (DS) designs. Parabolic trough is one of best inventions of the CSP technologies. Nowadays, two STPS such as solar power tower and trough solar power tower are widely used. In addition to this, other sources such as wind, fossil fuel, tidal can be added to make the hybrid system.

4 Control Technique

In the modern area of control theory application, the control inputs of three-area interconnected system u_1 and u_2 can be generated by linear combination of all states of the system. Controller means monitors the set point of the system, so that the system may decrease the large disturbance, errors, overshoot, and oscillation [13]. In this present study, proportional integral derivative (PID) controller is used to monitor the frequency deviation of each area and tie-line power.

4.1 PID Controller

PID controller is otherwise called proportional plus integral controller plus derivative controller. It is an extremely prominent feedback controller in industrial applications. It is a lively, viably appreciated controller that can give phenomenal control execution regardless of dynamic achievement of a plant.

4.2 Teaching–Learning-Based Optimization (TLBO)

Teaching–Learning-Based Optimization (TLBO) technique was first developed by [14]. This method is now very popular and most effective technique and employed in many fields of application. This technique has numerous advantages like lesser time is required for the best solution and gives more stable performance with multiple frequency constraint. TLBO is mainly divided into two phases.

a. First phase (teacher phase), and
b. Second phase (learner phase).

During the first phase, students learn from the teacher, and in the second phase, students learn by communicating with other students. The above two phases of TLBO algorithm can be described as follows:

Initialization:
Initially, generate the random number of population and number of dimension variable parameters, i.e., N_p and D. The mark secured by each number of students

in different subject can be written in matrix form, i.e., *i*th column with *i*th subject. Here initial population taken as X.

a. **First phase (teacher phase)**

   In the first phase, the influence of teacher on the students in a class. The teacher has to play a major role in this phase, and he must have very strong knowledge in their assigned subject and to teach the students to get the best score and performance. The best score shows the shine of a best teacher as compared to another teacher, i.e., X_{best}

   The mean value of result of each student in each subject can be estimated as:

$$M_d = [m_1, m_2, \ldots, m_D] \tag{7}$$

   The results can be compared with mean value of same subject, and the results of assigned teachers are given by

$$M_{diff} = \text{rand}(0,1)[X_{best} - T_F M_d] \tag{8}$$

   where T_F can be represented as teaching factor with $rand(0,1)$ and has been regarded as random number between 0 and 1. Value of T_F can be either 1 or 2 and is chosen randomly from Eq. (9)

$$T_F = \text{round}[1 + \text{rand}(0,1)] \tag{9}$$

   The old population is now calculated by Eq. (11)

$$X_{new} = X + M_{diff} \tag{11}$$

   New value of population is accepted (New) o if $f(X_{new}) < f(X)$ else X_{old} are accepted.

b. **Second phase (learner phase)**

   In the second phase, the learner can improve their knowledge in two ways. One can go through more discussion with the teacher, and other can communicate with other learner themselves. A learner takes advantages from other learner with random selection of learner and interacts with him or her. So the learner can get more knowledge. The process of this phase can be expressed as:

   Select two learners randomly, X_i and X_j where $i \neq j$.

$$X_{new} = X_i + \text{rand}(0,1)(X_i - X_j), \text{ if } f(X_i) < f(X_j) \tag{12}$$

   else

   $X_{new} = X_i + \text{rand}(0,1)(X_j - X_i)$. Accept X_{new} if it is best solution.

4.3 Choice of Objective Function

The main objective of the AGC is to reduce the area control error (ACE) with a very short period of time. In order to get low value of ACE, the cost function can be defined as

$$f = \int_0^t |\Delta w_1 - \Delta w_2| \cdot t dt \tag{13}$$

where dt is change in time, and Δw_1 and Δw_2 are the frequency deviation in area-I and area-II, respectively. In the present study, a step load disturbance of 1% is changed in area-I and starts simulation with TLBO optimization technique for 50 times to get the optimal values of gain of the PID controller. Two more parameters are put into algorithm, i.e., number of population and maximum number of iteration.

5 Results and Discussion

In this section, comparative analysis is done and shows the effectiveness of TLBO-PID controller. The optimum value of PID gains has been found out by TLBO optimization technique. Dynamic performance of two-area system is measured for 10% step load disturbance in area-I through LabVIEW environment (Fig. 2).

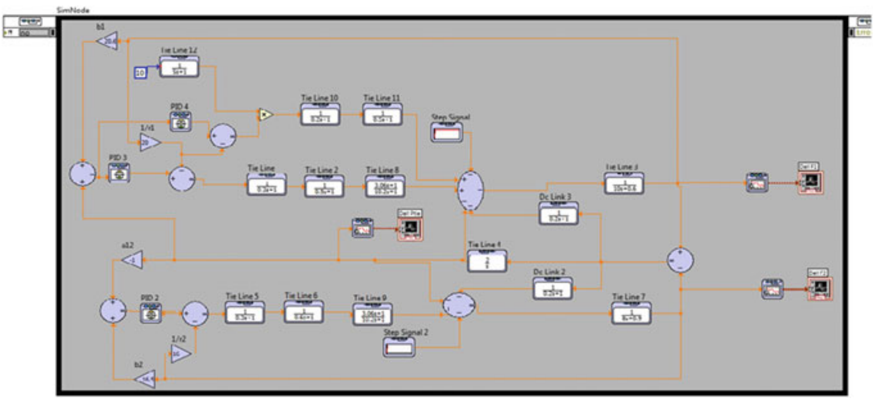


Fig. 2 LabVIEW front panel showing the detailed model

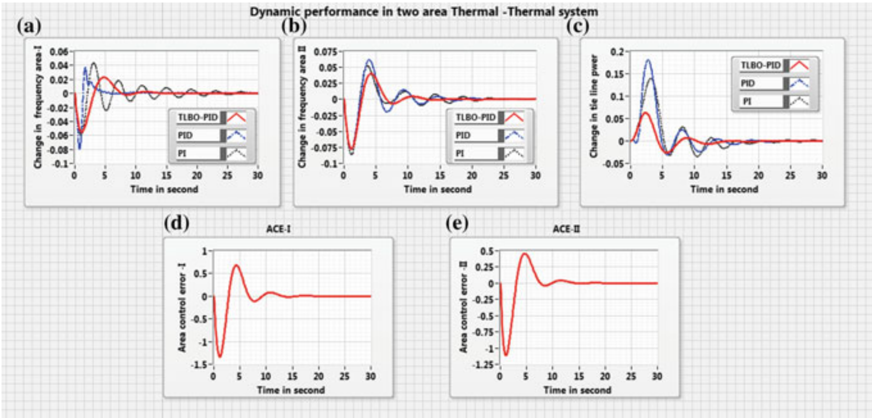


Fig. 3 Dynamic performance of two-area system without solar. **a** ΔF_1, **b** ΔF_2, **c** ΔP_{tie}, **d** ACE-I, **e** ACE-II

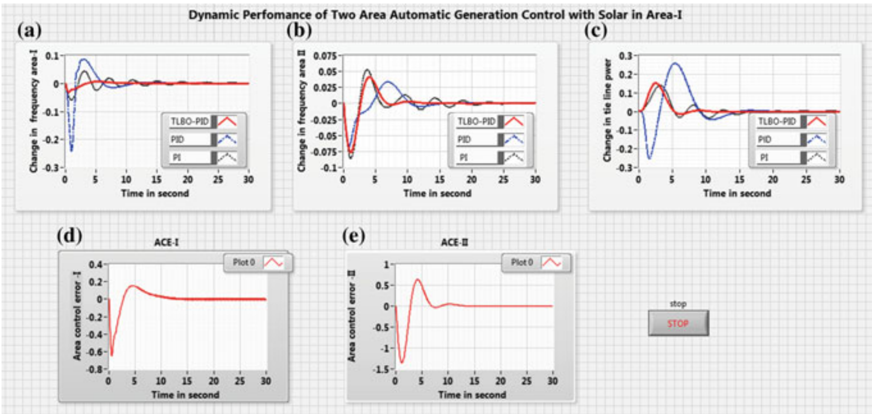


Fig. 4 Dynamic performance of two-area system with solar integration. **a** ΔF_1, **b** ΔF_2, **c** ΔP_{tie}, **d** ACE-I, **e** ACE-II

The frequency deviation, tie-line power deviation, and area control error response are shown in Figs. 3, 4, and 5 with different cases. It is clear that using TLBO-optimized PID controller gives more stable results and quite robust as compared to other controllers.

The dynamic performance of proposed hybrid system is analyzed with different cases.

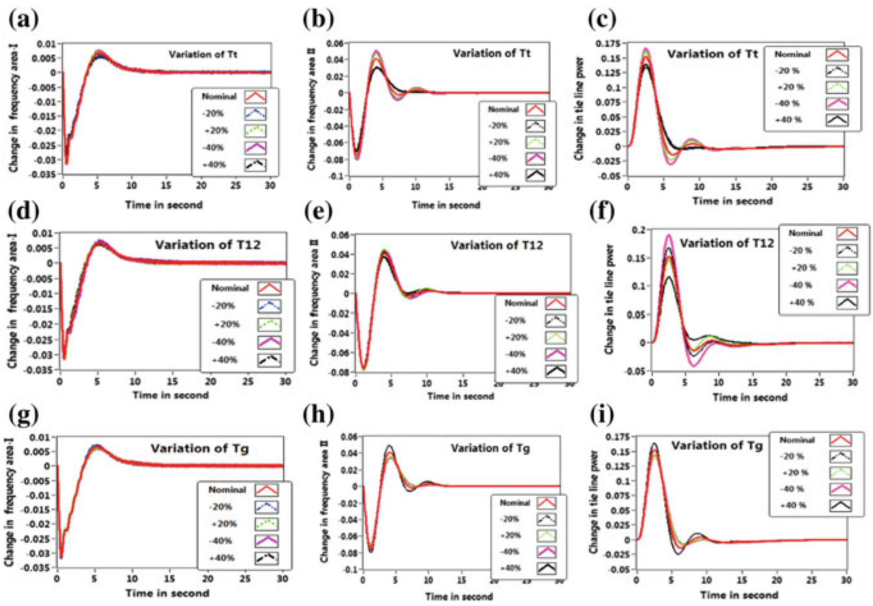


Fig. 5 Frequency deviation and tie-line power change with loading conditions and different parameters

A. 10% step load change in area-I
   (i) Thermal–thermal
   (ii) Thermal–CSP hybrid
B. Parameter variation
C. Random loading.

5.1 Case-A (10% Step Load Change in Area-I)

I. Thermal–Thermal

In the first case, a step increase in demand in the area-I is applied without solar integration. The change in frequency in area-I (ΔF_1), change in frequency in area-II (ΔF_2), tie-line power deviation (ΔP_{tie}), and area control error (ACE) is measured and shown in Fig. 4. Three different controllers are applied to the same system and compared as shown in Fig. 3a, b, c, d, and e. Optimum value of PID controller gain (K_P, K_I, and KD) is tuned with TLBO algorithm, which shows the better response. It is observed that using TLBO technique, system stability increases and decreases settling time and overshot. The settling time of frequency deviation, tie-line power, and area control error is shown in Table 1.

Table 1 Performance analysis

Controller	Settling time (s)				
	ΔF_1	ΔF_1	ΔP_{tie}	ACE_1	ACE_2
PI	29.12	28.35	29.32	28.22	27.92
PID	18.25	21.65	24.16	17.26	20.26
TLBO-PID	15.18	16.23	20.25	14.12	15.79

Table 2 Performance analysis

Controller	Settling time (s)				
	ΔF_1	ΔF_1	ΔP_{tie}	ACE-I	ACE-II
PI	26.32	25.04	27.50	26.69	24.32
PID	15.32	17.85	20.62	16.72	18.11
TLBO-PID	12.75	12.24	17.50	13.27	14.52

II. Thermal–CSP Hybrid
In this section, concentrating solar power (CSP) is added into the area-I which makes hybrid system, i.e., thermal–CSP hybrid system. In the same system, three different controllers such as PI, PID, and TLBO-PID are applied and compared as shown in Fig. 4. As seen in the result, it is clear that using CSP in area-I and TLBO-optimized PID bring the system more stable, smaller settling, overshoot decreases and zero steady-state error. Frequency deviation, tie-line power deviation, and area control error are shown in Fig. 4a, b, c, d, and e. The result demonstrates the impact of CSP on AGC in multi-area system, and settling time of different controllers is shown in Table 2.

5.2 *Case-B (Parameter Variation)*

This section describes the impact on frequency deviation, tie-line power by parameter variation, and working condition of the PID controller gains in the two-area CSP–thermal hybrid system. Here the variable parameter is T_{12} (synchronizing coefficient). T_{12} is varied ±50% from their nominal value. It is observed that, by varying T_{12} in the proposed system, a very small change in frequency deviation is noticed in area-I and area-II. Some impact on tie line power deviation is also noticed as shown in Fig. 5a, b, c, d, e, f, g, h and i and the performance analysis is depicted in Table 3.

Table 3 Performance analysis of sensitivity analysis

Parameter	Loading	Settling time in sec			Undershoot		
		ΔF_1	ΔF_1	ΔP_{tie}	ΔF_1	ΔF_1	ΔP_{tie}
T_g	Nominal	12.75	12.24	17.50	0.031	0.078	0.004
	+20%	12.94	12.31	17.89	0.031	0.078	0.0039
	−20%	12.71	12.22	17.34	0.032	0.079	0.004
	+40%	13.01	12.94	18.05	0.031	0.076	0.0041
	−40%	12.16	12.01	17.09	0.032	0.080	0.0039
T_t	+20%	12.84	12.34	17.61	0.031	0.079	0.0040
	−20%	12.61	12.22	17.80	0.033	0.077	0.0039
	+40%	13.03	12.91	17.41	0.031	0.078	0.0040
	−40%	12.26	12.08	18.15	0.030	0.081	0.0040
T_{12}	+20%	12.74	12.29	17.50	0.031	0.078	0.0041
	−20%	12.11	12.21	17.87	0.031	0.079	0.0038
	+40%	13.31	12.68	17.14	0.031	0.076	0.0041
	−40%	12.14	12.05	18.45	0.033	0.081	0.0039

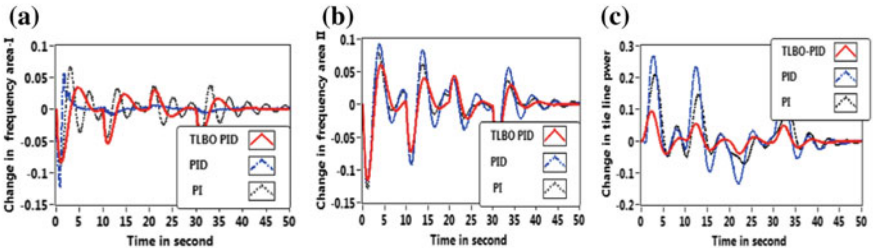


Fig. 6 Dynamic performance of two-area system with solar integration. **a** ΔF_1, **b** ΔF_2, **c** ΔP_{tie}

5.3 *Case-C (Random Loading)*

Figure 6 shows the random loading of CSP–thermal hybrid system. The change in frequency in area-I, change in frequency in area-II, and tie-line deviation is shown in Fig. 6a, b, and c with three different controllers.

6 Conclusion

The paper discusses an unequal two-area thermal CSP plant hybrid power system where TLBO algorithm is applied for tuning the gains of the PID controller. In this proposed hybrid system, the power balance can be maintaining automatically by governing action of thermal plants. The dynamic performance of two-area systems is measured with three different controllers such as PI, PID, and TLBO-PID. A TLBO-PID controller result shows more robust and better performance. In this study, also random load is applied to the same model and shows the performance of the system. At last, sensitivity analysis is done by ±50% loading.

References

1. Elgard, O.I., Fosha, C.: Optimum megawatt frequency control of multi area electric energy system. IEEE Trans. Power Syst. **89**(4), 556–563 (1970)
2. Das, D.C., Sinha, N., Roy, A.K.: GA based frequency controller for solar solar thermal—diesel-wind hybrid energy generation/energy storage system. Int. J. Electr. Power Energy Syst. **43**(1), 262–279 (2012)
3. Bevrani, H., Ghosh, A., Ledwich, G.: Renewable energy source and frequency regulation: survey and new perspective. IET Renew. Power Gener. **4**(5), 438–457 (2010)
4. Valenzuela, L., Zara, E., Berenguel, M., Camacho, E.F.: Direct steam generation in solar boiler. IEEE Control Syst. **24**(2), 15–29 (2004)
5. Saika, L.C., Nanda, J., Mishra, S.: Perfomance comparison of several classical controller in AGC multi area interconnected thermal system. Int. J. Electr. Power Energy Syst. **33**(3), 394–401 (2011)
6. Sahu, R.K., Sidharth, P., Kumar, R.U.: DE optimised parallel 2-DOF PID controller for load frequency control of power system with governer dead band nonlinearity. Electr. Power Energy Syst. **49**, 19–23 (2013)
7. Ghoshal, S.P.: Application of GA/GA-SA based fuzzy automatic generation control of a multi area thermal generating system. Electr. Power Syst. Res. **70**, 115–127 (2004)
8. Nanda, J., Mishra, S., Saikia, L.C.: Maiden application of bacterial foraging based optimization technique in multi area automatic generation control. IEEE Trans. Power Syst. **24**(2), 602–609 (2009)
9. Shabani, H., Vahidi, B., Ebrahimpur, M.: A roboust PID controller based on imperialist competitive algorithm for load frequency control of power system. ISA Trans. **52**(1), 88–95 (2013)
10. Bevrani, H.: Roboust power system frequency control. Springer (2009)
11. Mishra, D.K., Mohanty, A., Ray, P.K.: Multi area automatic generation control with FOPID and TID controllers. Int. J. Control Theory Appl. **10**(6), 397–410 (2017)
12. Kundur, P.: Power system stability and control. Mc Graw Hill (1993)
13. Shabani, H., Vahidi, B., Ebrahimpur, M.: A roboust PID controller based on imperialist competitive algorithm for load freqency control of power system, ISA transactions, 52(1), 88–95, (2013)
14. Rao, R.V., Savsani, V.J., Vakharia, D.P.: Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput. Aided Des. **43**(3), 303–315 (2011)

Maximal Stable Extremal Region Extraction of MRI Tumor Images Using Successive Otsu Algorithm

Priya Jyotiyana and Saurabh Maheshwari

Abstract The computer-aided diagnosis of magnetic resonance imaging (MRI) images using computer vision techniques is a facilitating area of interest. The segmentation model tries to differentiate tumor regions in an MRI section. The MRI is better than computerized tomography (CT) scan images since there is no radiation in the magnetic resonance operation, and furthermore, the outcomes obtained utilizing MR images are more useful and clear when compare with CT scan image. In this paper, we proposed a novel method of successive Otsu algorithm applied it into hierarchical levels to define segmentation threshold for MRI. Maximal stable extremal region (MSER) is more appropriately used for objective detection as well as tracking, as it detects the major MSER from both the depth image and the intensity image. Appling MSER initially, the input image is changed from a pixel-based to a region-based model by utilizing the MSER. The segmentation performance is evaluated by using four stability parameters, namely probability randomness indexing, variation often information, globally consistency errorless, PSNR measures. The purpose of segmentationing method is carried out by utilizing the successive Otsu algorithm. The outcome of segmented MRI images is used to extract the tumor area by introducing the MSER.

Keywords Image thresholding · Brain tumor · Magnetic resonance imaging
Image segmentation · Maximal stable extremal region

P. Jyotiyana (✉) · S. Maheshwari
Department of Computer Science and Engineering, Government Women Engineering College, Ajmer, India
e-mail: priya7653@gmail.com

S. Maheshwari
e-mail: dr.msaurabh@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_67

1 Introduction

Digital image processing has been developed a unique area of interest in research field. The significance of extracting data from images is essential on daily basis applications including satellite framework, text detection, face detection, computerized medical field, military services. The procedure of image processing describes the set of input data, or input images are utilized to produce the output image or the set of output data [1]. In Image segmentation technique, an image can be segmented into regions that usually keep in touch with object on images. Pixels are assembled into regions that distribute several general properties (e.g. color, intensity).

Image segmentation is a technique of image processing so that a single image is segmented into several parts. After applying the segmented process, every component of image is utilized for a particular reason, e.g., to recognize object or other important data. Various methods of image segmentation are used for segmentation technique modal-based method, histogram-based method, thresholding, region-based method, etc. [2]. Most trendy method for image segmentation is thresholding. In this, it takes input of grayscale image and changes each pixel with a black if intensity value is less than fixed constant value, or white pixel occur if the value of intensity is greater than the constant value. The new image which is generated from the values separates the bright region [3]. Thresholding is easy and resourceful method for image segmentation [4].

1.1 Our Contribution

This paper has made an evaluation of threshold value on the basis of Otsu and successive Otsu method. The paper addresses the problems of:

- Tumor identification at multiple hierarchical levels using successive Otsu method.
- MSER region: the area is monitored. Regions such that their dissimilarity with respect to the threshold is minimal are defined maximally stable.
- Evaluation of MSER region on the basis of four parameters namely improbability randomness indexing, globally inconsistency errorless, variationing of informationing, PSNR.
- Various algorithms suffer from the reality that there is no usual technique to find out the number of optimum thresholds. Later than applying the current algorithm, PSNR of the threshold image is found to saturate. This property can be used to find the appropriate number of thresholds.

2 Otsu's Algorithm

Otsu algorithm is a basic and well-known thresholding strategy for image segmentation which can be falls into clustering. Image histogram can be divided into two classes by utilizing a threshold or limit; e.g., the in-class changeability is small [5, 6].

$$g(x,y) = \begin{cases} 1, & f(x,y) \geq T \\ 0, & x < T \end{cases} \tag{2.1}$$

Pixels be divided into two classes; the value of within the class variance $\sigma_w^2(t)$ is minimized that can be define by Eq. (2.2). The threshold is defined by the variable (t) whose significance among 0 in addition to 255.

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t) \tag{2.2}$$

For every pixel value, a probability function is finding. Firstly, histogram of image can be calculated afterward standardization is performed which follows the probability distribution function [7]. Then values of pixels can be categorized into two namely C1 and C2 using threshold (t), and the probability function of class can be defines as q1(t) and q2(t) as in Eqs. (2.3) and (2.4)

$$\sigma_b^2(t) = q_1(t)q_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{2.3}$$

$$q_2(t) = \sum_{i=t+1}^{i} P_i \tag{2.4}$$

For those pixels, the intensity level of class C1 and C2 can be represented in interval, so for C1, the interval is [1, t] and for C2 the interval is [t + 1, i]; I represents the largest pixel whose values are 255. Then in mean can be calculated of C1 and C2 as follows

$$\mu_1(t) = \sum_{i=1}^{t} \frac{iP(i)}{q_1(t)} \tag{2.5}$$

$$\mu_2(t) = \sum_{i=t+1}^{i} \frac{iP(i)}{q_2(t)} \tag{2.6}$$

Afterward, variance of class C1 and C2 can be calculated as $\sigma_1^2(t)$ and $\sigma_2^2(t)$

$$\sigma_1^2(t) = \sum_{i=1}^{t} [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)} \tag{2.7}$$

$$\sigma_2^2(t) = \sum_{i=t+1}^{i} [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)} \tag{2.8}$$

Equations (2.7) and (2.8) describe the withings class's varianceing of C1 along with C2. Otsu algorithm tries theseing value can be minimized. The totalling varianceing knows how to compute by using the Eqs. (2.2), (2.7) and (2.8). The variance of image not depend on threshold is all the time stable significance so algorithm is reduce $\sigma_w^2(t)$ or maximize $\sigma_b^2(t)$

$$\sigma^2 = \sigma_w^2(t) + \sigma_b^2(t) \quad \text{Where}\ \sigma_b^2(t) = q_1(t)q_2(t)[\mu_1(t) - \mu_2(t)]^2 \tag{2.9}$$

2.1 Disadvantages of Otsu Algorithm

- In managing pictures broken on gray level, with conventional Otsu calculation, the acquired threshold could not locate a good combination to the global optimal.
- When identifying the target and backgrounds vary widely, Otsu calculation will not succeed properly.
- In Otsu calculation, the mean and fluctuation to locate the threshold value however Otsu calculation just gives the lower level threshold value so in MRI tumor can't visualize appropriately (Fig. 1).

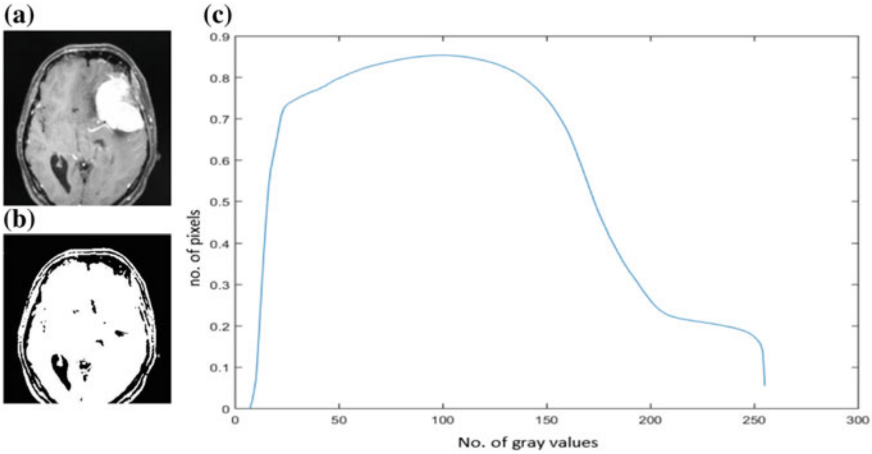


Fig. 1 Otsu algorithm result **a** original image; **b** Otsu algorithm result; **c** graphical representation of Otsu result where threshold value = 102

3 Proposed Methodology

The tumor area of MRI images can be segmented successive Otsu algorithm. The segmentation technique is higher PSNR when compared to other segmentation technique. Stability parameters used for MSER extraction determine the region between two levels which are "maximally steady". The tumor region is properly segmented we used to subtract the tumor region with its original image then again apply the successive Otsu algorithm.

3.1 Image Pre-processing

The visual appearance of information MRI pictures is made better. Image pre-preparing make three subcapacities: resizing of image and converting image into gray Scale. To begin with, resizing of the MRI image is done so that image can be enlarged. At that point, the gray scale transformation is obtained. After that noise is removed by using median filtering. This process is better because it filters the image without any modification in pixel intensity and features of an image. Therefore, whole noise of an image removes through filtering.

3.2 Successive Otsu Algorithm

The fundamental goal behind the division of the medical image is to isolate the tumor from the background [1]. This empowers the awareness of the picture toward the tumor, isolating the vital data from the background. Now for overcoming the Otsu problem, we can apply the successive Otsu for multiple hierarchical levels. In successive Otsu, we can divide the histogram peak into two parts, namely forward and backward direction. We can find the threshold in forward direction because we can find the higher-frequency information value which are not find in backward direction because it gives the lower frequency value [7, 8]. When we can apply successive Otsu in forward direction then by changing the slider location find the threshold values for each image at higher successive levels. Threshold value can be found till its successive level is not ended, or tumor can visualize it properly. In image tumor part separated from the background and visualizes properly then stops [9]. In a given image, let G(x, y) represents the gray level in the range of (0 to $R-1$). In this, R is the numbered of feasible dissimilar gray intensity. Then, apply the Otsu related segmentation method to describe a subroutine of OtsuAlgo:

$$\mathrm{Thresh} = \mathrm{OtsuAlgo}\ (\mathrm{h}, \mathrm{S} = \mathrm{begin}, \mathrm{E} = \mathrm{stop})$$

By using Otsu calculation, optimum threshold value can be calculated:

h histogram of an image
S image histogram of beginning value
E image histogram of last value

Let's in forward direction OtsuLast describe a subroutine for n successive segmentation and it can be signified by algorithm:

$$\text{FrwdThresh} = \text{OtsuLast}\,(\text{h}, \text{S}, \text{E}, \text{n})$$

Algorithm-1 For n successive segmentation in forward direction

```
FrwdThresh = OtsuLast (h, S, E, n)
S = 1; E = R; where (S = begin; E = stop)
For I = 1 to n
{
FrwdThresh (I) = OtsuAlgo (h, S = begin, E = stop);
S = FrwdThresh (I);
}
Stop = E
```

A vector representation describes by forward thresholding (FrwdThresh), the n consecutive segmentation for successive thresholding value. We can add the new parameter n to subroutine OtsuLast as contrast with OtsuAlgo. To calculate P, the earlier thresholding w.r.t. optimum thresholding similar method has been used. Let's define the OtsuPre subroutine in Algorithm-2 for execution of P value of successive histogram other than in the backward direction w.r.t. central optimum thresholding.

$$\text{PreThresh} = \text{OtsuPre}(\text{h}, \text{begin} = \text{S}, \text{stop} = \text{E}, \text{P})$$

Algorithm-2 Subroutine for P previous threshold values in backward direction with respect to central optimum threshold

```
PreThreshng = OtsuPre (h, S = begin, E = stop, P)
S = 1; E = R;
PresThresh (1) = OtsuAlgo (h, S, E);
For I = 2 to P
{
E = PreThresh (I − 1);
PreThresh (I) = OtsuAlgo (h, S, E);
}
Stop = E
```

Table 1 Threshold values of Fig. 2 images applied in forward direction using Algorithm-1

Otsu	Level-1	2	3	4	5	6	7	8	9	10
102	186	226	243	250	253	254	255	255	255	255

The vector PreThresh determines the P consecutive threshold valued in backward direction w.r.t. central optimum threshold. Once more, we can add the P new constraint to subroutine OtsuLast as contrast to OtsuAlgo. Therefore, executing the planned segmentation algorithm with parametric assessment n in forward direction as well as P in backward direction will outcome within $n + P - 1$ threshold values regarding the known histogram be able to subdivided (Tables 1 and 2).

3.3 Stability Parameters

A steady area has a little dissimilarity. The calculation discovers the area which is "maximally steady", implying that they have a lower dissimilarity than the areas one level below or above. Maximal stable extremal region is more suitable for object detection and tracking, as it detects the major MSER from both the depth image and the intensity image. But in our work, we can compare the level of segmentation between the two levels. For example, in Table 3, we can apply the level-1 and level-2, then by using the performance parameter, we can compare the segmentation level. Four stability parameters are used in our work namely

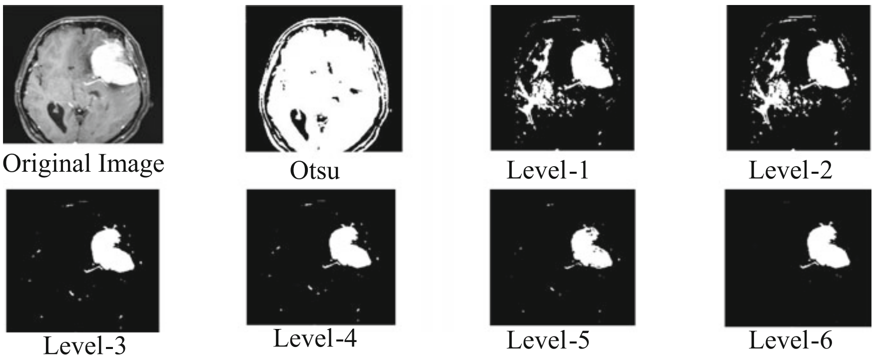


Fig. 2 Successive Otsu algorithm result

Table 2 Threshold values of Fig. 2 images applied in backward direction using Algorithm-2

Otsu	Level-1	2	3	4	5	6	7	8	9	10
102	43	22	15	11	9	8	7	NaN	–	–

Table 3 Stability parameter results of Fig. 2 at level-1–2, 3–4 and 5–6

PRI	GCE	VOI	PSNR
0.485136	0.344804	3.375443	4.058438
0.732321	0.118431	1.859812	11.64785
1	0	2.22E−16	65535

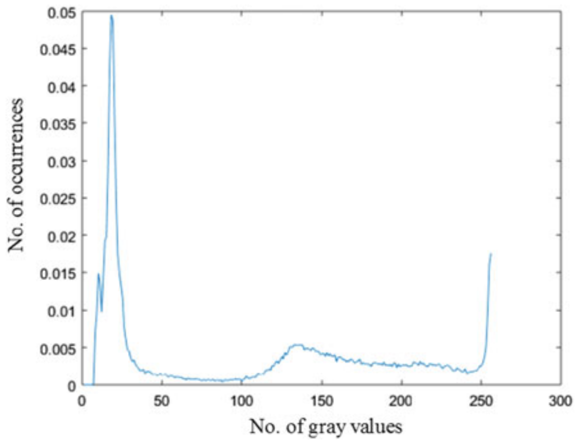


Fig. 3 Graphical representation of successive Otsu

randomness indexing, globally inconsistency errorless, variation of information, PSNR [10] (Fig. 3).

Now same as apply to all level and compare till the similarity in image can achieve; if we can get the similarity, then MSER region can achieved successfully. For example, we go to the higher successive levels and compare the level 5–6; now the comparison can stop at that level where the values of parameter achieve it successfully, and that region is called MSER region.

- **Probabilistic Random Index (PRI)** [11]—checks then division often sets often pixels who's marking aren't constant among the calculated division as well as then grounding fact, usually over numerous argument fact divisions toward represents the level variety. The Rand list lies among 0 as well as 1, 0 representing to facilitate then two fact groups don't concur on every match of pairs of point in addition 1 representing the information bunches be accurately similar. The random index (ri) is:

$$ri = \frac{p+q}{p+q+r+s} = \frac{p+q}{\frac{m}{2}} \tag{3.1}$$

Hence, in Eq. (3.1), p + q = number of conformity among x with y
r + s = number of dissimilarity among x with y.

- **Variation of information (VOI)** [12]—quantifies then entirety of data loss and achieve among the two module, as well as along these lines generally compute then degree toward single set be capable of clarify then other set. Then VOI metric is nonnegative; through lower value demonstrating well-known similarity. In this, where x = x1, x2, …, xk. Then, the VOI among two clustering is:

$$\mathrm{vi}(\mathrm{x};\mathrm{y}) = h(x) + h(y) - 2i(\mathrm{x},\mathrm{y}) \tag{3.2}$$

where in Eq. (3.2) h(x) is entropy often x and vi(x, y) is mutual information (MI) among x as well as y. The mutual information (MI) of two clustering is then loss often ambiguity often individual clustering if the other is known. Thus, MI is positive furthermore bounded by {h(x), h(y)} log2(n).

- **Global consistency error (GCE)** [13]—It is utilized to contrast the consequences of calculations with a database of segmented images. Individual segment is a suitable separation often other; at that point, pixel represents within a range often alteration, as well as the inaccuracy must be zero. The two areas cover in a conflicting way because here is no subset connection.

$$gce = \frac{1}{m}\min\left\{\sum_i e(S1,S2,PI),\ \sum_i e(S2,S1,PI)\right\} \tag{3.3}$$

- **Peak signal to noise ratio (PSNR)** [14]—is utilizes toward calculate then variation among two images, where b is the largest feasible significance often then indication (normally 255 or 1). The unit of PSNR is specified in decibel units (dB), which compute the fraction often then peak signal along with the dissimilarity among two images. Mean squared error and PSNR are defined in Eqs. (3.4) and (3.5):

$$mse = \frac{1}{pq}\sum_{i=0}^{p-1}\sum_{j=0}^{q-1}[\mathrm{a}(\mathrm{i},\mathrm{j}) - \mathrm{b}(\mathrm{i},\mathrm{j})]^2 \tag{3.4}$$

$$\begin{aligned} psnr &= 10\cdot\log_{10}\left(\frac{\max_I^2}{mse}\right) \\ &= 20\cdot\log_{10}\left(\frac{\max_I}{\sqrt{mse}}\right) = 20\cdot\log_{10}(max_I) - 10\cdot\log_{10}(mse) \end{aligned} \tag{3.5}$$

Here, maxI represents the maximum feasible pixel value often the image. When the pixels are signified with 8 bits per sample, this is 255. In Table 3 for comparing the segmentation at level 5–6, we can achieve the threshold value and segmentation results of MRI tumor images.

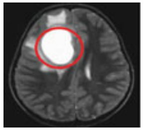
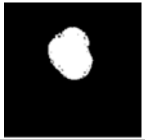
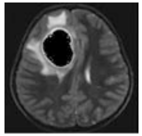
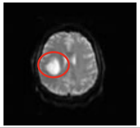
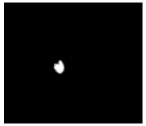
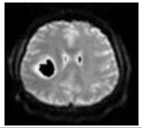
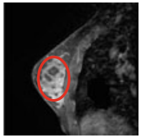
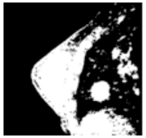
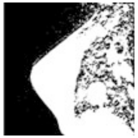
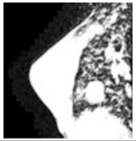
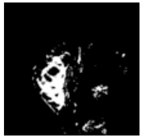
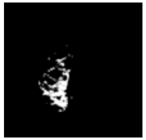
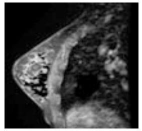
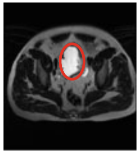
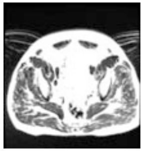
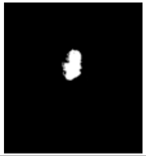
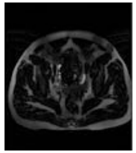
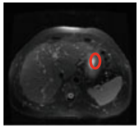
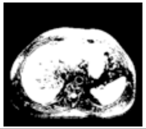
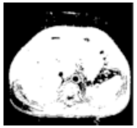
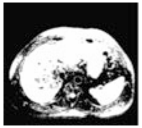
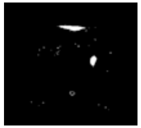
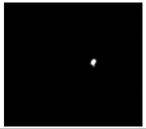
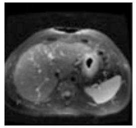
4 Implementation and Results Analysis

The MATLAB R2015a (version 8.5) is used for the simulation and implementation of the algorithm. This tool has been chosen for simulation because of this native support for image processing. The algorithm takes input the *dicom images generated for MRI scans. The standard database more than 50 images of brain, liver, breast and bladder are collected from TCIA [15] database has been used to train and verify the system. Many algorithms go through from the reality that there is no ordinary technique to find out the number of optimum thresholds. Later than applying the proposed algorithm in forward direction at hierarchical level, PSNR of threshold image has found to be saturate. This property can be used to obtain the appropriate number of thresholds. The execution time measurements for checking the feasibility of the strategy are picked as computational time so as to get a consideration of the complexity, while PSNR is utilized toward decide the value often threshold image. A segmented image having R gray levels utilized with n number of threshold value. Otsu's comprehensive follow technique looks for $R - 1$ Cn diverse arrangements of threshold values, for $n \ll R$. In any case, to discover threshold utilizing our successive Otsu calculation, we initially compute the mean of R gray levels which devour time O(R) that can follows the computation of fluctuation which again requires significant investment O(R). Because of this the time devoured at each level is additionally of O(R); note that threshold is figured at each progression. Accordingly the execution time taken for discovering n thresholds is constantly <O(nR) and thus the calculation is done in polynomial time. Comparisons between Otsu [6], entropy [16], FCM [17] and proposed approach gives the good segmentation results on MRI brain tumor images. Entropy segmentation as far as noise is superior to Otsu, but it demonstrates difference in provisions of maintenance of geometrical characteristics. It calculates its thresholding level at a lower level as contrast besides challenging methods. FCM algorithm is broadly chosen because often its extra flexibility which allows pixels go with several module by varying quantity. However then main running drawback of FCM technique is time consuming and sensitivity toward noise. We can apply our proposed algorithm on different tumor images dataset of MRI [15] e.g. brain, bladder, breast, liver in Table 5, now by considering the forward threshold value at level-5 and 6 we can calculate the comparing segmented result using four performance parameters in Table 4.

Table 4 Performance analyses of Table 5 images between level-5 and level-6

Images	PRI	GCE	VOI	PSNR	CPU time (ms)
(a)	1	0	−2.2E−16	65,535	23.167
(b)	0.9981	0.0010	0.0129	30.2749	33.140
(c)	1	0	4.72E−16	65,535	33.999
(d)	0.9958	0.0017	0.0251	31.8316	37.470
(e)	0.9988	4.3725e−04	0.0080	32.4010	35.587

Table 5 Comparisons between Otsu, entropy, fuzzy c-means based and proposed approach results on MRI brain tumor images

Images [17]	Original images	Otsu results	Entropy-based results	Fuzzy c-means results	Proposed approach		Subtract the tumor area
					Level-5	Level-6	
Image (a) brain tumor							
Image (b) brain tumor							
Image (c) breast tumor							
Image (d) bladder endothelial carcinoma							
Image (e) liver hepatocellular carcinoma							

Total no. of pixel	65536
No. of white pixel	3055
No. of black pixel	62481
Ratio is	0.0489
Area of tumor is	3.0736e+03

Fig. 4 Calculate the area of tumor

We contrast our planned process through previous methods, which including Otsu algorithm [6], entropy-based algorithm [16], and fuzzy c-means algorithm [17]. When identify the target and backgrounds vary widely, Otsu calculation will not succeed. It indicates that our algorithm gives the better segmented result on MRI images with high PSNR values. Moreover, although our proposed algorithm gives the little bit higher computational time as compare with Otsu algorithm, but its PSNR value is too low. Here, 256 $\times$ 256 jpg images are a maximum image size. The calculated tumor area is analysis on the basis of number of white pixels Fig. 4. Illustrate the final outcome for tumor area. The stage of tumor is based on the size of tumor.

4.1 Segmented Results

Figure 5 When we applying the proposed algorithm then successive Otsu segmentation results are obtained at hierarchical level image (b), now after this, we can apply the four parameters to know the stable region. In the proposed segmented image result at level-6, we can get the image (b) result and original result image (a) is used for subtraction of tumor part. In this subtracted resultant image (c), we can again apply the proposed algorithm (i.e., successive Otsu algorithm) and we can see the result in image (d), we have clearly seen that there is no tumor portion can be detected. At that point, we can say that tumor can be segmented properly and Otsu algorithm is also apply on image (c) then resultant image is (e), in this resultant image (e), we cannot see the tumor part clearly.

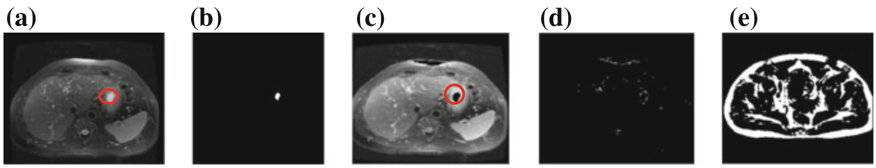


Fig. 5 Segmented results **a** original image; **b** segmented result; **c** subtracted result; **d** again applied proposed algorithm on subtracted image at level 3–4 (no tumor detected); **e** after applying Otsu on subtracted image

5 Conclusion and Future Work

In segmentation methods, thresholding is the easy and achieving highest speed technique for segment the foreground and background successfully. To alter the illustration of an image in easier way to inspect as well as to rearrange it properly is the main objective of segmentation. One important objective in tumor research study is to correctly assessment of growth area. In this paper, optimal hierarchical level images thresholding trouble we addressed the successive Otsu algorithm which is guided by Otsu. Stability based parameters is used such as probability random index, global consistency error, variation no of information, PSNR a new successive Otsu algorithm proposed to extend it into hierarchical levels to define the segmentation threshold for MRI image. While many of the pixels reduces in gray-level image, our algorithm generated higher peak signal to noise ratio in minimum levels, which give high gray-level intensity value to visualize the tumor properly. The major aims of image segmentation are to retrieve necessary information from the known image in a way that it will not affects the other features of that image.

In this study attempts have been made to reduce the time of image segmentation at hierarchical levels for developing diagnostic skills in medical imaging but this is just a beginning to make it more finer and effortless using dedicated algorithms. It will be effectively applied to training a huge database and video rescue method in real-time in future research. In future work, to expand the ability of the segment procedure applied the different segmentation techniques, for an intensity value modification method will give all the extra difficult as well as challenging features for us to filter the segment systems for MRI tumor image.

References

1. Singh, D.V.: Digital image processing with matlab and labview. Reed Elsevier India Private Limited (2013)
2. Bauer. S., Nolte, L., Reyes, M.: Segmentation of brain tumor images based on atlas-registration combined with a Markov-Random-Field lesion growth model In: 2011 IEEE International Symposium on Biomedical Imaging: from Nano to Macro. IEEE, pp. 2018–2021 (2011)
3. Mohsen, H., El-Dahshan, E.S.A., Salem, A.B.M.: A machine learning technique for MRI brain images. In: 2012 8th International Conference on Informatics and Systems (INFOS). IEEE, pp. BIO-161 (2012)
4. Yang, X., Shen, X., Long, J., Chen, H.: An improved median-based Otsu image thresholding algorithm. In: 2012 AASRI Procedia Conference on Modelling, Identification and Control. Elsevier, pp. 468–473 (2012)
5. Cai, H., Yang, Z., Cao, X., Xia, W., Xu, X.: A new iterative triclass thresholding technique in image segmentation. In: IEEE Transactions on Image Processing. IEEE, pp. 1038–1046 (2014)
6. Otsu, N.: A threshold selection method from gray-level histograms. In: IEEE Transactions on Systems, Man, and Cybernetics. IEEE, pp. 62–66 (1979)

7. Huang, M., Yu, W., Zhu, D.: An improved image segmentation algorithm based on the Otsu method. In: 2012 13th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel & Distributed Computing (SNPD). IEEE, pp. 135–139 (2012)
8. Chen, Q., Zhao, L., Lu, J., Kuang, G., Wang, N., Jiang, Y.: Modified two-dimensional Otsu image segmentation algorithm and fast realisation. IET Image Proc. **6**(4), 426–433 (2012)
9. Gautam, D., Ahmed, M.: Efficient fuzzy edge detection using successive Otsu's method. In: 2014 International Conference on Convergence of Technology (I2CT). IEEE, pp. 1–5 (2014)
10. Nistér, D., Stewénius, H.: Linear time maximally stable extremal regions. In: Computer Vision–ECCV 2008. Springer, pp. 183–196 (2008)
11. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE, pp. 929–944 (2007)
12. Meilă, M.: Comparing clusterings: an axiomatic view. In: Proceedings of the 22nd International Conference on Machine Learning. ACM, pp. 577–584 (2005)
13. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Eighth IEEE International Conference on Computer Vision (ICCV). IEEE, pp. 416–423 (2001)
14. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electr Lett (IET) **44**(13), 800–801 (2008)
15. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Tarbox, L.: The cancer imaging archive (TCIA): maintaining and operating a public information repository. J. Dig. Imaging **26**(6):1045–1057 (2013)
16. Chaudhari, A.K., Kulkarni, J.V.: Local entropy based brain MR image segmentation. In: 2013 IEEE 3rd International Advance Computing Conference (IACC). IEEE, pp. 1229–1233 (2013)
17. Szilágyi, L., Szilágyi, S.M., Benyó, Z.: A modified fuzzy c-means algorithm for MR brain image segmentation. In: International Conference on Image Analysis and Recognition. Springer, Berlin, Heidelberg, pp. 866–877 (2007)

Synthesis of Reversible Array Divider Circuit

Somnath Rakshit, Sraboni Mondal, Aruna Chakraborty, Apurba Sarkar and Dipak K. Kole

Abstract Progress made in quantum computation has resulted in logic synthesis with reversible circuits receiving considerable interest. This paper presents a synthesis of reversible array divider circuit based on non-restoring division algorithm implemented with k-CNOT gates. The execution time of such an array divider is low. Our aim is to propose a reversible array divider circuit which is efficient as regards to the number of garbage outputs and gates.

Keywords Divider · Array divider · Reversible synthesis · Reversible circuit Quantum computation

1 Introduction

One of the major goals in the field of digital circuit design and synthesis is the reduction of energy demand. Landauer [1, 2] observed that information loss and inherent energy dissipation take place in a circuit when traditional (irreversible) logic gates

S. Rakshit (✉) · S. Mondal · D. K. Kole
Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India
e-mail: somnath@cse.jgec.ac.in

S. Mondal
e-mail: sm1921@cse.jgec.ac.in

D. K. Kole
e-mail: dipak.kole@cse.jgec.ac.in

A. Chakraborty
Department of Computer Science and Engineering, St. Thomas College of Engineering and Technology, Kolkata, West Bengal, India
e-mail: aruna.stcet@gmail.com

A. Sarkar
Department of Computer Science and Technology, Indian Institute of Engineering Science and Technology, Shibpur, Shibpur, West Bengal, India
e-mail: as.besu@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_68

are used, regardless of its implementation. A system that does not lose any information is said to be reversible. According to Bennett [3, 4], zero-energy dissipation would occur mainly by the network when it comprises only reversible gates. Thus, in future circuit design, reversibility will play a very important role. Over the last decade, significant energy loss has been achieved with irreversible gates by means of advanced process technologies, higher levels of integration and low power-consuming design methods and tools. Thus, traditional procedures of low power design are more likely to prevail. However, if Moore's Law [5] followed in the field of growth of IC technology, then energy loss in irreversible design may become more dominant in the next decade. Also, it has been shown that reversible logic can have immense application in the relatively unexplored field of quantum computation [6–8]. This has the potential for solving exponentially hard problems in polynomial time. Reversible logic is mainly used to construct quantum circuits. The concept of reversibility is also used in adiabatic CMOS [9], optical logic [10, 11], thermodynamic technology [12], nanotechnology [13, 14] and DNA technology [11]. We know that the traditional logic gates like AND gate, OR gate or EXOR gate, although extensively used in digital design, are not reversible in nature. NOT gate is the only gate that is reversible. However, only reversible gates can be used for the purpose of designing a circuit. Proposal for many such gates has been made over the past decades like controlled-NOT (CNOT) which was proposed by Feynman [15], Toffoli [16], and Fredkin [17] gates. Also, fanout and feedback should be absent in a reversible circuit. This restriction conforms to the design requirement of quantum networks [6]. In traditional methods, the number of gates is used as a parameter to estimate the complexity of a circuit. However, in the case of reversible logic, the number of garbage outputs, which is an important parameter, should also be considered. Since reversible gates are used in reversible logic design methods and the number of inputs and the number of outputs are equal, in such a network, the total number of inputs is equal to the number of outputs. Though the synthesis of reversible adders using various types of reversible gates has appeared in [18], the reversible circuit synthesis for other arithmetic operations is an extremely interesting open problem.

This paper presents a reversible synthesis of array divider. The proposed circuit can divide a whole number by another whole number. It has been shown that the proposed circuit generates the reversible array divider having minimum number of garbage outputs. Also, the array divider tends to have reduced computational time at the expense of increased hardware complexity.

2 Preliminaries

All division algorithms can be implemented using an array of cells in such a way, where each step of the algorithm is computed by a separate row of cells. To implement a radix-2 division algorithm, n rows of cells with n cells per row are required. If the selected division algorithm is restoring algorithm, then the difference in each row

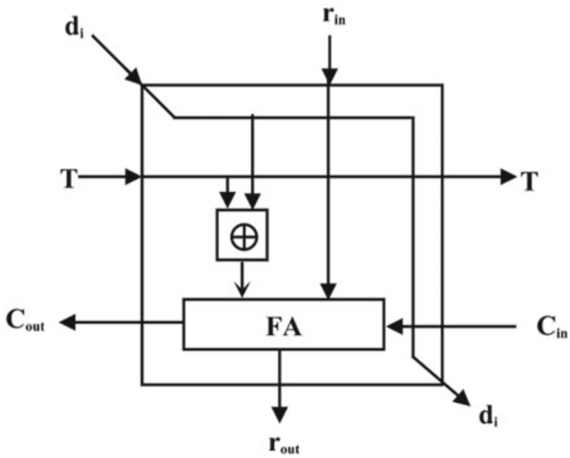


Fig. 1 Controlled add/subtract (CAS) cell

between the previous partial remainder and the divisor is generated and the quotient bit is generated according to the sign of this difference. There is no need to restore the partial remainder if the quotient bit is determined to be 0. We consider a non-restoring division array which has about the same speed as the restoring array. The ability to handle negative operands in a simple way is the only advantage of using a non-restoring array. An example of such an array divider is shown in Fig. 2 where $x_0x_1x_2x_3x_4x_5x_6$ is the dividend, $d_0d_1d_2d_3$ is the divisor, $q_0q_1q_2q_3$ is the quotient, and $r_0r_1r_2r_3$ is the final remainder. The signal T, which is shown in Fig. 1, controls the

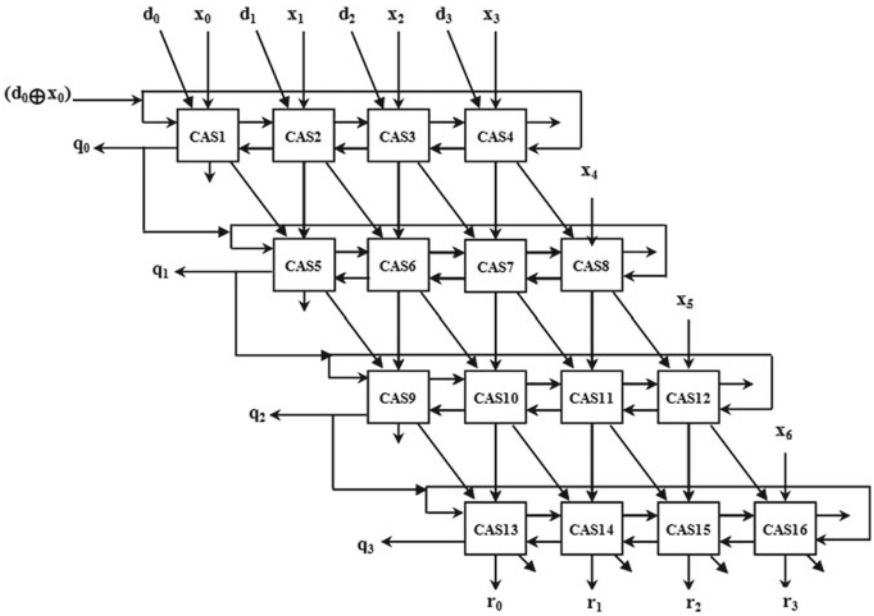


Fig. 2 Non-restoring array divider

operation (addition or subtraction) to be performed in a given row. When $T = 0$, an addition is performed, while if $T = 1$, a subtraction is performed by adding the two's complement of the divisor, which is assumed to be positive (i.e. $d_0 = 0$). The latter is done forming the one's complement of the divisor and forcing a carry of 1 into the rightmost cells by connecting it to T. The generated quotient and remainder are represented in two's complement. The quotient generated by the array divider in Fig. 1 is correct because in every step of the non-restoring division, the partial remainder and the multiple of the divisor ($\pm D$) have opposite signs always. Hence, the carry out from the leftmost cell is equal to 1 if the sign of the new partial product is 0 and vice versa. Hence, $C_{out} = 1$ implies that the operation in the next row should be subtraction ($T = 1$), since the divisor is assumed to be positive. Similarly, $C_{out} = 0$ generates $T = 0$, so an addition operation should be performed in the next row.

3 Reversible Synthesis of Array Divider

In this section, we propose the reversible synthesis of the array divider. To design the reversible non-restoring array divider, we need to design the reversible CAS cell circuit using only k-CNOT gates. In CAS cell circuit, the expressions for the output functions are given below.

$$r_{out} = (d_i \oplus T) \oplus r_{in} \oplus C_{in} \tag{1}$$

$$C_{out} = (d_i \oplus T) \cdot r_{in} + (d_i \oplus T) \cdot C_{in} + r_{in} \cdot C_{in} \tag{2}$$

From the expression of the r_{out} output function, we can easily map it in the reversible domain. Our aim is to express the C_{out} output function as a sum of exclusive-OR functions. The sum-of-products (SOP) expression for the output function Cout depends on the values of d_i, T, r_{in} and C_{in}. To achieve reversible synthesis of C_{out}, we derive the sum-of-EXOR expression for C_{out}.

$$C_{out} = (d_i \oplus T) \cdot r_{in} + (d_i \oplus T) \cdot C_{in} + r_{in} \cdot C_{in} \tag{3}$$

$$C_{out} = (d_i \oplus T) \cdot r_{in} \oplus (d_i \oplus T) \cdot C_{in} \oplus r_{in} \cdot C_{in} \tag{4}$$

since $a \cdot b + b \cdot c + c \cdot a = a \cdot b \oplus b \cdot c \oplus c \cdot a$, which is proved. To achieve the reversible synthesis of the CAS cell circuit, we need only one constant input 0 and one extra garbage output. The block diagrams of CAS cell and the reversible CAS cell using only k-CNOT gates are shown in Figs. 3 and 4. Also, the block diagrams of COPY circuit and the reversible COPY circuit are shown in Figs. 5 and 6.

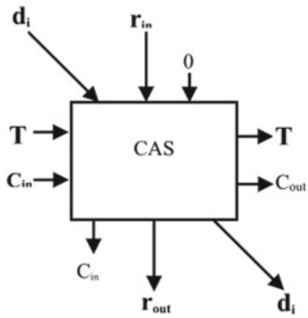


Fig. 3 Block diagram of CAS circuit

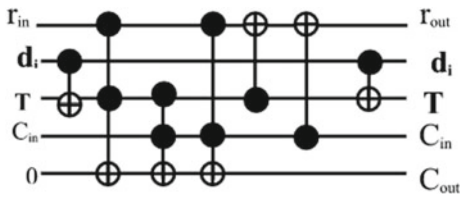


Fig. 4 Reversible CAS circuit

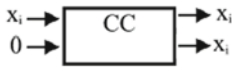


Fig. 5 Block diagram of COPY circuit

Fig. 6 Reversible COPY circuit

4 Results and Discussion

4.1 Analysis of Our Proposed Reversible Array Divider Circuit

Our proposed reversible array divider circuit generates less number of garbage outputs and uses less number of gates. We have shown the result of our proposed array divider circuit in Table 1. Our proposed design requires 24 gates (Fig. 7).

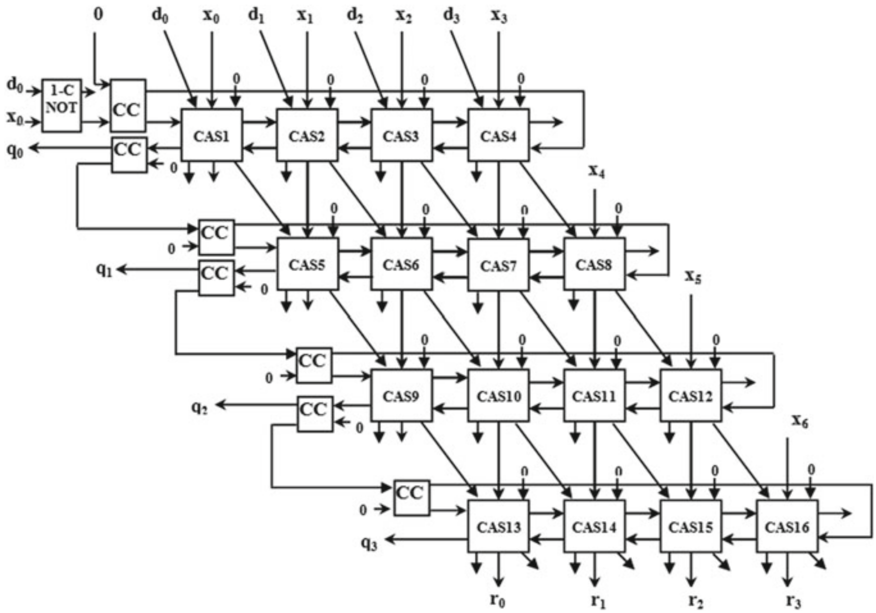


Fig. 7 Reversible non-restoring array divider

Table 1 Result of reversible array divider circuit

	No. of gates	No. of garbage outputs	Total no. of logical computations
Our proposed array divider circuit	$16\,CAS + 7\,CC + 1\,NOT = 24$	37	$103a + 48\beta$

Let,
a = A two-input EXOR gate computation
d = A NOT computation
T = Total logical computations
β = A two-input AND gate computation.

The number of output lines which are not used as a primary output or not used as an input to other gates is considered as garbage output lines. The major challenge in reversible logic synthesis is to reduce the number of garbage outputs. Thirty-seven garbage output lines are produced by our proposed reversible array divider circuit.

5 Conclusion

In this work, we have proposed a method for reversible synthesis of array divider based on non-restoring division algorithm implemented with k-CNOT gates and a reversible array divider circuit which is efficient with respect to number of gates and garbage outputs. Reversible array divider can have wide ranged applications in various fields like computer security and transaction processing, low power CMOS, quantum computation and nanotechnology.

References

1. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**(3), 183–191 (1961)
2. Robert, W.K., Landauer, R.: Minimal energy dissipation in logic. IBM J. Res. Dev. **14**(2), 152–157 (1970)
3. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**(6), 525–532 (1973)
4. Bennett, C.H.: Notes on the history of reversible computation. IBM J. Res. Dev. **44**(1/2), 270 (2000)
5. Moore, G.: Cramming more components onto integrated circuits. Electronics **38**(8) (1965)
6. Nielsen, M.A., Chuang, I.: Quantum Computation and Quantum Information (2002)
7. Gershenfeld, N., Chuang, I.L.: Quantum computing with molecules. Sci. Am. **278**(6), 66–71 (1998)
8. Preskill, J.: Lecture Notes in Quantum Computing (1998)
9. Athas, W.C., Svensson, L.J.: Reversible logic issues in adiabatic CMOS. In: 1994 Proceedings of Workshop on Physics and Computation, PhysComp'94, pp. 111–118. IEEE (1994)
10. Picton, P.: Optoelectronic multi-valued conservative logic. Int. J. Opt. Comput. **2**, 19–29 (1991)
11. Picton, P.: A universal architecture for multiple-valued reversible logic. MVL J. **5**, 27–37 (2000)
12. Merkle, R.C., Drexler, K.E.: Helical logic. Nanotechnology **7**(4), 325 (1996)
13. Merkle, R.C.: Reversible electronic logic using switches. Nanotechnology **4**(1), 21 (1993)
14. Merkle, R.C.: Two types of mechanical reversible logic. Nanotechnology **4**(2), 114 (1993)
15. Feynman, R.P.: Quantum mechanical computers. Optics news **11**(2), 11–20, 1985
16. Toffoli, T.: Reversible computing. In: Automata, Languages and Programming, pp. 632–644 (1980)
17. Fredkin, E., Tommaso, T.: Conservative logic. Int. J. Theor. Phys. **21**(3), 219–253 (1982)
18. Maslov, D., Dueck, G.W.: Garbage in reversible design of multiple output functions. In: 6th International Symposium on Representations and Methodology of Future Computing Technologies, pp. 162–170 (2003)

Finding Right Doctors and Hospitals: A Personalized Health Recommender

M. Sridevi and R. Rajeshwara Rao

Abstract In today's era, most of the population rely on the Internet to know about health information by accessing several online health portals, videos, articles, and blogs regarding diseases, diagnoses, and treatments published by hospitals, medical associations, etc. This change has been observed throughout the globe. However, it proved to be non-personalized and challenging in finding relevant information that resulted in taking adequate decisions regarding personal health status. Therefore, a personalized healthcare recommender system provided a solution by recommending the best services as per the user needs. This paper proposes an improved recommender in health care by exploiting user demographic features, similar users predicted rating, rating based on total number of reviews, and ratings based on average ratings using collaborative filtering. The proposed system is evaluated using generic data and is found to provide accurate recommendations, and the quality of the recommendation is improved by keeping our system up to date.

Keywords Health care · Recommender system · Collaborative filtering

1 Introduction

The tremendous growth on the Web resulted in a profound impact not only on e-commerce [1] Web sites but also on health information systems [2]. It has posed a great challenge of exposing users with more number of options. Finding the right choice that meets their requirements is hardly acceptable. In the modern era, technology has reduced the barriers of information overload dramatically by providing only the information which is most valuable to us. Recommender systems

M. Sridevi
Anurag Group of Institutions, Hyderabad, Telangana, India
e-mail: siri_235@yahoo.com

R. Rajeshwara Rao (✉)
JNTU, Kakinada, Kakinada, Andhra Pradesh, India
e-mail: raob4u@yahoo.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_69

[3, 4] is one of the most promising technologies in several applications like information filtering, e-commerce, health information systems. The demand for personalized real-time recommender system is truly required whose main aim is to filter the relevant information to the users which finds interesting to them.

It is said that health is wealth; thus, our proposed system helps in finding appropriate doctors and hospital for the needful users to diagnose and treat them. The system is overloaded with all kinds of doctors and hospitals from different sources. But only a subset of doctors and hospitals are filtered for a given health issue. At this point, our system learns the ratings of filtered doctors and hospitals given by the users interests and thus serving as a decision maker.

In real-world scenarios, a person seeks friend's, family, or a neighbor's opinion when making decision about the doctors and hospitals for their treatment of diseases. We will often turn to our close friend's opinion by assuming they are trustworthy. On the other hand, there is less likelihood of having same medical history with that of friends. Therefore, in our proposed system users seek public preferences and their ratings on doctors and hospitals.

Health recommender systems provided a solution by recommending the best services as per the patient needs. Users, when faced with vast amount of medical data available on various Web sources, will be in quandary state. Therefore, a personalized healthcare recommender system guides the users in making decisions. The proposed recommender system is patient-centric, providing laymen friendly interaction and delivering highly relevant data suitable for the user profile. The goal of the proposed system is to gather, disseminate the information and knowledge to empower various stakeholders like patients, healthcare providers, and lead them toward patient-centric environment.

The paper is organized as follows: In Sect. 2, we brief about related work, and Sect. 3 presents an overview of the proposed approach. In Sect. 4, our approach implementation is presented, and finally, in Sect. 5 we conclude our paper by giving directions for future work.

2 Related Research

2.1 Personal Health Record (PHR)

Today, patients are changing their way of engaging with healthcare providers to maintain their health conditions. Personalized health record (PHR) system [5] is used to gather and keep track of the user's health. It allows the data to be shared between the patients and healthcare providers. It is user friendly and provides online health solutions and empowers their ability to learn more about their own healthcare records.

PHR allows patients to coordinate with several care providers by facilitating continuous care. PHR enables to improve the quality and speed of communication between care providers and users, tracking the services provided and be informed of medications, pharmacy, allergies, etc.

In the modern era, patients are no longer passive in the medical world. They get actively involved in their medical decisions. This motivated our proposed system to personalize medical decisions. Though there exist several techniques and methodologies whose roots are from information retrieval system and recommender systems, we believe that our proposed approach could benefit both patient care provider and could be a value-added system to existing system.

2.2 Electronic Health Record (EHR)/Electronic Medical Record (EMR)

Advantages in technology had numerous advantages like managing large amount of patient's health-related data like data analysis, clinical statistics data, and reuse of health data. Though EHR, EMR, and PHR are used interchangeably, there are some important differences between them.

An electronic health record [6] is an electronic record of medical care created and managed by healthcare organizations. All authorized clinicians access information of the patient. EHR consists of the following elements:

Health data: It holds the patient's problems, medication lists, and results of the test.

Result data: Laboratory results, reports, and X-ray images.

Appointments: Patient can schedule appointment.

Reporting database: It holds the database of the patients.

Patient support: Online questionnaire can be posted and can receive educational material.

Electronic medical record (EMR) [7] is an electronic record of health information of a user created and managed by the clinicians within a healthcare organization. It is considered as healthcare organization-centric. The factors like customization of data, sharing the information, compatibility between different hospital work cultures have made the adoption of EHR a serious issue in the current health organizations.

2.3 Recommender Systems

Recommender systems reduce the complexity of making decision. Users can refer to recommender system if they find trustworthy. Satisfaction is one of the key factors in building user's trust in the system. It can be measured directly from the user feedback. Lee et al. [8] developed an application that recommends medical products as per their health records. Huang et al. [9] derived an optimal longtime treatment plan by analyzing patient similarity. A recommender system is said to be efficient when it helps in making good decisions faster. Recommender systems can be categorized into collaborative filtering recommender system [10], content-based recommender system [11], and hybrid-based recommender system [12]. Among the existing techniques, collaborative filtering is one of the widespread successful recommender systems in the Web.

Collaborative Filtering. Collaborative filtering, the traditional recommender system, is based on rating structure usually represented as user-item rating matrix. Each cell value represents the rating of an item by the user. It predicts the ratings based on similarity measures like Pearson correlation coefficient, cosine similarity, Euclidean distance measure. Collaborative filtering can be classified into two types: memory-based CF and model-based CF [13]. Memory-based CF predicts the ratings using the entire user-item database of users who are similar to the active user, whereas model-based CF predicts the ratings by using the constructed model. The key challenges of collaborative filtering are to handle large number of users and items, deal sparse data and cold start issue.

Content-Based Recommender System. Content-based recommendations are based on the user individual preferences and tastes. It recommends the movies preferred by the user in the past. Content-based recommender [14] system often suffers from the following issues:

Limited content analysis It is difficult to recommend if there is a limited content about the user profile.

Overspecialization restricts users to items similar to the ones defined in their respective profiles, and thus new items and other options are not discovered.

Hybrid Recommender System. Many recommender systems use either collaborative filtering or content-based approach. Each of these approaches has its own strengths and weaknesses. In order to improve the performance of recommendations, collaborative- and content-based methods have been combined to build hybrid recommenders [12].

Hybrid recommendation incorporates multiple techniques to reduce certain limitations of each of them and to improve the quality and performance of the recommendations. It helps to overcome some of the shortcomings of collaborative filtering and content-based filtering, namely first-rater problem, tackling the sparsity of data, handling scalability, and overspecialization. However, hybrid recommender systems suffer from changing user preferences. As people’s tastes and preferences are subjected to change, it should deal with changing user preferences.

Eysenbach [15] known for e-health and consumer health informatics initiated the health information through online and Internet technologies. E-health characterizes thinking globally to improve health by utilizing communication and information technology. *MedWatch* [16] uses public databases available online and disseminates safety information to promote the health of the public. Additionally, MedWatch offers information on drugs, changes in labeling of drugs, etc., through different delivery methods which include email, RSS, Web site which will be updated at regular intervals. *PatientLikeMe* [17] enables patients to share healthcare experiences with other users, care providers, pharmacologists, etc., sharing the data, and makes the change possible by fixing and helping to promote better treatments. The patients experience when shared with industry helps to develop better products and speed up the care and services.

3 Proposed Methodology

Personalization of healthcare systems increases the patient satisfaction and better understands their medical condition and requirements. The proposed system provides a laymen friendly interaction for the users by tailoring their personal health records. It empowers the user by guiding the user in their lifestyle decision making, interactively using similar user's opinion. Though the sources of obtaining information about the patient are extensively growing, there is a risk of lacking in quality and relevance. It is important to create a relevant recommender system. Thus, our proposed system considers ratings of similar users and their predictions to increase the relevance and accuracy of recommendations. The aim of the proposed system is to improve the quality of life and help patients in their decision making about health care and reduce information overload of doctors and hospitals.

The intention of people to search for hospitals and doctors is often influenced by the services provided and can be effectively served by the health recommender systems. Majority of the users are concerned about the quality of the health recommender systems. The proposed system considers the consumer's usefulness and recommends the hospitals and doctors.

Initially, the patient when experiencing a health issue will be too weak in decision making of doctor's and hospital's choice. Though recently, the improved digital information has changed drastically but is often not personalized. Thus, a personalized recommender system is required that supports users in finding relevant medical content.

A framework is proposed that enables users to enter ratings based on their experience to doctors and hospitals and receive reliable recommendations for their conditions from the recommender system. Our proposed system is realized using collaborative filtering technique in order to derive benefits. The proposed framework consists of three important phases: data source phase, preprocessing phase, similarity computation, and finally recommendation phase as depicted in Fig. 1.

3.1 Data Source

This phase considers the input from the patients in the form of user profile. To make the system as patient-oriented, the input is considered from various sources that include demographic data and health parameters as shown in Table 1. Demographic attributes of a user and health factors play an important role for the user's health, and user rating data on doctors and hospitals expresses the reviews on them which

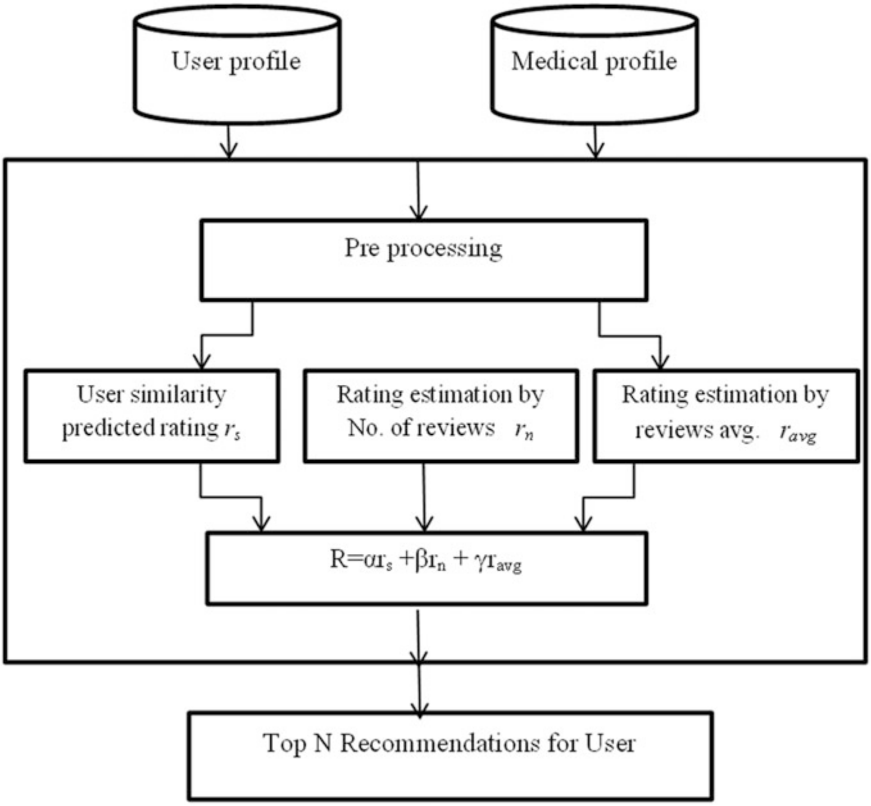


Fig. 1 Framework of the proposed system

Table 1 Sample user demographic data

Uid	Name	Gen	Age	Marital status	Blood group	Surgical history
u_1	John	M	21	S	O+	None
u_2	Jose	F	24	S	B	None
u_3	Robin	M	35	M	O+	None
u_4	Jenni	F	38	M	A+	None
u_5	Jacob	M	20	S	B	None
u_6	Cherry	F	40	M	O+	None

enables the system to enhance and provide a personalized health recommender. The user provides input in the form of explicit rating to the subset of hospitals and doctors provided in the list, thereby creating a custom taste profile as shown in Table 2. Input ratings are represented in the form of user–hospital/user–doctor rating matrix where users and hospitals are represented in row and column,

Table 2 User rating matrix

u	h				
	h_1	h_2	h_3	h_4	h_5
u_1	3	4	4	5	5
u_2	4	5	?	3	?
u_3	3	5	5	1	4
u_4	5	?	4	?	3
u_5	5	4	4	4	5

respectively, and the entry represents the ratings given by the user 'u' to the hospital or doctor. Thus, the system continuously learns the user preferences and generates accurate recommendations.

3.2 Preprocessing Phase

The demographic data and health issues are used by the system to filter the doctors and hospitals list. The filtered data helps to get faster recommendations. Preprocessing phase after collecting the input data filters to provide records of the doctor name and hospital name based on user's choice via the user's profiles. The dependency between user, doctor, and hospital is to be determined by means of health issue concerned with user. At this point, collaborative filtering technique is not used as they need user rating data.

3.3 User–User Similarity Computation

In this phase, users similar to the active user are selected as per the health history and rating data. User rating data on hospitals and doctors is analyzed to find the users similar to the active user. Once the user profile is constructed, the similarity between users is computed by using similarity measure such as Pearson correlation coefficient. Similarity between two users is given by the following equation:

$$s(u,a) = \frac{\sum_{(i=1)}^{n} (R_{u,i} - R_{\overline{u}})(R_{a,i} - \overline{R_a})}{\sqrt{\sum_{i=1}^{n} (Ru_i - R_{\overline{u}})^2 \sum_{i=1}^{n} (R_{a,i} - \overline{R_a})^2}} \tag{1}$$

where $R_{u,i}$ is the rating of user u on hospital/doctor i, $R_{a,i}$ is the rating of target of user a on hospital/doctor i, $R_{\overline{u}}$ and $R_{\overline{a}}$ are the average ratings of user 'u', and target user 'a' and 'n' are the number of hospitals/doctors rated by 'u' and 'a'. The proposed system assumes that users with similar rating data have a high probability of recommending the resources accurately. The system is provided with the user's current health condition and requires the system to recommend the doctor and hospitals for the patient to improve their health conditions.

3.4 Recommendation Phase

In this phase, the combined ratings of similar users identified in previous phase, ratings based on the numbers of reviews, and rating based on average ratings of hospitals/doctors are used to provide recommendations. The goal of the proposed system is to design health system that learns data from user profile and guide him in finding doctors and hospitals as per user's request. Similar user's predictions for hospitals and doctors are generated for the target users that are not rated based on their similarity score and their ratings. Similar user's prediction is given by the following equation:

$$r_s = \hat{R}a + \frac{\sum_{u \in S(u)} s(u,a) * \left(R_{u,h} - \widehat{R_u}\right)}{\sum_{u \in S(u)} s(u,a)} \tag{2}$$

where $S(u)$ is the set of users similar to target user a, and r_s is the predicted rating for target user a for hospital h or doctor d. The weightage on the number of ratings based on the number of reviews is given as mentioned in Table 3.

The combined rating is the weighted solution of similar users ratings r_s, number of reviews r_n, and average ratings r_a, i.e.,

$$R = \alpha * r_s + \beta * r_n + \gamma * r_a \tag{3}$$

where

R Resultant rating for target user
r_s Predicted rating based on user similarity
r_n Rating based on number of reviews
r_a Rating based on average

α, β, and γ are the weights of r_s, r_n, and r_a, respectively, and the values of α, β, and γ are given as

$$\alpha + \beta + \gamma = 1 \tag{4}$$

Table 3 Weightage on number of votes

Ratings	Weight		
	Ifvotes <= 100	Ifvotes > 100 && votes <=1000	Ifvotes > 1000
5	5	10	15
4	4	8	12
3	3	6	9
2	2	4	6
1	1	2	3

4 Proposed System Implementation

In the proposed health recommender, the system employs collaborative filtering and computes recommendations for the users. In our domain, initially, for every health issue doctors and hospitals are grouped and are available to the user. The proposed system is simple and interactive by using PHP and Java script. Figure 2 shows the Web interface of our proposed system. The system to perform efficiently, all the information is available in the user profile and medical database which contains information about the hospitals, doctors, etc.

Initially, the user registers by providing the information such as age, gender, and health parameters like blood group, surgical history. When a user logs into our system, he/she is given the option to input values. Based on the input, database is searched and a suitable list of hospitals and doctors based on health issues are filtered. The user then gives ratings to the known physicians/hospitals. The rating reflects patient satisfaction with the doctor and hospital. The rating varies from 1 to 5 (1—low, 5—high). The system then computes similar users. Finally to recommend the top-N hospitals/doctors to the target user, the combined ratings of similar users predicted rating, rating based on number of reviews for each hospital/doctor, and the average ratings of hospitals/doctors are considered.

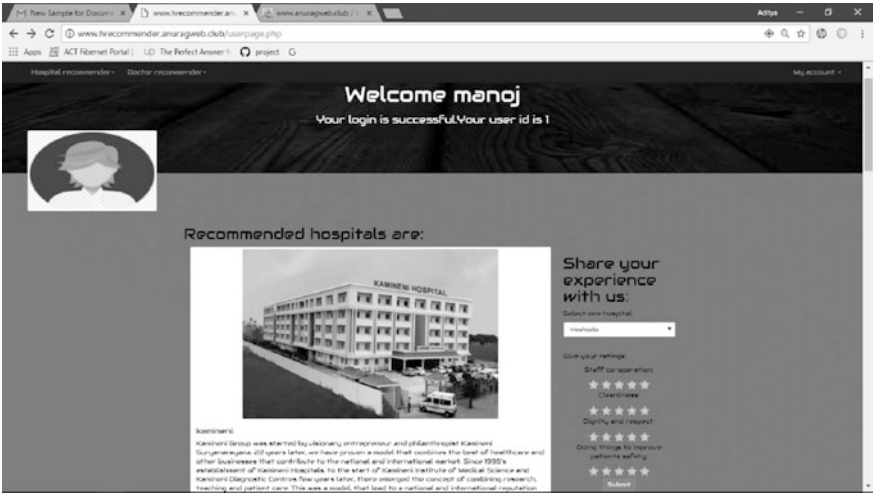


Fig. 2 Web interface for recommending hospitals

5 Conclusion

This paper has proposed a personalized tailored data regarding the healthcare providers and hospitals in a simple, personalized, and organized way to the users by helping them in controlling their health issues in time. The user is provided the option of building his profile with input rating, and then, the system recommends him a list of top-rated hospitals/doctors using collaborative filtering. A combined approach is proposed by considering the demographic data, similar users rating prediction, rating based on total number of reviews, and ratings based on average ratings help to solve cold start new user problem.

Our system provides accurate recommendation, and the quality of recommendation is improved by keeping our system up to date. The experiments conducted for a small subset of users have proven to be positive. In future, larger data set of users and the geographical distance between the users and hospitals will be considered to recommend and improve the performance. Additionally, we are presently working with real-world patient data and to evaluate the system performance.

References

1. Schafer, J.B., Konstan, J., Riedl, J.: Recommender systems in e-commerce. In: Proceedings of 1st ACM Conference on Ecommerce, 1999, pp. 158–166
2. Ballegaard, S.A., Hansen, T.R., Kyng, M.: Healthcare in everyday life: designing healthcare services for daily life. In: Proceedings of Conference on Human Factors in Computing Systems, USA, ACM, pp. 1807–1816 (2008)
3. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state of the art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
4. Bobadilla, J., Ortega, F., Hernando, A., Gutierrez, A.: Recommender systems survey. Knowl. Based Syst. 109–132 (2013)
5. Tang, P.C., Ash, J.S., Bates, D.W., Overhage, J.M., Sands, D.Z.: Personal health records: definitions, benefits and strategies for overcoming barriers to adoption. J. Am. Med. Inf. Assoc. 121–126 (2006), DOI: 10.1197
6. Health Informatics-Electronic Health Record—Definition, Scope and Context, International Organization for Standardization (2005)
7. Baird, A., North, F., Raghu, T.: Personal health records (PHR) and the future of the physician-patient relationship. In: ACM Conference, pp. 281–288 (2011)
8. Lee, C., Lee, M., Han, D.A.: Framework for personalized healthcare service recommendation health. In: Proceedings of 10th International Conference on E-Health Networking, Applications and Services, pp. 90–95 (2008)
9. Huang, Z., Lu, X., Duan, H., Zhao, C.: Collaboration based medical knowledge recommendation. Artif. Intell. Med. 13–24 (2011), DOI: 10.1016
10. Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. The Adapt Web, 291–324, 2007
11. Balabanovic, M., Shoham, Y., FAB: content based, collaborative recommendation, communications of the ACM, pp. 66–72 (1997)

12. Hu, R., Lu, Y.: A hybrid user and item-based collaborative filtering with smoothing on sparse data. In: 16th International Conference on Artificial Reality and Telexistence-Workshops ICAT'06, 2006, pp. 184–189. IEEE, Nov 2006
13. Cacheda, F., Carneiro, V., Fernández, D., Formoso, V.: Comparison of collaborative filtering algorithms: limitations of current techniques and proposals for scalable, high-performance recommender systems. ACM Trans. Web (TWEB) **5**(1) (2011)
14. Pazzani, M.J.: A framework for collaborative, content based and demographic filtering. Artif. Intell. Rev. 393–408 (1999)
15. Eysenbach, G.: Design and evaluation of consumer health information web sites. Consum. Health Inf. (2005). https://doi.org/10.1007/0-387-27652-14
16. http://www.fda.gov/Safety/MedWatch
17. https://www.patientslikeme.com

An Improved Hybrid Offset-Based Min-Sum Decoding Algorithm

Michaelraj Kingston Roberts and Elizabeth Sunny

Abstract In this work, a new hybrid approach based on offset min-sum decoding algorithm is proposed for decoding irregular low-density parity-check (LDPC) codes. In the proposed algorithm, incorporations are made in both the nodes of the Tanner graph with varying error correcting factors. In addition, to the error correcting factors, an adaptive weighting factors are utilized in CNU and VNU process to mitigate the numerical instabilities and negative correlation effects. These modifications in turn enhance the error correcting performance, which ultimately results in a better decoding efficiency. Through exhaustive simulation and comparison, it is shown that the proposed algorithm with a six-bit non-uniform quantization scheme exhibits good error correcting outcome at BER of 10^{-5} with an acceptable computational complexity.

Keywords Irregular LDPC · Offset min-sum decoding · Decoding efficiency Adaptive weighting factors · Hybrid approach

1 Introduction

Low-density parity-check (LDPC) codes [1] have been a revelation in the field of error control coding (ECC) since their comeback in early 1990s [2]. In the current scenario of LDPC decoding, the min-sum algorithm (MSA) [3] and its variants are widely employed for better noise immunity and efficient hardware realization. An innovative approach in MSA is proposed by utilizing two different modification factors in check node process (MNMSA) [4]. This methodology uses the theory of order statistics to derive suitable modification factor to enhance the bit error rate

M. K. Roberts (✉) · E. Sunny
Department of Electronics and Communication Engineering,
Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India
e-mail: king.pane@gmail.com

E. Sunny
e-mail: elizabethsunny1990@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_70

(BER) performance. This method is enhanced to a certain extent by utilizing self-adaptive error correction factors (SAMSA) instead of order statistics [5]. This process achieves good reduction in computational complexity compared to [4]. Even though this process has fewer merits, it is not preferred for larger codeword length and irregular LDPC codes. The simplified version of 2D-MSA (SDMSA) [6] is proposed with an objective to achieve good error correcting performance with minimal numerical instability. Furthermore, in this numerical instability is reduced by replacing conventional multiplication operations with addition and shifting operations. Recently, an improvement to MSA using density evolution (IMSA) was proposed by Wang et al. [7]. This algorithm has a twofold approach, first was by utilizing probability density function to calculate the suitable error correction factor. Then, the selected error correction factor is calculated using the weighted average. In summary, these algorithms show that by utilizing different methodologies ranging from high-order statistics to probability density function achieves better improvement in error correcting performance and complexity reduction. However, these algorithms do suffer from severe error floors when decoded with larger block length irregular LDPC codes.

In this paper, an improved hybrid algorithm based on [8] is presented. In order to subdue the shortcomings of the former algorithms, some new incorporations are carried out in both the nodes of bipartite graph, i.e., check node unit (CNU) and variable node unit (VNU). In general, the magnitude overestimation issue occurs when a single decoding algorithm is used to decode multiple codeword lengths of irregular LDPC codes. This often results in numerical instabilities which increases the requirement for more decoding iterations. To overcome this drawback, multiple error correction factors, along with iteration dependent compensation (weighting) factors are employed in the CNU. These error correcting factors are adaptive and are dynamically used on the switching basis to enhance the convergence speed. Also, in the bit node update process along with error correction factor a penalty (weighted) factor is used to minimize the negative correlation effects of the iterative information updating process. However, the usage of weighting factors enhances the computational complexity moderately. Furthermore, this proposed algorithm is shown to be compatible with different types of irregular LDPC codes with varying code length and rates. Also, this decoding algorithm can be employed in designing resource efficient multi-standard LDPC decoders.

2 Framework

In general, let the set of variable nodes associated to the cth check node be N(c), and the subset excluding the νth bit node from N(c) be denoted as N(c)\\$\nu$. Similarly, the set of check nodes connected to the νth bit node be M(ν), and the subset excluding the cth check node from M(ν) be denoted as M(ν)\\$c$ The following notations are utilized to signify the decoding process:

L_{ch}: The log-likelihood ratio (LLR) information according to the received channel value.

$\alpha_{cv}^{(i)}$: The outgoing LLR data from check node c to variable node ν.

$\beta_{vc}^{(i)}$: The outgoing LLR information from variable node ν to check node c.

$\beta_{v}^{(i)}$: Aposteriori LLR information computed at each iteration.

$h_n^T = (h_{n,1}, h_{n,2}, \ldots, h_{n,N})$ n = 1, 2, …, N: Rows of the H matrix.

$\sigma_{d_c(m)}^{(i)}$: Offset factor for check node update processing.

$\theta_{d_c(m)}^{(i)}$: Iteration dependent weighting factor.

$\zeta_{d_c(m)}^{(i)}$: Stifling factor in check node processing.

$\xi_{d_c(m)}^{(i)}$: Optimally adaptive offset factor for CNU.

$\delta_{d_v(n)}^{(i)}$: Offset factor for variable node processing.

μ: Iteration dependent penalty factor for variable node.

2.1 Decoding Methodology of the Proposed Algorithm

The decoding flow of the proposed algorithm is described briefly below

Step 1: Initialization Process

$$\beta_{\nu c}^{(0)} = L_{ch} \tag{1}$$

Step 2: Check node function

$$\alpha_{c\nu}^{(i)} = \prod_{n \in N(c)\backslash \nu} \text{sgn}(\beta_{nc}^{(i-1)}) \cdot \max(\theta_{d_c(m)}^{(i)} \cdot \min_{n \varepsilon N(c)\backslash \nu} \left|\beta_{nc}^{(i-1)}\right| - \sigma_{d_c(m)}^{(i)}, 0) \tag{2}$$

Message stifling if $\left(h_n^T \cdot x = \, = 1 \text{ mod } 2\right)$, where x = sign $(\beta_v^{(i)})$

$$\alpha_{cv}^{'(i)} = \prod_{n \in N(c)\backslash v} \operatorname{sgn}(\beta_{nc}^{(i-1)}) \cdot \max\left(\min_{n \in N(c)\backslash v} \left|\beta_{nc}^{(i-1)}\right| - \xi_{d_c(m)}^{(i)}, 0 \right) \quad (3)$$

$$\text{where } \zeta_{d_c(m)}^{(i)} = \left(1 + \zeta_{d_c(m)}^{(i)}\right) \cdot \sigma_{d_c(m)}^{(i)} \quad (4)$$

Step 3: Bit node function

$$\beta_{vc}^{'(i)} = \beta_{vc}^{(0)} + \operatorname{sgn}\left\{ \sum_{m \in M(v)\backslash c} \alpha_{mv}^{(i)} - \alpha_{mv}^{'(i-1)} \right\} \cdot \max\left\{ \left| \sum_{m \in M(v)\backslash c} \alpha_{mv}^{(i)} - \alpha_{mv}^{'(i-1)} \right| - \frac{\delta_{d_v(n)}^{(i)}}{\mu} \right\} \quad (5)$$

$$\beta_{v}^{'(i)} = \beta_{vc}^{(0)} + \operatorname{sgn}\left\{ \sum_{m \in M(v)} \alpha_{mv}^{(i)} - \alpha_{mv}^{'(i-1)} \right\} \cdot \max\left\{ \left| \sum_{m \in M(v)} \alpha_{mv}^{(i)} - \alpha_{mv}^{'(i-1)} \right| - \frac{\delta_{d_v(n)}^{(i)}}{\mu} \right\} \quad (6)$$

Step 4: Judge and terminate using end conditions
If $(H^T \cdot \hat{c} = = 0)$ output x and the decoding process is terminated.

3 Simulation Results

In order to validate the decoding efficiency of the proposed methodology, two different irregular rate 1/2 LDPC codes belonging to Wi-MAX and WLAN standards have been considered, respectively. Figures 1 and 2 show the error correcting performance of the proposed algorithm and its comparison with few recent decoding algorithms. For performance comparison purposes, the maximum decoding iterations are fixed at 10 and a six-bit non-uniform quantization scheme is adopted. Furthermore, in the proposed decoding algorithm with the error correction factor values are found using heuristic simulations [9] and are found to be $\left(\alpha_{d_c(m)}^{(i)}, \xi_{d_c(m)}^{(i)}, \delta_{d_v(n)}^{(i)}\right) = (0.21, 0.33, 0.37)$ for rate 1/2, Wi-MAX Standard and (0.18, 0.29, 0.33) for rate 1/2, WLAN Standard irregular LDPC code, respectively. In light of results illustrated in Figs. 1 and 2, it is clearly evident that the proposed scheme outperforms the recently proposed schemes convincingly in terms of BER when measured at BER of 10^{-5}. Therefore, clearly it is demonstrated using experimental simulations that the proposed hybrid scheme achieves better error correction phenomenon with only a small increase in the implementation complexity. This outstanding error correcting phenomenon trait exhibited with irregular LDPC codes of larger codeword length is clearly suited for many emerging wireless communication standards.

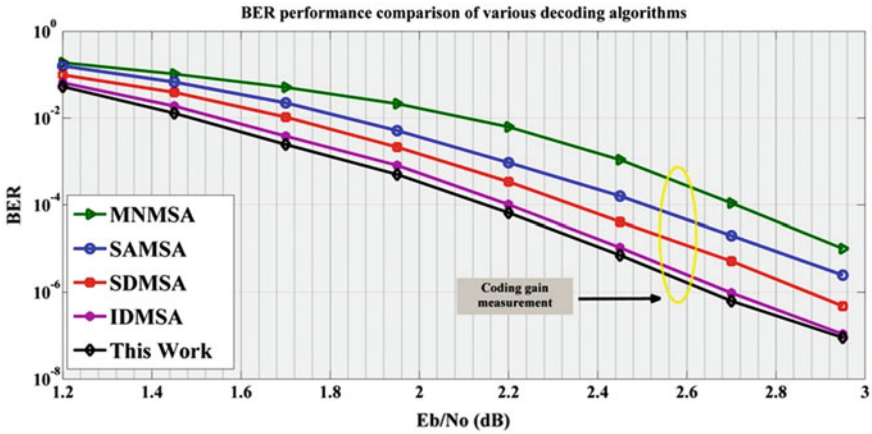


Fig. 1 BER plots for the rate 1/2 (2304,1102) irregular LDPC codes of Wi-MAX standard

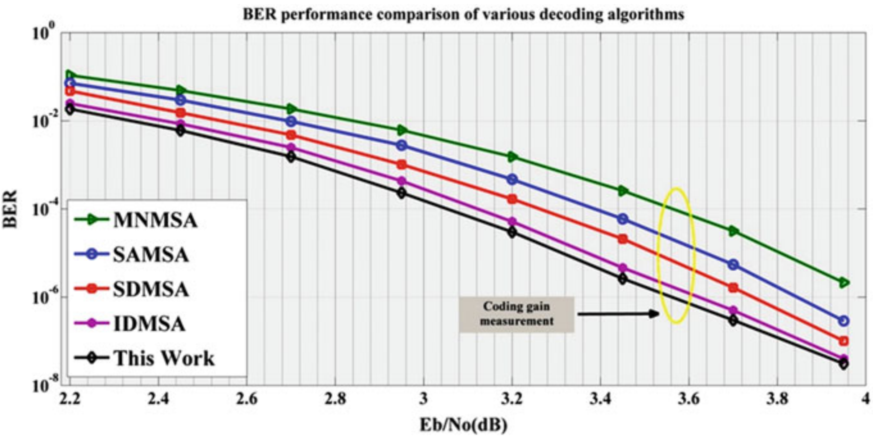


Fig. 2 BER plots for the rate 1/2 (1944,972) irregular LDPC codes of WLAN standard

4 Conclusion

In this paper, an improved hybrid offset decoding algorithm based on adaptive weighting factors is introduced. The proposed methodology is applied to irregular LDPC codes belonging to Wi-MAX and WLAN standard. Through exhaustive simulations, it is clearly seen that the intended algorithm can achieve considerable coding gain improvement with minimum increase in the computational complexity. By analyzing the performance with recent MSA-based decoding algorithms, the proposed scheme exhibit better decoding efficiency without any additional implementation overhead.

Acknowledgements This research work is funded by the Department of Science and Technology, Government of India through Early Career Research Award (Young Scientist) scheme of Science and Engineering Research Board (SERB) (Grant no. ECR/2016/001275).

References

1. Richardson, T., Shokrollahi, A., Urbanke, R.: Design of capacity approaching irregular low-density parity-check codes. IEEE Trans. Inf. Theory **47**(2), 619–637 (2001)
2. Mackay, D.J.C.: Good error-correcting codes based on very sparse matrices. IEEE Trans. Inf. Theory **45**, 399–431 (1999)
3. Fossorier, M., Mihalijevic, M., Imai, H.: Reduced complexity iterative decoding of low-density parity-check codes based on belief propagation. IEEE Trans. Commun. **47**(5), 673–680 (1999)
4. Xue, W., Ban, T., Wang, J.: A modified normalized min-sum algorithm for LDPC decoding using order statistics. Int. J. Satell. Commun. Netw. (Wiley) **35**(2), 163–175 (2017)
5. Cho, K., Chung, K.-S.: Self-adaptive termination check of min-sum algorithm for LDPC decoders using the first two minima. KSII Trans. Internet Inf. Syst. **11**(4), 1987–2001 (2017)
6. Cho, K., Chung, K.-S.: Simplified 2-dimensional scaled min-sum algorithm for LDPC decoder. J. Electr. Eng. Technol. (JEET) **12**(1), 1921-718 (2017)
7. Wang, X., Cao, W., Li, J., Shan, L., Cao, H., Li, J., Qian, F.: Improved min-sum algorithm based on density evolution for low-density parity check codes. IET Commun. **11**(10), 1582–1586 (2017)
8. Roberts, M.K., Jayabalan, R.: A modified optimally quantized offset min-sum decoding algorithm for low-complexity LDPC decoder. Wirel. Pers. Commun. J. **80**(2), 561–570 (2015)
9. Chen, J., Fossorier, M.P.C.: Density evolution for two improved BP-based decoding algorithms of LDPC codes. IEEE Commun. Lett. **6**(5), 208–210 (2002)

Simulation of Phasor Measurement Unit for Research and Teaching Applications

Hannan Ahmad Khan, Mohd. Rihan and Mohammad Zaid

Abstract The dynamic behavior of the modern power system requires accurate monitoring of certain important parameters such as bus voltage, frequency, phase angle, rate of change of frequency. With the advent of PMU, precise measurement of these parameters can be achieved. This paper is an attempt to develop a simple simulator to provide a basic idea about PMU working and algorithms involved in its phasor estimation. PMU simulator is being developed, and its performance under different test conditions is observed. This paper discusses the basic design and algorithm used in phasor measurement unit (PMU). Estimation of the rate of change of frequency (ROCOF) is also made with precision. The average filter is implemented, and its effectiveness to die out the second harmonics is revealed introduced by the complex gains. It is expected that the work will be helpful for teaching the PMU concept at advance level as well as initiating the researches in this area.

Keywords Phasor measurement unit · Phasor estimation techniques
Rate of change of frequency (ROCOF)

1 Introduction

Power demand is growing tremendously, and the power grid is maturing with time [1]. The reliable electrical power is required to keep the world moving. Continuous and accurate measurement of certain critical parameters is therefore very important [2]. For instance, SCADA provides only magnitudes; power flow between the transmission lines primarily depends on the phase difference of the bus voltages.

H. A. Khan (✉) · M. Rihan · M. Zaid
Department of Electrical Engineering, Aligarh Muslim University, Aligarh, India
e-mail: hannanahmadkhan@gmail.com

Mohd. Rihan
e-mail: m.rihan.ee@gmail.com

M. Zaid
e-mail: mohammad.zaid@zhcet.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_71

The PMU is becoming the backbone of the modern power system monitoring and control due to its capability of measuring the voltage and current synchronized to the same time reference through the GPS. The reporting rate of PMU is much faster than the conventional SCADA system.

Data from different PMUs is sent to phasor data concentrators (PDCs). This data is available for local application in real time [3]. After blackout of 1996 and 2003, this technology gained momentum and became a wide area of research [4]. Using the PMU simulator developed in the present work, one can get the physical interpretation of the effect of disturbances on the ideal supply signal [5]. It helps to understand the working of the algorithm to develop the practical PMU with better processing and results.

2 Different Phasor Estimation Techniques

The streaky nature of disruption affects the synchrophasor, frequency, ROCOF. Many algorithms have been developed to estimate them. These algorithms are based on different techniques, in frequency and time domain. Some techniques are discrete Fourier transform (DFT), direct model matching, filtering and demodulation, zero crossing, sliding DFT, etc.

2.1 Discrete Fourier Transform (DFT)

DFT is a most used technique for phasor calculations. DFT has two algorithms: recursive and non-recursive.

Let a sine input signal given by Eq. (1).

X_M represents the peak value of the signal.

$$x(t) = X_M \sin(\omega t + \theta) \tag{1}$$

This sinusoid is symbolized by a phasor X in Eq. (2).

$$X = \left(\frac{X_M}{\sqrt{2}}\right)e^{j\theta} = (X_M/\sqrt{2})(\cos\theta + j\sin\theta) \tag{2}$$

Assuming that x(t) of 60 Hz is sampled N times per cycle to produce the set of samples $\{x_m\}$ as given in Eq. (3).

$$x_m = \sqrt{2}\sin(2\pi m/N + \theta) \tag{3}$$

The DFT contains a fundamental frequency component given by Eq. (4) [6].

$$X^m = \frac{\sqrt{2}}{N}\sum_{m=0}^{N-1} x_m \cos(2\pi m/N) + j\frac{\sqrt{2}}{N}\sum_{m=0}^{N-1} x_m \sin(2\pi m/N) \tag{4}$$

A program that saves computation utilizing the previous information is termed as recursive discrete transform algorithmic program [7, 8]. The signal of equation offers the basic frequency part as given in Eq. (5).

$$X^{N+1} = X^N + \sqrt{2}/(x_{N+m} + x_N)e^{-jm\theta} \tag{5}$$

Discrete Fourier transform (DFT) converts the signal from time domain to frequency domain. Time and frequency domains have the identical information, but the form is different. By using the information of one domain, the information of other domain can be derived. If x(t) is a sine signal in time domain sampled N times in a cycle to generate a set of x(m), then DFT of input can be estimated using Eq. (6).

$$X(k) = \frac{1}{N}\sum_{m=0}^{N-1} x(m)e^{-j2\pi mk/N} \tag{6}$$

Frequency domain X(k) have two parts, real part of X(k) and imaginary part of X(k), with phase and magnitude of the data at every sample can be estimated using Eqs. (7) and (8).

$$Magnitude = \sqrt{(\mathrm{Re}\,X(k))^2 + (\mathrm{Im}\,X(k))^2} \tag{7}$$

$$Phase = \tan^{-1}(\mathrm{Im}\,X(k)/\mathrm{Re}\,X(k)) \tag{8}$$

3 PMU Simulator Architecture

PMU is used to monitor and control the disturbances in data of power system and is time synchronized; i.e., each sample is time stamped and has the same reference of time. PMU data is initialized to satisfy the Nyquist criterion by using GPS. This paper is focused on measurement part of PMU only [9] (Fig. 1).

3.1 Phasor and Frequency Calculation

In phasor calculation, the amplitude and phase of the input signal have been calculated by using non-recursive DFT at the fundamental frequency of the signal.

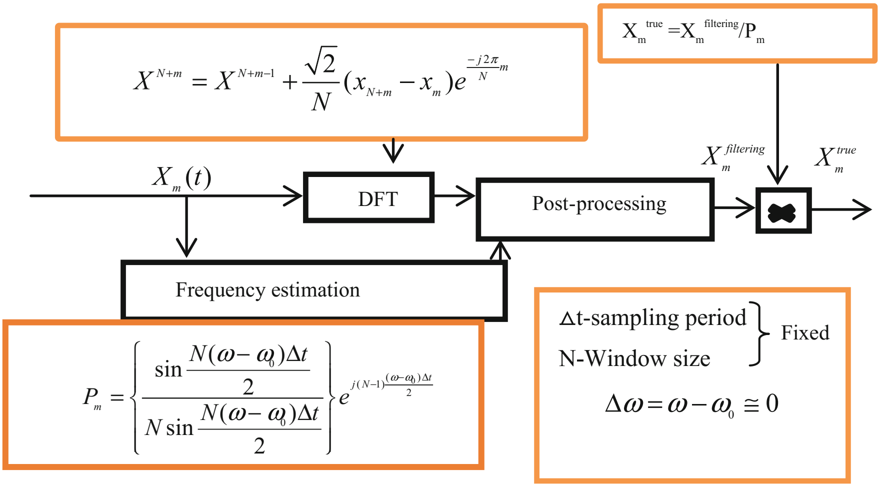


Fig. 1 PMU simulator architecture

In steady-state conditions, frequency remains at nominal value but in practical cases power system frequency changes [10]. A power system is subjected to diverse disturbances which affect the frequency of power system. If changes can be monitored, then abnormal conditions can be predicted earlier and corrective measures may be taken to avoid such conditions. This paper uses the method proposed by Fan and Centeno [11].

Let, N = the size of the data window for phasor estimation.

W = the size of the data window for frequency estimation.

In this method, series can be written as (Fig. 2)

$$(\phi^{(D-E)} - \phi^{(A-B)})/W\Delta t = \Delta\omega_{r+((W+N-1)/2)} \tag{9}$$

Generally r = 0,

$$(\phi^{(D-E)} - \phi^{(A-B)})/W\Delta t = \Delta\omega_{((W+N-1)/2)} \tag{10}$$

Equations (9) and (10) estimate the frequency deviation obtained from phasor estimation from window A–B, and D–E corresponds to the instantaneous frequency deviation of sample $((W+N-1)/2)$ or at the time $(((W+N-1)/2)\Delta t)$.

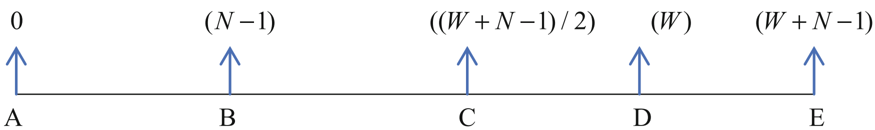


Fig. 2 Phasor-based frequency estimation

This technology assigns the proper time to the estimated frequency and is supported by the traditional method.

3.2 Post-Processing

In post-processing, data has been corrected which is already affected by windowing, leakage, etc., to improve TVE. The calculated phasor is attenuated by two complex gain (P_m) and (Q_m). The average filter is one of the most useful and effective filters to minimize the distortions and harmonics introduced by the complex gain (Q_m) and second harmonics, but some traces of the second harmonics remain in the signal. Therefore, second harmonic does not die out completely, and some ripples remain [12].

$$P_m = \left\{ \frac{\sin \frac{N(\omega - \omega_0)\Delta t}{2}}{N \sin \frac{N(\omega - \omega_0)\Delta t}{2}} \right\} e^{j(N-1)\frac{(\omega - \omega_0)\Delta t}{2}} \tag{11}$$

$$Q_m = \left\{ \frac{\sin \frac{N(\omega + \omega_0)\Delta t}{2}}{N \sin \frac{N(\omega + \omega_0)\Delta t}{2}} \right\} e^{-j(N-1)\frac{(\omega + \omega_0)\Delta t}{2}} \tag{12}$$

4 Results and Discussion

This section considers various cases of the input signal. The outputs of three phase as well as single-phase model have been presented in the section.

The balanced three-phase sinusoidal supplies of unit magnitude (RMS) are taken as inputs. Two different cases with balance supply have been considered here.

1. In first case, step disturbance has been used and frequency of balanced three-phase signal changes from 60 to 58.5 Hz after some time.
2. In second case, ramp disturbance has been used and frequency of balanced three-phase signal changes from 60 to 59 Hz after some time.

4.1 Results from Step Disturbance

Frequency deviation and rate of change of frequency. When frequency deviates from 60 to 58.5 Hz, it takes some time to reach its final value. The sampling steps are uniform and it can give information of frequency change from initial to final value. Steps can be counted by using subroutine. By using this method, the rate of

Fig. 3 Frequency deviation

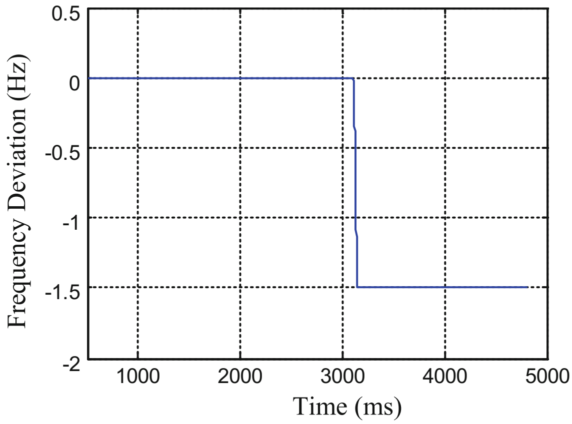
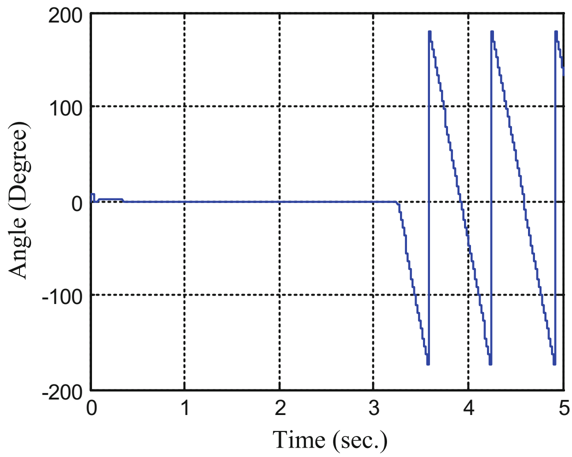


Fig. 4 Downsampled positive sequence angle from step disturbance



change of frequency which comes out to be 30.612344556637 Hz/s is correct up to sixth decimal. Frequency deviation is using (P_m) for estimation.

Downsampled positive sequence angle and magnitude. When step disturbance is applied, the resulting positive sequence angle after downsampling and positive sequence magnitude are shown in Figs. 4 and 5. Sample at every N interval is being chosen to decrease sampling rate from N time of nominal frequency to nominal frequency (Fig. 3).

Single-phase processing. After applying step signal the effect of complex gain, Q_m is observed in single-phase phasor calculation during the frequency change with and without filtering and is shown in Figs. 6 and 7.

Effect of filter. When ramp disturbance is applied, the single-phase magnitude is shown in the Figs. 8 and 9. The effect of the filter is seen physically; it dies out the harmonics introduced in the input.

Fig. 5 Positive sequence magnitude

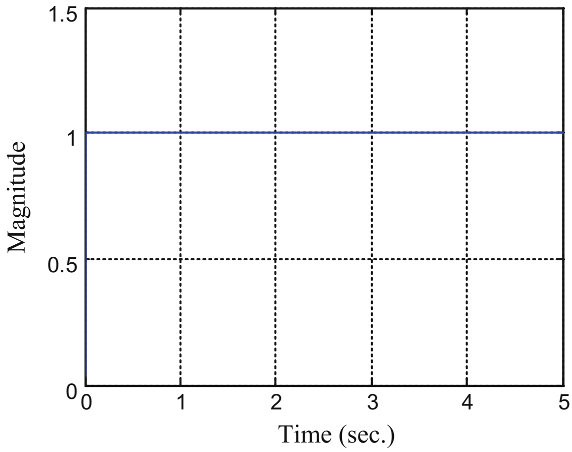


Fig. 6 Single-phase magnitude with filtering

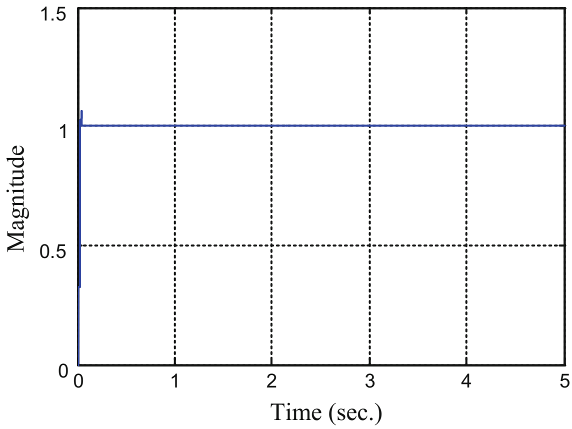
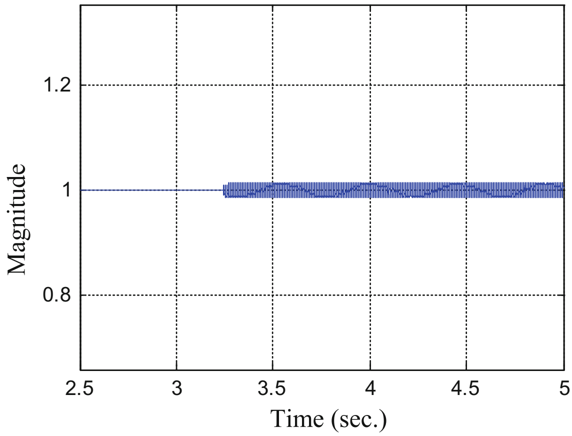


Fig. 7 Single-phase magnitude without filtering



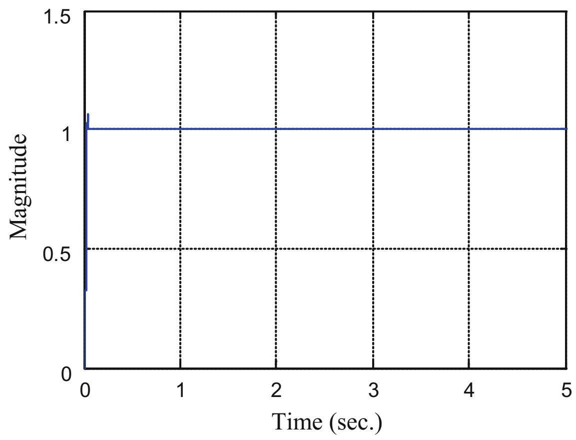


Fig. 8 Single-phase magnitude without filtering

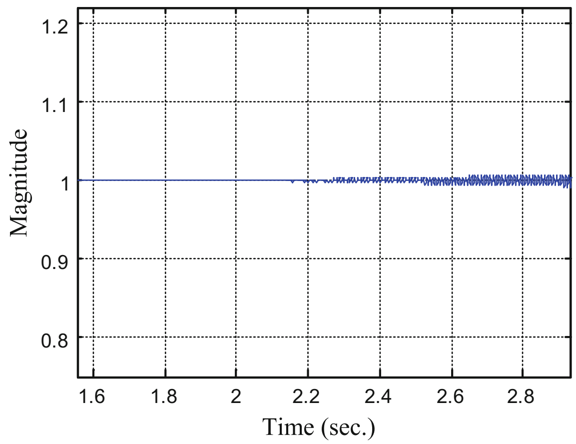


Fig. 9 Single-phase magnitude without filtering from ramp disturbance

Downsampling. The angle starts changing in positive sequence downsampled data when ramp signal is applied. Curve is shown in the Fig. 10 which gives the information about the change of angle.

At the time of starting of any system, transients are present. The work presented in this paper is fast enough to detect those transients. Graphs showed gives information about different parameters and the effect of transients on them.

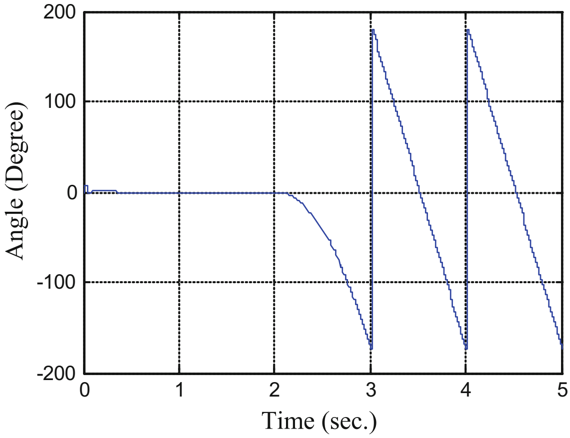


Fig. 10 Downsampled positive sequence angle from ramp disturbance

5 Conclusion

In this work, Simulink-based PMU simulator is presented. The motive of this simulator is a physical interpretation of signals to understand the effects of different disturbances on the input signal at different stages of processing. With the real data or by introducing the time lag in frequency change, accuracy of rate of change of frequency can be calculated. This simulator helps to test different algorithms. The estimated data were compared with the results given in the different literature. The models available are not fast to detect transients. The next step is to develop the new algorithms for non-uniform sampling and reduce the computation used in estimation with enhanced precision.

References

1. Soonee, S., Narasimhan, S., Porwal, R., Kumar, S., Kumar, R.A., Pandey, V.: Application of phase angle measurement for real time security monitoring of Indian electric power system—an experience. Power grid corporation of India Limited, CIGRE Working Group C2–107, Cigre session (2008)
2. Singh, B., Sharma, N., Tiwari., A., Verma, K., Singh, S.: Applications of phasor measurement units (PMUs) in electric power system networks incorporated with FACTS controllers. Int. J. Eng. Sci. Technol. **3**(3), 64–82 (2011)
3. Monti, A., Muscas, C., Ponci, F.: Phasor measurement units and wide area monitoring systems. Elsevier AP, Amsterdam (2016)
4. Haridas, R.: Synchrophasor measurement technology in electrical power system. Int. J. Eng. Res. Technol. **2**(6), 2063–2068 (2013)
5. Dotta, D., Chow, J., Bertagnolli, D.: A teaching tool for phasor measurement estimation. IEEE Trans. Power Syst. **29**(4), 1981–1988 (2014)
6. Sedlacek, M., Krumpholc, M.: Digital measu.of phase difference-a comparative study of DSP algorithms. Metrol. Meas. Syst. (1), 15 (2005)

7. Mondal, S., Murthy, C., Roy, D., Mohanta, D.: Simulation of phasor measurement unit (PMU) using labview. In: 2014 14th International Conference on Environment and Electrical Engineering (2014)
8. Nguyen, T., Li, X.: A fast and accurate method for estimating power systems phasors using DFT with interpolation. In: IEEE Power and Energy Society General Meeting (2006)
9. Phadke, A., Thorp, J.: Synchronized phasor measurements and their applications. Springer, Boston, MA (2008)
10. Premerlani, W., Kasztenny, B., Adamiak, M.: Development and implementation of a synchrophasor estimator capable of measurements under dynamic conditions. IEEE Trans. Power Deliv. **23**(1), 109–123 (2008)
11. Fan, D., Centeno, V.: Phasor-based synchronized frequency measurement in power systems. IEEE Trans. Power Deliv. **22**(4), 2010–2016 (2007)
12. Dotta, D., Chow, J., Vanfretti, L., Almas, M., Agostini, M.: A MATLAB-based PMU simulator. In: 2013 IEEE Power & Energy Society General Meeting (2013)

Maximum Network Flow Aware Multi-lightpath Survivable Routing in WDM Networks

Dinesh Kumar Tyagi and V. K. Chaubey

Abstract In this paper, a survivable routing and wavelength assignment strategy using flow network concepts of network graph theory has been proposed. The proposed approach determines a route and the wavelength simultaneously for incoming lighpath connection traffics. Generally, all algorithms of routing and wavelength assignment problem solve the route selection and channel assignments problem separately. In the proposed approach, flow network represents the wavelength resources in such a way to solve the survivable routing and wavelength assignment problem for WDM networks efficiently. This strategy provides all the possible wavelength continuous flow paths to maximize the lightpaths between pair of source and destination node of the requests. It chooses a suitable lightpath from the available multiple link-disjoint lighpaths to transmit the data for an improved network performance. Simulation results taken over different network topologies confirm that the maximum network flow-based lightpath routing for survivable network has significant improvement of call acceptance ratio when compared against the conventional fixed alternate routing, shortest path and K-shortest path routing approach.

Keywords Routing and wavelength assignment · Lightpath · WDM
Survivable routing · Maximum flow network · Optical network

D. K. Tyagi (✉)
Department of CSE, MNIT Jaipur, Jaipur, India
e-mail: dktyagi.cse@mnit.ac.in

V. K. Chaubey
Department of EEE, BITS Pilani, Pilani, India
e-mail: vkc@pilani.bits-pilani.ac.in

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_72

1 Introduction

The advancements in the technology of wavelength division multiplexing (WDM) have enabled systems which have capability of providing huge amounts of required bandwidth. In order to meet the traffic demand of existing and future communication applications, wavelength-routed WDM optical networks have been considered to be a viable solution of current and next-generation large backbone networks [1–3].

A wavelength-routed optical network accomplished a connection request through a lightpath. For an incoming connection request, a wavelength continuous path requires to be established between each pair of source–destination node of request. Routing and wavelength assignment (RWA) problem is the process of selection of route and wavelength to accomplish the establishment of lightpath [4–6]. RWA problem may be solved as two sub-problem in which first route is determined, and thereafter, a wavelength channel is assigned. It is necessary to employ good RWA algorithm for wavelength-routed WDM optical networks, to determine lightpaths in an efficient manner for efficient network operation [2].

A wavelength channel in WDM optical network can provide data rate of up to terabits per second to the end users and applications [3, 7]. Such high-speed WDM optical network system may suffer from enormous loss of data due to the network failures of node, link, fiber cuts, or channel faults, etc. [8]. Hence, in the designing, the high speed WDM networks, survivability has an important issue of next-generation networks. For the past decade survivability issues have received intensive attention from the optical network researchers in the literatures [9–15]. Generally, survivability is an important concern of high-speed WDM optical network. The problem of survivable routing and wavelength assignment is critically necessary to ensure the survivability of WDM all optical networks [16]. The purpose of SRWA problem is to assign a link disjoint primary and backup lightpaths to each lightpath request such that the total number of request accepted is maximized [17]. In this work, a survivable lighpath RWA strategy using maximum flow network concepts of network graph theory has been proposed. This approach determines a route and also the wavelength simultaneously for incoming lighpath connection requests. Mostly, all routing and wavelength assignment algorithms solve the route selection and channel assignment problem separately. The proposed strategy chooses a suitable pair of lightpath from the available multiple link-disjoint lightpaths to transmit the data for an improved network performance with survivability.

The rest of the paper is organized as first maximum flow network concept and proposed strategy of survivable routing, and wavelength assignment is described in Sect. 2. The performance analysis of the proposed strategy is demonstrated in Sect. 3, and Sect. 4 finally concludes the presented work.

2 System Model of Flow Networks

The maximum flow problem is the problem of determining of a flow of maximum value on a network from a source node s to a sink node t. It is an important problem which is applicable in a vast variety of engineering and scientific problems. The maximum network flow problem has been studied extensively for various applications. A flow network can model the scenarios of flowing liquids through pipes, flow of assembly lines parts, flow of current in electrical network, flow of data or information in infrastructure of communication networks, etc. Here, we introduce some basics of flow network theory.

In network graph theory, a flow network G(V, E) is represented as a directed graph such that each edge (u, v) $\in$ E of graph has a value of nonnegative capacity C(u, v) $\geq$ 0 having one special source vertex s, and another one is the sink (destination) vertex t [18]. Each edge of flow network receives a flow. A flow in flow network is a nonnegative real-valued function f: V $\times$ V $\rightarrow$ R. The three properties that flow must full fill are:

$$\text{Capacity restriction: } f(\mathrm{u},\mathrm{v})\ \mathrm{C}(\mathrm{u},\mathrm{v}), \forall(u,v) \tag{1}$$

$$\text{Skew restriction:} f(\mathrm{u},\mathrm{v}) = -f(\mathrm{v},\mathrm{u}), \forall(u,v) \tag{2}$$

$$\text{Conservation of Flow: } \sum_{v \in \mathrm{V}} f(u,v) = 0,\ \forall \mathrm{u}\ \in \mathrm{V} - \{\mathrm{s},\mathrm{t}\} \tag{3}$$

Using the above notations, maximum flow can be described as an objective function which maximizes the flow through the network, i.e.

$$\text{Maximize } \sum_{v \in \mathrm{V}} f(s,v) \tag{4}$$

subjected to restrictions $C(u,v) = 0$, $\mathit{if}\ (u,v) \notin \mathrm{E}(u,v)$ and Eqs. 1–3.

Flow networks problem described above can be utilized to find all possible edge disjoint flow paths of graph from a source s to a destination sink d. All the paths with available capacity are traced; these paths are known as augmenting paths. The results we will get finally are the residual network with maximum flow [17, 19]. Various algorithms are suggested in the literature to solve max flow network problem to get the maximum flow [17–21]. But Boykov method based on augmenting approach has most preferable due to its more feasibility to the problem for network routing [21].

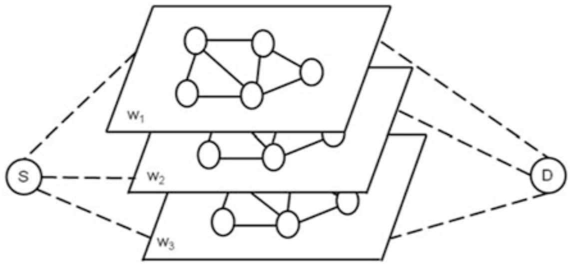
2.1 Proposed Approach

The maximum flow-based algorithm has proposed which is used for a wavelength continuity constraints survivable routing in WDM optical network. It is used to establish a wavelength continuity constraint pair of disjoint lightpaths between given source–destination node pair. Thus, proposed survivable routing uses wavelength continuity constraint route selection and channel assignment process simultaneously and makes it more efficient and reliable in a large-scale network environment. To solve the SRWA problem, we represent the physical topology in the form of a flow network and thenceforth, used max network flow graph theory to determine all max flow paths to determine both a route and wavelength simultaneously.

2.2 Conversion of Physical Topology

The physical network topology which is used as a flow network needs to be converted into a graph which allows us to solve the survivable routing and wavelength assignment problem under the restriction of wavelength continuity constraint. A given network having N nodes and L link having total W wavelength can be represented as a W-layered graph between any two node's pair as shown in Fig. 1. This is similar to a wavelength graph where each edge in the graph represents a wavelength. The total number of edges in such graphs can be found by the product of the number of links L and number of wavelengths W. At any point of time, a connected edge having capacity value 1 indicates the presence of that wavelength; while value of the capacity of 0 to an edge shows that the wavelength is busy and not available. A connected edge having assigned a capacity value of

Fig. 1 Logical topology



other than 1or 0, indicates that this wavelength is sharable and can be assigned to the backup protection paths. Such links are associated with a capacity value of larger than or equal to 2, indicating that how many time this wavelength is shared among backup paths. For example, capacity value of 2 indicates that single backup path uses the wavelength. When a backup path uses a wavelength which is already assigned to other backup paths, its shareability count is incremented by one. Hence for each new lightpath which is established on a certain wavelength, the edge corresponding to 0 indicates the unavailability of this wavelength at that moment. Similarly, when a lightpath terminates, the corresponding edges capacity are restored to value of 1 in the wavelength graph to indicate their availability for next lightpath request. Primary paths can use only free wavelengths, while backup path may use either free or sharable wavelengths. However, sharable wavelengths are preferred for backup paths to optimize the resources.

Converted physical topology enables us to determine both the route and wavelength for a lightpath simultaneously instead of separately as two-step process. Thus, route selection and a wavelength assignment problem take place simultaneously. Survivable routing and wavelength assignment problem becomes more complex due to the wavelength continuity constraint. By translating the network topology in the form which is suitable to the maximum flow problem and wavelength continuity, constraint is considered into the network topology representation.

Proposed SRWA Algorithm

Proposed SRWA algorithm

```
Input:Physical  network  topology  G(Node,Link,Wavelength);
   Set of connection request,source and destination node.
Output: Pairs of disjoint primary and backup lightpath.
---------------------------------------------------------------
RAMPR_SRWA(G(N,L,W) )
 Begin
  For (∀n ∈N of physical topology G) do
   Replicate node n of G up to n1, n2,......nw  in G'
   logical topology.
 endfor
For (∀ link (n, m) ∈ L of physical topology) do
Create edges(n1, m1),(n2, m2),...,(nw, mw) into G' topology.
endfor
For (each incoming connection request Rsd) do
 For(∀edge (nλ,mλ)∈ L of Logical topology G'where λ∈ W)do
     if (λ is free)
        Assign capacity value of edge link (nλ, mλ) to 1.
     else if (λ is busy)
        Assign capacity value of edge link (nλ, mλ) to 0.
    endif
  endfor
Compute all paths using BoykoKolmogorov max flow approach
over transformed topology.
if (Paths not found) then
  Drop and reject the connection request from network.
else
  Arrange all the wavelength continuous paths according
   to path length.
  Select the optimal shortest path as primary lightpath.

endif
For(∀edge (nλ, mλ)∈L of Logical topology G'where λ∈ W) do
  if (edge (n, m) belongs to primary lightpath )
     Assign capacity value of edge link (nλ, mλ) to 0λ.
  if (λ is sharable)
 Assign capacity value of edge link (nλ,mλ)equal to number
  of backup lightpaths using that λ.
else if (λ is free)
     Assign capacity value of edge link (nλ, mλ) to 1.
   else
    Assign capacity value of edge link (nλ, mλ) to 0.
  endif
endfor
Use the transformed Physical topology for Backup path
computation.
Compute all paths using BoykoKolmogorov max flow approach
 over transformed topology.
if (Paths not found) then
    Drop the connection request from network
    Release all the wavelengths assigned to primary path
else
 Arrange all the wavelength continuous paths according to
  its maximum path flow value.
 Select the best path which has a largest maximum flow
 value as Backup lightpath.
All remaining paths may kept as protection path at source
 node which may be tried in the restoration process.
 endif
endfor
```

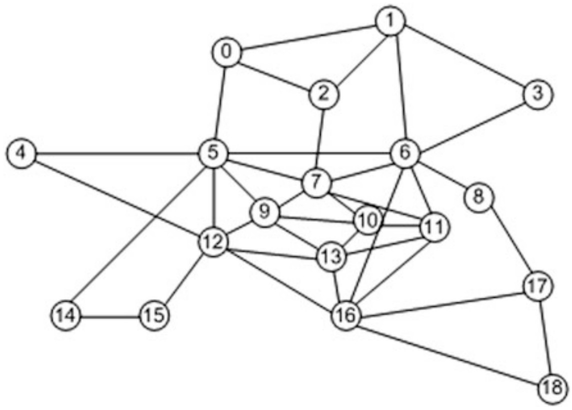


Fig. 2 EON network topology

3 Performance Analysis

Extensive simulation has been performed to evaluate the performance of our proposed strategy which is compared with other three routing strategies. We have simulated the dynamic network environment over three representative test network topology of Figs. 2 and 3. The EON network topology of Fig. 2 consists of 19 nodes and 35 links, and NSFNET topology of Fig. 3 has 14 Nodes and 22 links. Each link in the network is assumed to have similar number of W wavelength channels. It is assumed that connection requests arrive randomly for a source and destination node pair. The bandwidth requirement of the connection request is single wavelength, and there is no waiting queues in network. Hence when a connection request is not fulfilled due to limited network resource or route cannot

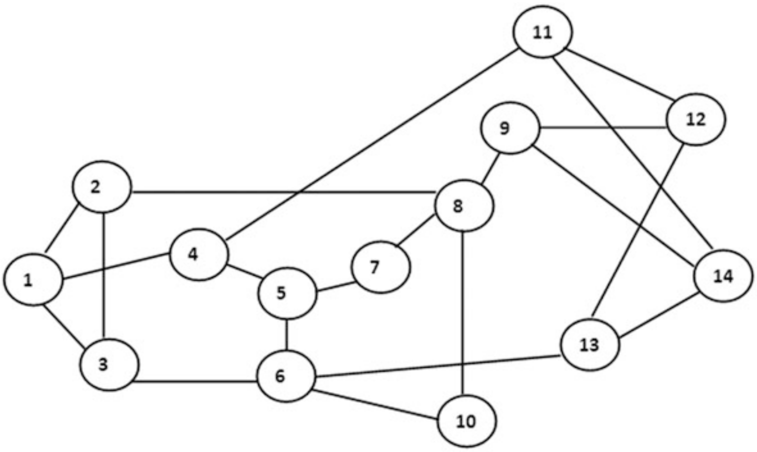


Fig. 3 NSFNET network topology

be determined, then it is abandoned immediately from the network and blocked. In simulation, 400 requests are generated and averaged results of over 20 such simulation used for performance analysis. The two metrics used to measure the performance are the blocking probability and improvement gain in blocking probability.

3.1 Blocking Probability Performance

The blocking performance of the proposed approach is shown in Figs. 4 and 5. In results, blocking probability demonstrated as a function of incoming connection request on EON and NSFNET topologies of Fig. 2 and Fig. 3, respectively.

The performance of proposed method is considerably higher as compared to fixed alternate routing, shortest path and K-shortest path routing strategies. Blocking performance increases with the increase in traffic load of the network as expected. The blocking probability of proposed method is significantly lower even at higher traffic load of incoming connection requests as compared to other three survivable routing strategies. At higher traffic load, less spare resources are available in the network so blocking probability performance also increases as the traffic load increases. In particular, plots show that proposed method for survivable routing can significantly perform better and resource efficient than other three conventional approaches.

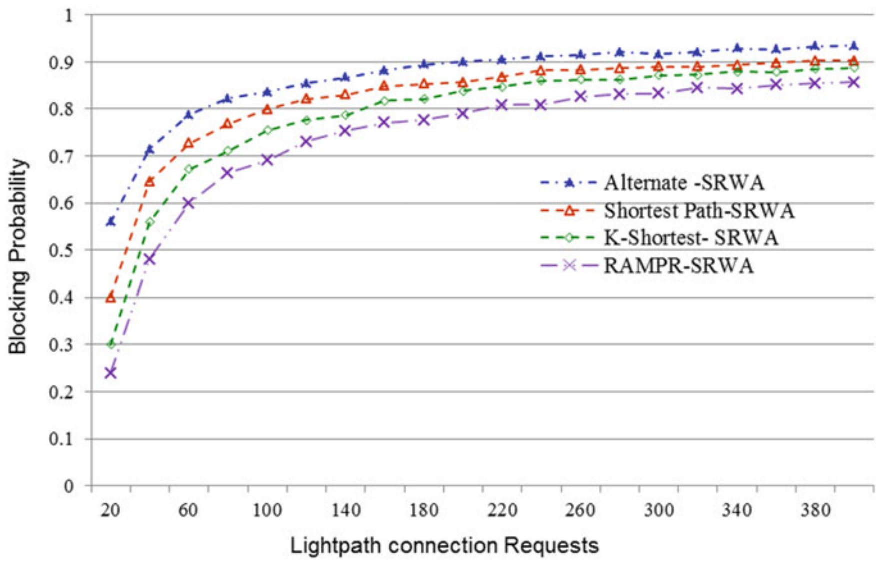


Fig. 4 Blocking performance on EON topology

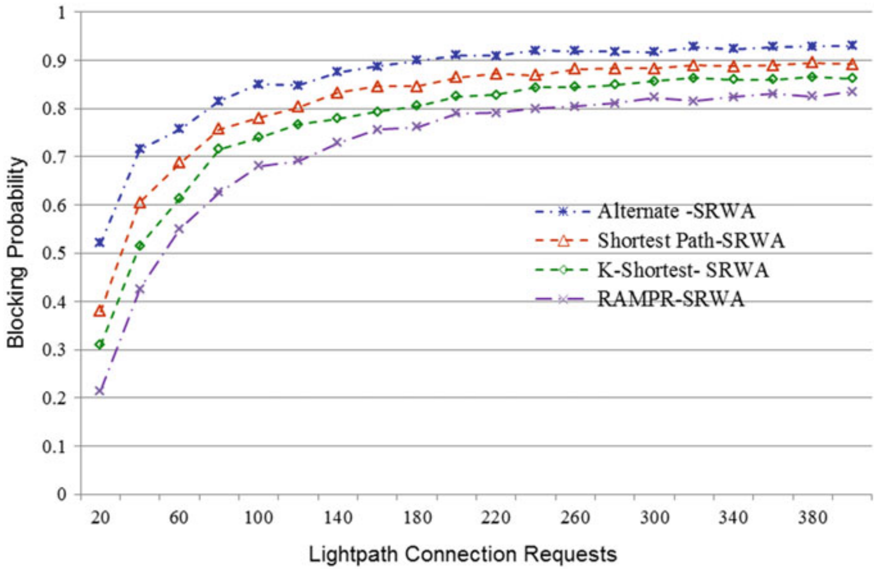


Fig. 5 Blocking performance on NSFNET topology

3.2 Blocking Probability Improvement Performance

In this section, the percentage improvement of blocking probability of maximum flow network-based survivable routing as a function of incoming connection requests for EON and NSFNET network topology have investigated and simulation results have presented in Fig. 6 and Fig. 7, respectively. In all the plots, percentage improvement in blocking performance of proposed algorithm relative to alternate routing algorithm higher than the performance improvement relative to shortest and K-shortest path routing algorithm.

Maximum flow concept-based routing algorithm performs significantly better relative to all the other three approaches. Plots show that there is decrease of percentage improvement gain of blocking probability as rate of incoming lightpath connection requests increases. At low requests arrival rate in the network, probability is low and relative benefit in terms of blocking improvement is high in comparison to higher requests arrival rate.

It is observed that call blocking probability increases at higher arrival rate of connection requests, but relative percentage improvement of blocking probability is not increasing as fast as equivalent to blocking probability itself. That is why it is evident from graphs that percentage improvement is decreasing in all the three network topologies. Overall observation is that the performance of maximum network flow-based survivable lightpath protection routing is significant and promising.

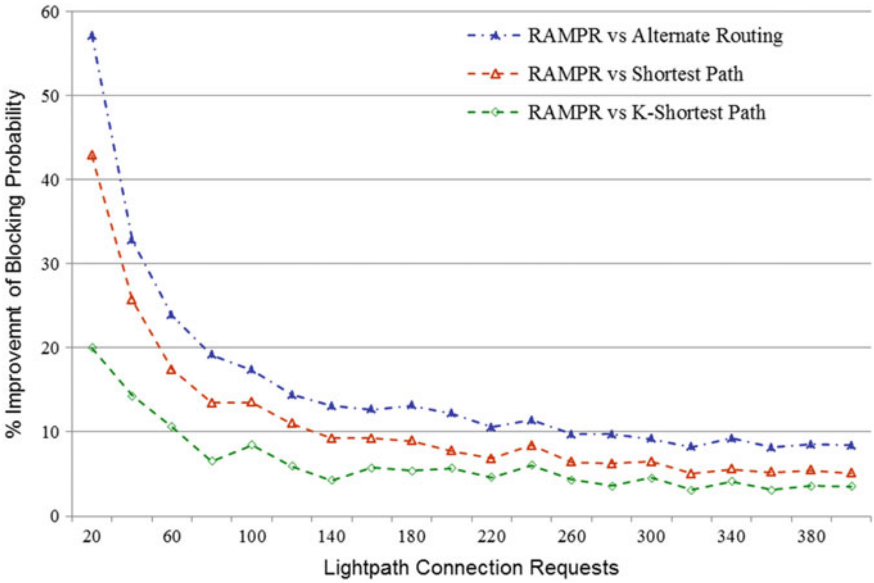


Fig. 6 Blocking probability improvement on EON topology

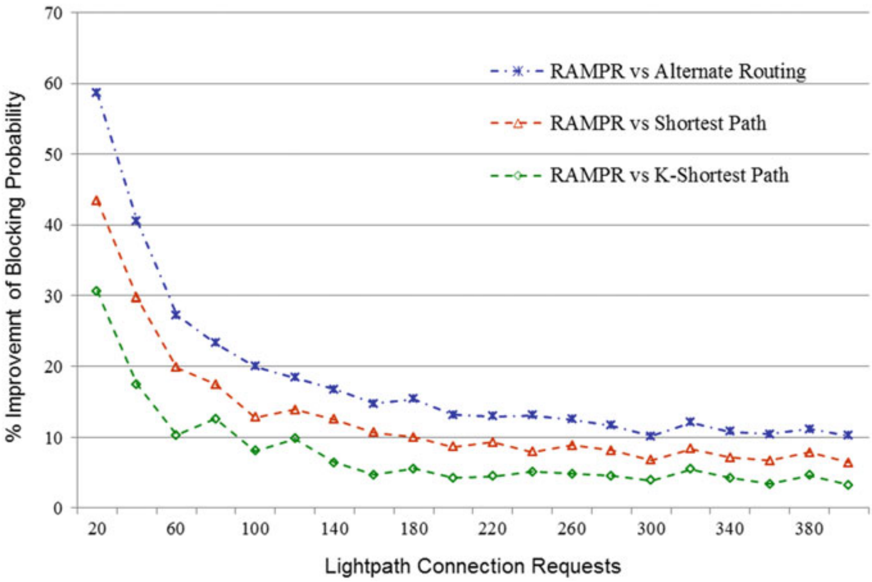


Fig. 7 Blocking probability improvement on NSFNET topology

4 Conclusion

In this paper, we presented a survivable routing and wavelength assignment strategy using maximum flow concepts of network graph theory to improve the network performance in WDM optical networks. Multi-lightpath mechanism improves the wavelength resource utilization intelligently. The survivable lightpath routing performed using maximum network flow concepts determines the routes and available free wavelength channels simultaneously for an incoming traffic requests. Extensive simulation results over sample network topologies confirms that maximum flow network-based routing for the survivable network has lower call blocking probability performance compared to shortest path routing, alternate fixed routing and k-shortest path routing approach.

References

1. Banerjee, D., Mukherjee, B.: A practical approach for routing and wavelength assignment in large wavelength-routed optical networks. IEEE J. Sel. Areas Commun. **14**(5), 903–908 (1996)
2. Asuman, E., Bertsekas, D.P.: Routing and wavelength assignment in optical networks. EEE/ACM Trans. Netw. **11**(2), 259–272 (2003)
3. Berthold, J., Saleh, A.A.M., Blair, L., Simmons, J.M.: Optical networking: past, present, and future. J. Lightwave Technol. **26**(9), 1104–1118 (2008)
4. Choo, H., Valdimar, V.S., Mukherjee, B.: Routing and wavelength assignment in optical WDM networks with maximum quantity of edge disjoint paths. Photonic Netw. Commun. Springer Sci. **12**(2), 145–152 (2006)
5. Shekhawat, V.S., Saini, A.K., Chaubey, V.K.: Efficient heurstic algorithms for routing and wavelength assignment in optical WDM networks. IUP J. Telecommun. **4**(3), 31–43 (2012)
6. Chatterjee, B.C., Sarma, N., Sahu, P.P.: Review and performance analysis on routing and wavelength assignment approaches for optical networks. IETE Tech. Rev. **30**, 12–23 (2013)
7. Bejerano, Y., Koppol, P.V.: Resource reservation comparison of fault resilient routing schemes. In: 39th Annual IEEE Conference on Local Computer Networks, Edmonton, AB, 2014, pp. 81–89 (2014)
8. Ramamurthy, S., Sahasrabuddhe, L.: Survivable WDM Mesh networks. J. Lightwave Technol. **21**(4) (2003)
9. Beshir, A., Kuipers, F., Van Mieghem, P., Orda, A.: On-line survivable routing in WDM networks. In: 21st International Teletraffic Congress, Paris (2009)
10. Ruan, L., Zheng, Y.: Dynamic survivable multipath routing and spectrum allocation in OFDM-based flexible optical networks. IEEE/OSA J. Opt. Commun. Netw. **6**(1), 77–85 (2014)
11. Ho, P.-H., Tapolcai, J., Mouftah, H.T.: On achieving optimal survivable routing for shared protection in survivable next-generation Internet. IEEE Trans. Reliab. **53**(2), 216–225 (2004)
12. Liu, Y., Guo, L., Wei, X.: Optimizing backup optical-network-units selection and backup fibers deployment in survivable hybrid wireless-optical broadband access networks. J. Lightwave Technol. **30**(10), 1509–1523 (2012)
13. Wu, J., Liu, Y., Yu, C., Wang, X., Zhao, J.: Multicast protection based on enhanced intelligent p-cycle method in optical mesh networks. Opt. Eng. **52**(1), 015008-1–015008-7 (2013)

14. Guo, L., Cao, J., Hongfang, Yu., Li, L.: Path-based routing provisioning with mixed shared protection in WDM Mesh networks. J. Lightwave Technol. **24**(3), 1129–1141 (2006)
15. Lu, R., Nan, X.: Survivable multipath routing and spectrum allocation in OFDM-based flexible optical networks. J. Opt. Commun. Netw. **5**(3), 172–182 (2013)
16. James Sterbenz, P.G., Hutchison, D., et al.: Resilience and survivability in communication networks: strategies, principles, and survey of disciplines. Comput. Netw. **54**(8), 1245–1265 (2010)
17. Beshir, A., Kuipers, F., Van Mieghem, P., Orda, A.: On-line survivable routing in WDM networks. In: 21st International Teletraffic Congress, Paris, pp. 1–8 (2009)
18. Cormen, T.H., Leiserson, C.E., et al.: Introduction to Algorithms, 3rd ed. PHI Learning Pvt. Ltd. (2009)
19. Gurzì, P., Nowé, A., Colitti, W., Steenhaut, K.: Maximum flow based routing and wavelength assignment in all-optical networks. In: 2009 International Conference on Ultra Modern Telecommunications & Workshops, St. Petersburg, pp. 1–6 (2009)
20. Goldberg, A.V., Tarjan, R.E.: A new approach to the maximum-flow problem. J. ACM **35**(4), 921–940 (1988)
21. Boykov, Y., Kolmogorov, V.: Experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Pattern Anal. Mach. Intell. **26**(9), 1124–1137 (2004)

Impact of Feature Selection on EEG Based Motor Imagery

Mridu Sahu and Sneha Shukla

Abstract An EEG based motor imagery translates the motor intention of any subject into control signal by classifying EEG data of different imagination tasks such as hand and feet movements. As indicated by study, it is found that there are almost around 1 in 50 individuals living with loss of motion roughly 5.4 million individuals. For this sort of inability, EEG based BCI for motor imagery of right hand and feet movement imagination is acquired and classified. Short time Fourier transform and wavelet features are extracted and classified with and without feature selection. Ranking method is used for feature selection. Both classification outcomes are comparatively analyzed and observed that there is an increment in classification accuracy when features are classified after feature selection.

Keywords Motor imagery · Electroencephalography · Short time Fourier transform · Wavelet · Feature selection

1 Introduction

Motor imagery is the activation of neural system, where there is an imagination of body movement without physically performing that movement. In other words, motor imagery is motion in imagination and mental rehearsal of a motor act without any overt motor output. An electroencephalography (EEG) based brain–computer interface (BCI) system is used to analyze and operate the transient EEG changes during different types of motor imagery (like imagination of left hand or right hand or foot movement) [1]. Brain–computer interface (BCI) enables the brain signals (or EEG) to communicate with computer or other external device without using any peripheral nerve. This technology has been converted into various everyday life

M. Sahu (✉) · S. Shukla
Department of IT, National Institute of Technology, Raipur, Raipur, Chhattisgarh, India
e-mail: mrisahu.it@nitrr.ac.in

S. Shukla
e-mail: sneha.shukla989@gmail.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_73

applications from laboratory research which are widely involved in clinical fields. Neurorehabilitation is one of the clinical BCI applications related to neurological disorder (such as stroke) suffered patients. Motor imagery becomes fruitful for stroke patients because in the absence of any motor function, they can imagined or even attempt to move their plegic hand [2]. Motor imagery has been mostly used and studied with functional MRI (fMRI) techniques. This technique compares the functional anatomy of nervous system of motor execution and imagery with the help of a task that assesses imagery performance [3]. BCI requires brain signals from the multiple site of scalp to achieve good performance but applying multiple EEG channels may include noisy and redundant signals or artifacts that degrade the BCI performance. As a result of various researches, it has been observed that central EEG channels C_3, C_Z, and C_4 give good accuracy [4]. Feature selection is one of the best approaches for performance improvement of any device or classifier [5].

1.1 Literature Review

Motor imagery can be described as an imagination of movement also known as motor action without any neuromuscular activity. Thought of movement in brain is intimately linked with action. Many researches and studies have already been done in this field. Townsend et al. [6] proposed a work on continuous EEG classification during motor imagery simulation of an asynchronous BCI. Qin and He [7] proposed a work on a wavelet-based time–frequency analysis approach for classification of motor imagery for BCI applications. Wang et al. [8] presented common spatial pattern method for channel selection in motor imagery-based brain–computer interface. Herman et al. [9] comparatively analyzed the spectral approaches to feature extraction for EEG based motor imagery classification. Bhattacharyya et al. [10] studied and worked on automatic feature selection of motor imagery EEG signals using differential evolution and learning automata.

1.2 Overview and Flow Diagram of Proposed Method

The presented article shows comparative analysis between changes in the EEG classification performance with and without feature selection technique. Refer to Fig. 1, EEG signals are acquired and preprocessed in the first signal acquisition stage. This section includes noise removal, amplification, and digitization of EEG signal. EEG data, which is taken from BNCI Horizon 2020, is now segmented into multiple parts.

In the feature extraction stage, different features such as wavelet, short time Fourier transform, statistical are extracted by applying different extraction methods on each part. After combining all the extracted features, these two steps will be followed:

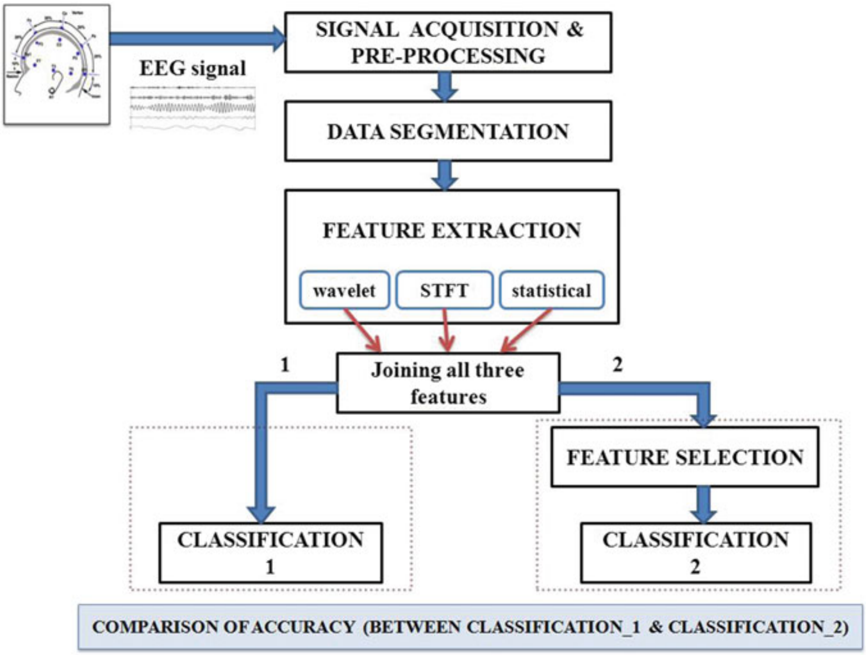


Fig. 1 Flow diagram of proposed method

- Classify the extracted features using different classifiers and get accuracy.
- Apply feature selection method on extracted features and then classify the selected features and get accuracy.

At the end, both classification outcomes have been compared.

2 Materials and Techniques

2.1 *Dataset*

For signal acquisition, the data is taken from BNCI Horizon 2020. It is based on cue-guided Graz BCI training paradigm [1]. Hence, within a single session recording, training, and feedback was performed. Each session consist of eight runs, in which five of them are taken for training and three with feedback for validation. There was a composition of 20 trials for each run. By taking them together, total record of 100 trials per class for training and 60 trials per class for validation as shown in Fig. 2.

As per the cue instruction, participants had the task of performing sustained (5 s) kinesthetic motor imagery (MI) of the right hand and of the feet. At a sampling

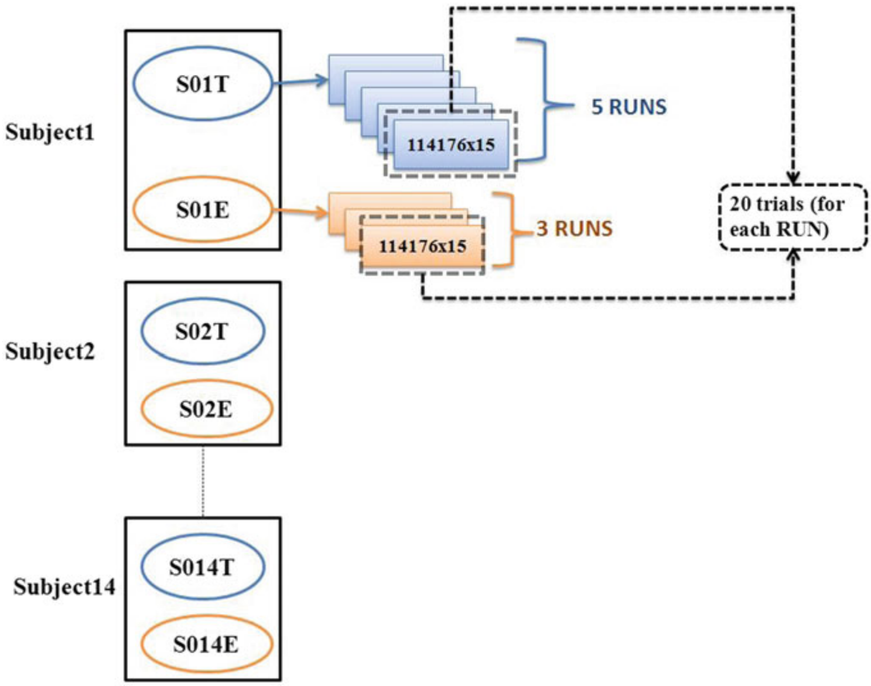


Fig. 2 Dataset description

frequency of 512 Hz, EEG signal is amplified and measured with active Ag/AgCl electrodes. The electrodes placement is based on 10/20 system. According to this, center electrodes at positions C_3, C_Z, and C_4 are selected with a distance of 2.5 cm, out of 15 electrodes total. The reference electrode was mounted on the left mastoid and the ground electrode on the right mastoid [11].

2.2 *Feature Extraction*

Feature extraction is a technique to extract features from the recorded and segmented EEG data. These features are intended to be informative and non-redundant, which leads to better human interpretation. This technique is related to dimensionality reduction. In the presented work, short time Fourier transform, wavelet, and six statistical features are extracted from the measured EEG dataset.

Short time fourier transform features. Fourier transform is one of the best techniques for spectral analysis, which decomposes the time domain signal into frequency components and vice versa [12], but it does not provide time and frequency localization at the same time. Hence, it is not useful for non-stationary signal analysis. To overcome this problem, Short Time Fourier Transform (STFT) is used, in which Fourier transform of windowed signal (original signal multiplied by window function) is considered. It is also called as windowed Fourier or Gabor transform. STFT is a technique in which non-stationary signal (EEG) is segmented into narrow time intervals (enough to be considered stationary) and take Fourier transform of each segment.

Mathematically given as,

$$STFT_f(\omega, \tau) = \int \mathrm{w}(t-\tau) f(t) e^{-j\omega t} dt \tag{1}$$

where w is windowing function and f(t) is time domain signal [13]. STFT coefficient is calculated by spectrogram function, which gives approximated magnitude of input signal. STFT is frequently used for speech processing applications [14].

Wavelet features. Wavelet method is another approach for analyzing non-stationary signal. The term "wavelet" was first used by Alfred Haar in 1909, and wavelet concept was first proposed by Jean Morlet [15]. Wavelet transform decomposes the input signal into set of building blocks or basis functions, termed as wavelet. These wavelets are obtained from a prototype or single generating function (called as mother wavelet) by scaling and shifting operations [16]. There are two types of wavelet used for signal processing, continuous and discrete.

- **Continuous Wavelet Transform** (**CWT**) is used for continuous variation of both shifting and scaling factor. Mathematically expressed as,

$$CWT(x, y) = \int_{-\infty}^{+\infty} z(t) \cdot \varphi^{*}_{x,y}(t) dt \tag{2}$$

where z(t) is an unprocessed EEG signal. x is a dilation factor and y is a translation factor. The term $\varphi^{*}_{x,y}(t)$ is given by,

$$\varphi^{*}_{x,y}(t) = \frac{1}{\sqrt{x}} \cdot \varphi\left(\frac{t-y}{x}\right) \tag{3}$$

where $\varphi(t)$ is wavelet [17].

- **Discrete Wavelet Transform** (**DWT**) uses finite impulse response filter to decompose the signal into approximation and detail coefficients. Segmented EEG data is applied to the low-pass and high-pass filter and down sampled by two as shown in Fig. 3. "Haar" is a discrete wavelet transform used in this proposed article [17].

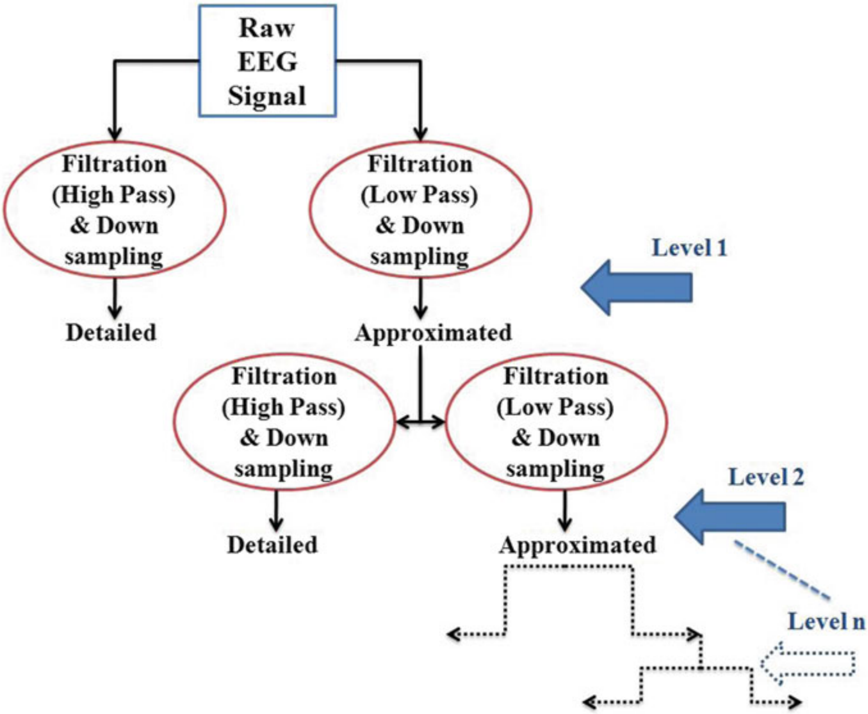


Fig. 3 DWT method up to "n" level

Statistical features. In order to reduce the dimension of input signal, statistical analysis is performed over the segmented data. The statistical feature represents the time–frequency distribution of EEG signal. Following six features were extracted in the proposed method [18].

- **Max Peak** is maximum amplitude of any signal.
- **Min Peak** is minimum amplitude of any signal.
- **Peak to Peak** is the difference between maximum and minimum amplitude of any signal.
- **Mean** is the average amplitude of any signal over given time interval.
- **Median** is measure of the center of any signal, or it is the value at which half of the observations in the sample have values smaller than the median and half have values larger than the median.
- **RMS** is root mean square or effective value of any signal over varying frequency or any other non-stationary parameter.

2.3 Feature Selection

Feature selection is a process to remove those extracted features of the data, which are irrelevant or redundant with respect to the task, to be performed. The advantage of selecting most relevant feature can include the increment in classification accuracy as well as reduction in data dimension. "Supervised" feature selection is focused on classification with labeled data (availability of class information), whereas "Unsupervised" feature selection is focused on clustering with unlabeled data [19]. In this proposed work, ranking method is used for feature selection. On the basis of different ranking methods, features are ranked from higher to lower priority and selection is performed according to the ranking list. Two types of feature ranking methods are applied on the extracted features for feature selection.

Correlation based ranking. Correlation method ranked the features on the basis of "how much they are highly correlated with the class and low intercorrelated with each other," from higher ranking to lower ranking. Correlation based feature selection involves evaluation of subsets of attributes rather than individual attribute [20]. There are basically two types of correlation related to the features and the class, named as C-correlation and F-correlation.

- **C-correlation** involves correlation between any feature and class. This is also known as "Feature–Class" correlation.
- **F-correlation** involves correlation between any pair of features. This is also known as "Feature–Feature" correlation.

Aiming to achieve increased accuracy, C-correlation is calculated for each feature and select those relevant features which are highly correlated with the class [21].

InfoGain-based ranking. InfoGain or information gain ranked the features on the basis of information gain values, which varies from 0 (no information) to 1 (maximum information). Attribute with higher information gain value is taken in the higher priority list because of contributing more information, whereas those attribute that contribute less information (lower Information gain value) is taken in the lower priority list [22]. Let us assume that a set of classes $Y = \{y_1, y_2, y_3, \ldots, y_k\}$ is taken with the information gain IG(k) of feature k and entropy H, then IG(k) can be expressed as,

$$IG(k) = H(Y) - H\left(\frac{Y}{k}\right) \tag{4}$$

where

$$H(Y) = -\sum_{n=1}^{m} P(y_n) \cdot log\, P(y_n) \tag{5}$$

and

$$H\left(\frac{Y}{k}\right) = -P(k)\sum_{n=1}^{m} P\left(\frac{y_n}{k}\right)\cdot log\, P\left(\frac{y_n}{k}\right) \tag{6}$$

where P defines probability. This method is useful for handling high dimensional data.

Real life applications of feature selection involve text categorization, image retrieval, intrusion detection, customer relationship management, etc. [19].

3 Experiment and Result

The proposed method involves evaluation of EEG signal with motor imagery. For this raw data has been taken from BNCI Horizon 2020 [23]. The data consists of signals acquired from 14 subjects with 15 electrodes each. In this presented article, only four subjects with central electrodes or channels (C_3, C_Z, and C_4) have been taken. Each subject is divided into training (3 runs) and testing (5 run) part. Each run consists of data, trial, and class value. The class labels 1 and 2 indicate right hand and feet movement, respectively. Data is segmented into 20 trials for each channel. In the next phase, feature extraction is applied on each segmented data. MATLAB R2016a software platform is used for segmentation and feature extraction. Wavelet features are extracted using "Haar" discrete wavelet transform and detail (Ed1, Ed2, Ed3, Ed4, Ed5) and approximate (Ea) coefficients are calculated. Statistical features were obtained using sptool. Short time Fourier transform has been calculated using spectrogram function which gives absolute value of frequency. Classification on these extracted features was performed using two approaches. In the first case, classification is applied on all 39 extracted features using 10 fold cross-validation which gives the accuracy for all 4 subjects as shown in Table 1.

In the second case, 25 features are selected out of 39 for classification which gives the accuracy as shown in Tables 2 and 3. Features are selected using correlation and InfoGain-based ranking method. After that comparatively analyze both the classification results (refer to Fig. 4), Weka 3.8 software platform is used for feature selection and classification. The proposed method compares classification accuracy before and after feature selection and shows that there is an increment in accuracy after feature selection.

Each subject has taken into account and observed their performance individually. Figure 5 shows the best performance of each subject with their graphical view. Classification accuracy is increased after feature selection (Fig. 5).

Table 1 Classification on extracted features of four subjects before feature selection

Classifier	Subject 1			Subject 2			Subject 3			Subject 4			Class
	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	
1. Naive Bayes	49.375	0.700	0.713	60.625	0.688	0.475	45.625	0.775	0.863	54.375	0.313	0.225	1
		0.288	0.300		0.525	0.313		0.138	0.225		0.775	0.688	2
2. Logistic	51.875	0.538	0.500	50.625	0.538	0.525	45.000	0.500	0.600	55.000	0.563	0.463	1
		0.500	0.463		0.475	0.463		0.400	0.500		0.538	0.438	2
3. J48 tree	48.125	0.775	0.813	55.000	0.750	0.650	46.875	0.813	0.875	43.750	0.350	0.475	1
		0.188	0.225		0.350	0.250		0.125	0.188		0.525	0.650	2
4. LogitBoost	56.250	0.525	0.400	56.875	0.613	0.475	51.250	0.588	0.563	50.000	0.525	0.525	1
		0.600	0.475		0.525	0.388		0.438	0.413		0.475	0.475	2
5. Fine KNN	54.375	0.550	0.463	56.250	0.638	0.513	51.250	0.588	0.563	61.875	0.638	0.400	1
		0.538	0.450		0.488	0.363		0.438	0.413		0.600	0.363	2
6. AdaBoost MI	51.875	0.425	0.388	57.500	0.713	0.563	50.625	0.563	0.550	50.875	0.600	0.563	1
		0.613	0.575		0.438	0.288		0.450	0.438		0.438	0.400	2
7. LWL	55.000	0.513	0.413	53.125	0.800	0.738	52.500	0.600	0.550	46.875	0.400	0.463	1
		0.588	0.488		0.263	0.200		0.450	0.400		0.538	0.600	2
8. Bagging	47.500	0.500	0.550	59.375	0.650	0.463	49.375	0.513	0.525	49.375	0.488	0.500	1
		0.450	0.500		0.538	0.350		0.475	0.488		0.500	0.513	2
9. Random committee	47.500	0.588	0.638	53.750	0.650	0.575	47.500	0.575	0.625	46.250	0.525	0.600	1
		0.363	0.413		0.425	0.350		0.375	0.425		0.400	0.475	2
10. Iterative classifier optimizer	47.500	0.838	0.888	48.750	0.938	0.963	46.875	0.825	0.888	49.375	0.950	0.963	1
		0.113	0.163		0.038	0.063		0.113	0.175		0.038	0.050	2

Table 2 Classification on extracted features of four subjects after feature selection (using correlation method)

Classifier	Subject 1			Subject 2			Subject 3			Subject 4			Class
	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FPrate	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	
1. Naive Bayes	49.750	0.625	0.638	61.250	0.650	0.425	47.500	0.800	0.850	54.875	0.413	0.375	1
		0.363	0.375		0.575	0.350		0.150	0.200		0.625	0.588	2
2. Logistic	46.875	0.475	0.538	56.875	0.575	0.438	48.125	0.475	0.513	56.875	0.550	0.413	1
		0.463	0.525		0.563	0.425		0.488	0.525		0.588	0.450	2
3. J48 tree	48.750	0.800	0.825	55.750	0.775	0.700	47.500	0.813	0.863	44.375	0.413	0.525	1
		0.175	0.200		0.300	0.225		0.138	0.188		0.475	0.588	2
4. LogitBoost	56.250	0.450	0.450	58.125	0.600	0.438	51.625	0.550	0.538	51.250	0.500	0.475	1
		0.550	0.550		0.563	0.400		0.463	0.450		0.525	0.500	2
5. Fine KNN	54.500	0.525	0.475	58.125	0.675	0.513	51.875	0.538	0.500	62.875	0.600	0.438	1
		0.525	0.475		0.488	0.325		0.500	0.463		0.563	0.400	2
6. AdaBoost MI	53.750	0.463	0.388	60.625	0.788	0.575	56.875	0.625	0.488	51.250	0.375	0.413	1
		0.613	0.538		0.425	0.213		0.513	0.375		0.588	0.625	2
7. LWL	55.000	0.438	0.388	54.375	0.813	0.725	55.625	0.650	0.538	47.750	0.275	0.400	1
		0.613	0.563		0.275	0.188		0.463	0.350		0.600	0.725	2
8. Bagging	48.125	0.488	0.525	59.375	0.663	0.475	51.250	0.513	0.488	50.625	0.500	0.488	1
		0.475	0.513		0.525	0.338		0.513	0.488		0.513	0.500	2
9. Random committee	52.500	0.613	0.563	59.375	0.713	0.525	47.500	0.575	0.663	46.875	0.513	0.575	1
		0.438	0.388		0.475	0.288		0.338	0.425		0.425	0.488	2
10. Iterative classifier optimizer	47.500	0.838	0.888	50.000	0.938	0.500	48.125	0.825	0.863	49.500	0.850	0.900	1
		0.113	0.163		0.063	0.500		0.138	0.175		0.100	0.150	2

Table 3 Classification on extracted features of four subjects after feature selection (using InfoGain method)

Classifier	Subject 1			Subject 2			Subject 3			Subject 4			Class
	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	Accuracy (%)	TP rate	FP rate	
1. Naive Bayes	49.375	0.65	0.675	60.625	0.638	0.438	45.875	0.750	0.863	57.500	0.313	0.163	1
		0.325	0.35		0.563	0.363		0.138	0.250		0.838	0.688	2
2. Logistic	55.625	0.525	0.413	55.000	0.538	0.438	48.125	0.475	0.513	55.625	0.575	0.463	1
		0.588	0.475		0.563	0.463		0.488	0.525		0.538	0.425	2
3. J48 tree	50.000	0.825	0.825	56.250	0.788	0.663	48.125	0.838	0.875	43.750	0.300	0.450	1
		0.175	0.175		0.338	0.213		0.125	0.163		0.550	0.700	2
4. LogitBoost	56.875	0.575	0.438	58.125	0.575	0.413	52.500	0.588	0.538	50.000	0.550	0.575	1
		0.563	0.425		0.588	0.425		0.463	0.413		0.425	0.450	2
5. Fine KNN	59.375	0.588	0.4	56.250	0.613	0.488	57.500	0.613	0.463	60.000	0.550	0.525	1
		0.6	0.413		0.513	0.388		0.538	0.388		0.475	0.450	2
6. AdaBoost MI	58.125	0.538	0.375	61.250	0.725	0.500	53.750	0.513	0.438	51.250	0.463	0.438	1
		0.625	0.463		0.500	0.275		0.563	0.488		0.563	0.538	2
7. LWL	55.375	0.488	0.43	53.500	0.800	0.750	52.500	0.675	0.650	50.625	0.438	0.425	1
		0.6	0.51		0.250	0.200		0.350	0.325		0.575	0.563	2
8. Bagging	48.125	0.5	0.538	59.500	0.575	0.425	50.625	0.525	0.513	46.250	0.488	0.563	1
		0.463	0.5		0.575	0.425		0.488	0.475		0.438	0.513	2
9. Random committee	48.125	0.538	0.575	53.750	0.588	0.513	47.875	0.600	0.663	46.250	0.588	0.700	1
		0.425	0.463		0.488	0.413		0.338	0.400		0.300	0.413	2
10. Iterative classifier optimizer	50.625	0.9	0.888	49.000	0.838	0.888	48.125	0.863	0.900	49.500	0.800	0.850	1
		0.113	0.1		0.113	0.163		0.100	0.138		0.150	0.200	2

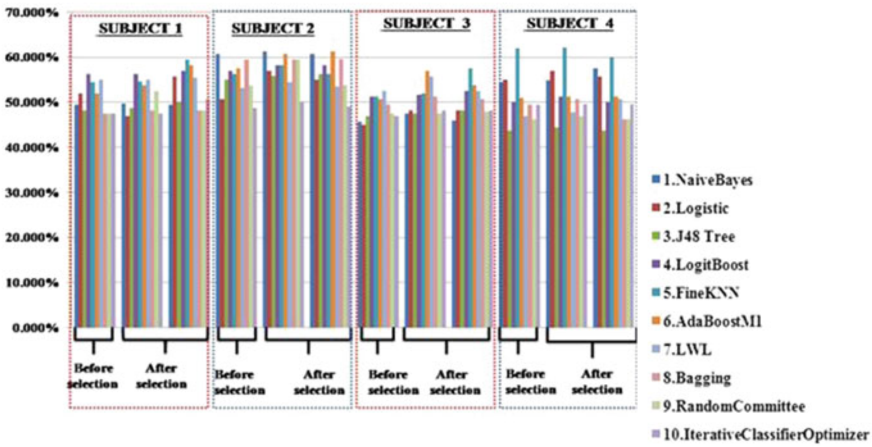


Fig. 4 Classification accuracy of four subjects using 10 different classifiers before and after feature selection

	CLASSIFIER	ACCURACY (before selection)	INCREASED ACCURACY (after selection)	SELECTION METHOD
Subject 1	Fine KNN	54.375%	59.375%	InfoGain Evalution
Subject 2	Naïve Bayes	60.625%	61.250%	Correlation Evalution
	AdaBoost M1	57.500%	61.250%	InfoGain Evalution
Subject 3	Fine KNN	51.250%	57.500%	InfoGain Evalution
Subject 4	Fine KNN	61.875%	62.875%	Correlation Evalution

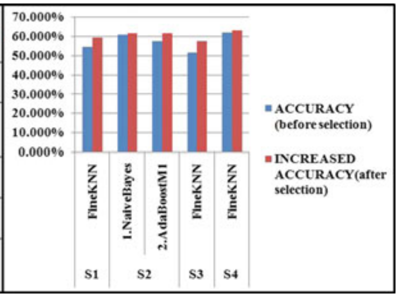


Fig. 5 Best performance of each subject with graphical view

4 Conclusion

Brain is one of the most important organs of the body. So, its structure and functions have become a great source for the research. Although a lot of work has already been done in this field, it is still a challenging area. Motor imagery study focused on motor skill, performance, or strength improvement by different techniques. Analyzing the effect of feature selection on motor imagery data with extreme precision was the goal of research. For achieving this, various extraction, selection, and classification techniques have been studied. For better classification accuracy, those features have been selected that can be useful for giving best result. The increased accuracy has been obtained when feature selection is used for classification. The accuracy is increased up to 62.875% with fine KNN classifier and correlation based selection method for subject 4 data. In future work, accuracy will be further increased by using some additional feature extraction and selection techniques.

Acknowledgements This work was supported by a grant from the National Institute of Technology, Raipur. The authors acknowledge the Head of Department of Information Technology, National Institute of Technology, Raipur. The authors also acknowledge the constructive criticisms by several anonymous reviewers of an earlier draft of this paper.

References

1. Pfurtscheller, G., Neuper, C.: Motor imagery and direct brain-computer communication. Proc. IEEE **89**(7), 1123–1134 (2001)
2. Ang, K.K., Guan, C., Chua, K.S. G., Ang, B.T., Kuah, C., Wang, C., Zhang, H.: Clinical study of neuro rehabilitation in stroke using EEG-based motor imagery brain-computer interface with robotic feedback. In: Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE, pp. 5549–5552. IEEE, Aug 2010
3. Hanakawa, T., Immisch, I., Toma, K., Dimyan, M.A., Van Gelderen, P., Hallett, M.: Functional properties of brain areas associated with motor execution and imagery. J. Neurophysiol. **89**(2), 989–1002 (2003)
4. Arvaneh, M., Guan, C., Ang, K.K., Quek, C.: Optimizing the channel selection and classification accuracy in EEG-based BCI. IEEE Trans. Biomed. Eng. **58**(6), 1865–1873 (2011)
5. Brunner, C., Naeem, M., Leeb, R., Graimann, B., Pfurtscheller, G.: Spatial filtering and selection of optimized components in four class motor imagery EEG data using independent components analysis. Pattern Recogn. Lett. **28**(8), 957–964 (2007)
6. Townsend, G., Graimann, B., Pfurtscheller, G.: Continuous EEG classification during motor imagery-simulation of an asynchronous BCI. IEEE Trans. Neural. Syst. Rehabil. Eng. **12**(2), 258–265 (2004)
7. Qin, L., He, B.: A wavelet-based time-frequency analysis approach for classification of motor imagery for brain computer interface applications. J. Neural. Eng. **2**(4), 65 (2005)
8. Wang, Y., Gao, S., Gao, X.: Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the (pp. 5392–5395). IEEE, Jan 2006
9. Herman, P., Prasad, G., McGinnity, T.M., Coyle, D.: Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Trans. Neural. Syst. Rehabil. Eng. **16**(4), 317–326 (2008)
10. Bhattacharyya, S., Sengupta, A., Chakraborti, T., Konar, A., Tibarewala, D.N.: Automatic feature selection of motor imagery EEG signals using differential evolution and learning automata. Med. Biol. Eng. Comput. **52**(2), 131–139 (2014)
11. Steyrl, D., Scherer, R., Frstner, O., Mller-Putz, G.R.: Motor imagery brain-computer interfaces: random forests vs regularized LDAnon-linear beats linear. In Proceedings of the 6th International Brain-Computer Interface Conference, Sept 2014
12. Canal, M.R.: Comparison of wavelet and short time Fourier transform methods in the analysis of EMG signals. J. Med. Syst. **34**(1), 91–94 (2010)
13. Gorur, D., Halici, U., Aydin, H., Ongun, G., Ozgen, F., Leblebicioglu, K.: Sleep spindles detection using short time Fourier transform and neural networks. In: Proceedings of the 2002 International Joint Conference on Neural Networks, IJCNN'02, vol. 2, pp. 1631–1636. IEEE (2002)

14. Almeida, L.B.: The fractional Fourier transform and time-frequency representations. IEEE Trans. Signal Process. **42**(11), 3084–3091 (1994)
15. Akin, M.: Comparison of wavelet transform and FFT methods in the analysis of EEG signals. J. Med. Syst. **26**(3), 241–247 (2002)
16. Cvetkovic, D., Beyli, E.D., Cosic, I.: Wavelet transform feature extraction from human PPG, ECG, and EEG signal responses to ELF PEMF exposures: a pilot study. Digit. Signal Proc. **18**(5), 861–874 (2008)
17. Soltani, S.: On the use of the wavelet decomposition for time series prediction. Neurocomputing **48**(1), 267–277 (2002)
18. Faust, O., Acharya, U.R., Adeli, H., Adeli, A.: Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure **26**, 56–64 (2015)
19. Liu, H., Yu, L.: Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng. **17**(4), 491–502 (2005)
20. Hall, M.A., Holmes, G.: Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans. Knowl. Data Eng. **15**(6), 1437–1447 (2003)
21. Yu, L., Liu, H.: Efficient feature selection via analysis of relevance and redundancy. J. Mach. Learn. Res. **5**(Oct) 1205–1224 (2004)
22. https://machinelearningmastery.com/perform-feature-selection-machine-learning-data-weka/
23. http://bnci-horizon-2020.eu/database/data-sets

Various Machine Learning Algorithms for Twitter Sentiment Analysis

Rishija Singh and Vikas Goel

Abstract In present time, the use of social networking Web site has increased rapidly. Social networking site, Twitter, produces a lot of opinion data these days. Opinion analysis is mainly concerned with the analysis of feelings and opinions of the text. Opinion mining is an existing field of research in the text mining. This article focuses on various techniques of machine learning that are used in opinion analysis and opinion exploration. In this paper, we have studied various machine learning techniques and compared various machine learning algorithms like support vector machine, Naive Bayes, and maximum entropy with their evaluation parameters: precision, accuracy, recall, and f-measure. During analysis of techniques, there are a number of problems and challenges that must be addressed. These issues are also presented in this paper.

Keywords Opinion analysis · Machine learning · Supervised learning

1 Introduction

Emotional analysis is described in the form of determining the concept of people about different institutions. Today, people use social sites to review and comment on this product, known as consumer ideas, emotions, emotions, behavior, or thinking. The opinion is a way to find ways in which the ideas are expressed. The opinion is typically analyzed at three levels: the level of document, the level of sentence and the level of aspect [1]. Document level arranges whether the records conclusion is positive, negative or neutral. It considers the entire document a basic

R. Singh (✉) · V. Goel
Department of CS&E, Ajay Kumar Garg Engineering College, Ghaziabad, India
e-mail: Singh.rishija@gmail.com

V. Goel
e-mail: rvikasgoel@yahoo.com

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3_74

data unit (talking about one topic). Sentence level decides if the sentence communicates any negative, positive or impartial assessment. Sentence level SA means to arrange sentiment expressed in each sentence. The initial step is to recognize whether the sentence is subjective or objective. Aspect level spotlights on all statements of sentiment present inside document and the aspect to which it refers. There is no key contrast amongst document and sentence level characterizations since sentences are simply short records. Classifying content at the document level or at the sentence level does not give the fundamental detail required conclusion on all parts of the element which is required in numerous applications, to obtain these details; we need to go to the aspect level. Aspect level SA aims to classify the opinion with respect to the specific aspects of entities. Revenue owners can give different ideas on different aspects of the same unit. Therefore the purpose of the SA is to take care of these ideas and then classify their differences [2].

2 Sentiment Analysis Techniques

Passion analysis means that natural language processing, text analysis, statistical linguistics and biometric use, diagnostic states, and identify identities, removal, and studying. There are some techniques for analyzing emotions:

2.1 Machine Learning Approach

Machine learning techniques use several series to learn and classify. The training set includes the names of the input feature vectors and their respective class. Using this training set, a classification model has been developed so that the input facility can be classified into week class texts. After this, a test set is used to validate the model by predicting the posts of the invisible feature vectors. In our ML approach, text classification methods can be divided into almost unethical-false, supervised, and semi-supervisor [3].

2.2 Unsupervised Learning

The main objective of text rating is to rate the documentation in many predefined types. To do so, a large number of writing material is inspected. In text rating, these labeled training materials may sometimes be difficult, but it is easy to collect poor documentation. Methods of jumbled learning can stun these hitches [4].

2.3 Supervised Learning

Supervised machine learning method acquaintance with the usage of a noticeable article set to retain some classification purpose and comprises learning of function from the experiment along with its input and output [5]. Controlled learning is task of assuming a function labeled trained data set. Training data set includes set of training examples; each and all example consists of pair of an input data as well as predictable output [6].

2.3.1 Decision Tree Classifier

In the decision tree classification, the inner node is marked with the features, and the left edges of the nodes have been designated as a test on the data set weight [5]. The decision of decision-making tree has been estimated that as a result of the expected value of this element, he shares an element of information.

2.3.2 Linear Classifier

To classify the input vector in the classical classroom in linear classical, they use linear dairy shaped margins. There are several types of linear classical support vector machines, one of them. This rating offers a good linear spread between the different [5].

- Support Vector Machine

Support vector machine (SMS) is known as the best classifier, which provides the most accurate results in speech rating issues. After getting a hyperplane with maximum distance for the closest trained example, he realized that according to the rest of the training data, there is no access to the eligibility rating, because the rating of SVM has successfully been implemented for the purpose of text-class purposes and is used in various settings.

The SMM classifier uses large differences for classification. It separates tweets with hyperplane; SVM uses a defined function.

$$\mathrm{g(X)} = w^{T}\phi(\mathrm{X}) + \mathrm{b} \tag{1}$$

The x feature is vector, w is the weight vector, and b is the bias vector. W training and B training are automatically learned on the set [3].

- Neural Network

The neural network consists of several neurons in which the neuron is its primary unit. The multi-level neural network was used with online margins; the results of the neuron in the previous layer are given as input to the next layer. In such a classifier, creating a data set is more complex, as the defects should be propagated for different layers [5].

2.3.3 Rule-Based Classifiers

In rule-based classical, the database is created with a set of rules. Shows a bet on the feature set described in the lower position on the left, while labeled on the right side labeled. Terms are on the word presence; word shortage is used at least because it is not informative in scattered data; there are several standards for making laws, two most commonly supported support and trust [7].

2.3.4 Probabilistic Classifier

This classification uses a variety of ratings, as part of this mixture, there are different types of varieties, and all the different elements are productive models, in which a different word for this element. It is necessary to observe. The probability is that this classifier is also known as the ingredient rating; some potential sizes are equally unique, bio-networks, and maximum entropy capabilities [5].

- Naive Bayes Classifier (NB).

The Naive Bayes classifier is the easiest and most used rating, which works with the origin of the model boot function, which ignores the condition of the word in the document. Without the repeated evaluation of the current parameter, it is easy to use the comfortable beige model, which makes it particularly interesting for larger data. The unconditional possibility for allegiance can be described as probable;

$$P(\mathrm{X}|\mathrm{y_j}) = \prod_{\mathrm{i}=1}^{\mathrm{m}} P(\mathrm{x_i}|\mathrm{y_j}) \tag{2}$$

The X attribute is the vector, which is called x = fx1, x2, xmg, and y_j class tag [8]. The main advantage of the Navy Bayes classifier is that it analyzes each function independently; therefore, this feature uses all the features in the vector.

- Maximum Entropy (ME)

Probability distribution estimation is based on technique for doing maximum entropy classification. In this technique, the first classified feature sets are converted to defined vectors using one of the encoding schemes. Second, the use of this encoded vector is used to calculate the weight for each extracted characteristics, which can collectively support it to determine the most probable label for a set of attributes. It depends on a possible approach like Naive Bayes. The fundamental concept of maximum entropy is that if more information about the data is not known then distribution should every similar

$$P_{\mathrm{ME}}(c|d) = \frac{\exp[\sum_i \lambda_i f_i\,(c,d)]}{\sum_{c'} \exp[\sum_i \lambda_i f_i\,(c,d)]} \quad (3)$$

where is c the class, d is color, and is a weight vector. Weight vectors decide the importance of a feature in the classification.

3 Evaluating Factors of Sentiment Classification

The classification of opinion can be evaluated using four indexes calculated on the basis of the following equations: accuracy, precision, recall, and f-measure [9].

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (4)$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (5)$$

$$\text{F Measurement} = 2 * \text{ Precision } * \text{ Recall}/(\text{Precision} + \text{Memory}) \quad (6)$$

$$\text{Accuracy} = \text{TP} + \text{TN}/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (7)$$

4 Analyses and Comparative Study

In this section, we are analyzing various algorithms w.r.t. data set, precision, recall, f-measure, and accuracy for accomplishing sentiment analysis. Table 1 depicts evaluating parameters: precision, recall, and f-measure of the SVM sentiment classifier. Various authors have worked on support vector machine (SVM) classifier. However, Abhinash Tripathy et al. have worked on movie data set. The work of Abhinash et al. is more precise with precision rate 0.88, and their accuracy is also better. In case of twitter data, the work of Bac Le et al. is more precise with

Table 1 Precision, recall, and f-measure of the SVM sentiment classifier

Classifier	Authors	Data set	Precision	Recall	F-measure	Accuracy
Support vector machine (SVM)	Bac Le et al.	Twitter	0.79	0.79	0.79	79.5
	Abinash Tripathy et al.	Movie reviews	0.88	0.87	0.88	94
	Geetika Gautam et al.	Twitter	0.39	0.85	0.53	85.5
	Hailong Zhang et al.	Twitter	0.72	0.67	0.70	71.00

precision rate 0.79 and f-measure rate is also better with rate 0.79, but Geetika Gautam et al. have better recall with recall rate 0.85.

Figure 1 depicts graph of precision, recall, and f-measure, and Fig. 2 depicts accuracy of the SVM sentiment classifier of different authors. Abinash Tripathy et al. have high precision, recall, f-measure, and accuracy as compared to others.

Table 2 depicts evaluating parameters: precision, recall, and f-measure of the NB sentiment classifier. Various authors have worked on Naive Bayes (NB) classifier. However, Abhinash Tripathy et al. have worked on movie data set. The work of Abhinash et al. is more precise with precision rate 0.84, and their accuracy is also better. In case of twitter data, the work of Bac Le et al. is more precise with precision rate 0.80 and f-measure rate is also better with rate 0.79, but Geetika Gautam et al. have better recall with recall rate 0.88.

Figure 3 depicts graph of precision, recall, and f-measure, and Fig. 4 depicts accuracy of the Naive Bayes. Work of Abinash Tripathy et al. have high precision, recall, f-measure, and accuracy as compared to other paper.

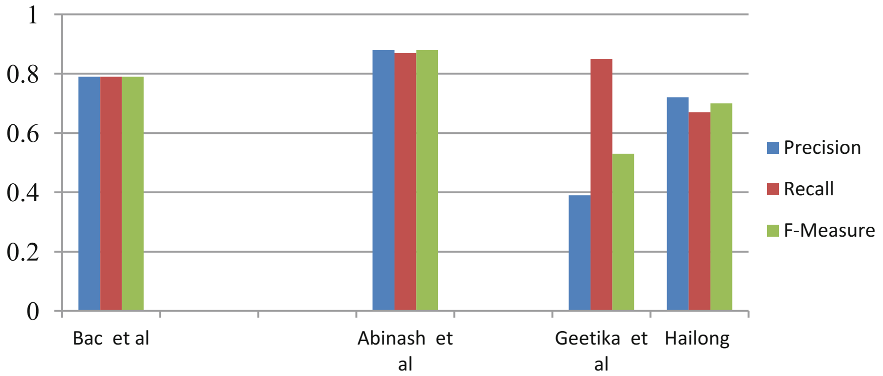


Fig. 1 Graph of precision, recall, and f-measure of the SVM sentiment classifier

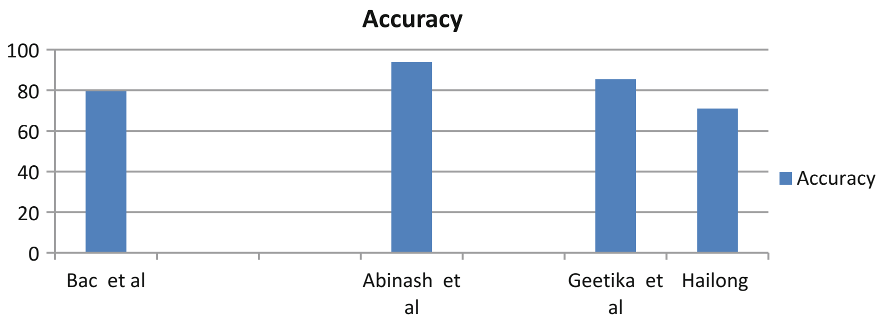


Fig. 2 Graph of accuracy of the SVM sentiment classifier

Table 2 Precision, recall, and f-score of the naive Bayes classifier

Classifier	Authors	Data set	Precision	Recall	F-measure	Accuracy
Naive Bayes (NB)	Bac Le et al.	Twitter	0.80	0.79	0.79	80
	Abinash Tripathy et al.	Movie reviews	0.84	0.83	0.83	89
	Geetika Gautam et al.	Twitter	0.44	0.88	0.58	88.2
	Hailong Zhang et al.	Twitter	0.65	0.78	0.68	68.75

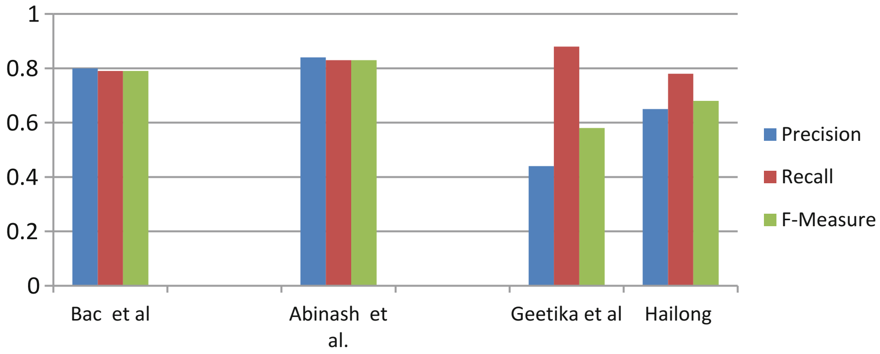


Fig. 3 Graph of precision, recall, and f-measure of the naive Bayes

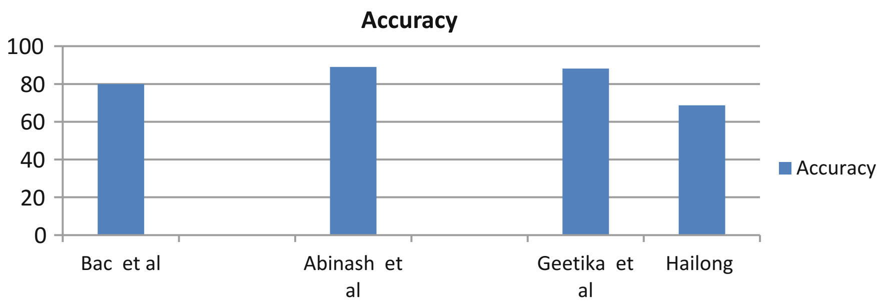


Fig. 4 Graph of accuracy of the naive Bayes

Table 3 depicts precision, recall, and f-measure of the ME sentiment classifier. Various authors have worked on maximum entropy (ME) classifier. However, the work of Neethu M et al. was more precise with precision rate 0.87, and their accuracy was also better in comparison to other.

Table 3 Precision, recall, and f-measure of the maximum entropy

Classifier	Authors	Data set	Precision	Recall	F-measure	Accuracy
Maximum entropy (ME)	Geetika Gautam et al.	Twitter	0.37	0.83	0.51	83.8
	Neethu et al.	Twitter	0.87	0.93	0.89	90

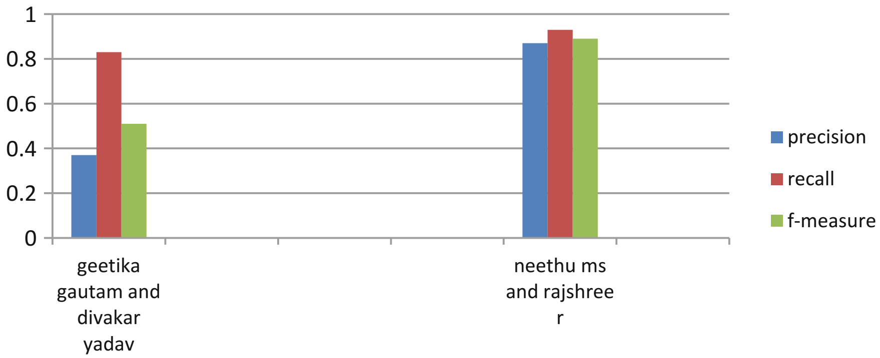


Fig. 5 Graph of precision, recall, and f-measure of the maximum entropy

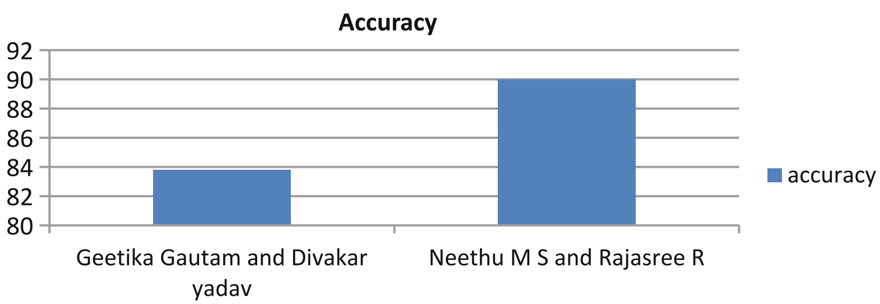


Fig. 6 Graph of accuracy the maximum entropy

Figure 5 depicts graph of precision, recall, and f-measure, and Fig. 6 depicts accuracy of the maximum entropy. Neethu et al. [3] has high precision, recall, f-measure, and accuracy as compared to other paper.

5 Conclusion and Future Work

The text of the emotional analysis (SA) is an existing research area in the mining area. This paper focuses on many machine learning techniques used in the analysis of emotions and opinion-taking. We compared and analyzed support vector machine (SVM), Naïve Bayes (NB), and maximum entropy (ME) with their evaluating parameters: accuracy, precision, recall, and f-measure. We can conclude that, in case of SVM classifier, the work of Bac Le et al. performs better in precision, recall, and f-measure, but accuracy of Geetika Gautam et al. is better. In case of NB classifier, the work of Bac Le et al. performs better in precision, f-measure, but Geetika Gautam et al. work performs better in recall and accuracy. In case of maximum entropy, the work of Neethu et al. performs better than Geetika Gautam et al. work.

In our study, we found that many authors have accuracy issues and trying to solve data inconsistencies and big data. Along with this, we have also found the problem with multilingual tweets and handling techniques. The possibility of future emotion analysis for other languages may be considered. As most research work only addresses monolingual tweets. Therefore, there is a need to analyze all tweets available in different languages.

References

1. Tripathy, A., Agrawal, A., Rath, S.K.: Classification of sentiment reviews using n-gram machine learning approach. Sci. Direct **57**, 117–126 (2016)
2. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014)
3. Neethu, M.S., Rajasree, R.: Sentiment analysis in twitter using machine learning techniques. In: Proceedings of Conference on Communications and Networking Technologies (ICCCNT), pp. 1–5. IEEE (2013)
4. Youngjoong, K., Jungyun, S.: Automatic text categorization by unsupervised learning. In: Proceedings of Conference on Computational Linguistics (COLING-00), vol. 1, pp. 453–459 (2000)
5. Bhavitha, B.K., Rodrigues, A.P., Chiplunkar, N.N.: Comparative study of machine learning techniques in sentimental analysis. In: Proceedings of Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 216–221. IEEE (2017)
6. Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., Liu, B.: Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. Hewlett-Packard Development Company, L.P. (2011)
7. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Proceedings of Conference on Knowledge Discovery and Data Mining (KDDM), pp. 80–86 (1998)
8. Upadhyay, N., Singh, A.: Sentiment analysis on twitter by using machine learning technique. Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) **4**(5), 488–494 (2016)
9. Zhang, H., Gan, W., Jiang, B.: Machine learning and lexicon based methods for sentiment classification: a survey. In: Proceedings of IEEE Conference on Web Information System and Application (WISA), pp. 262–265 (2014)

10. Desai, M., Mehta, M.A.: Techniques for sentiment analysis of twitter data. In: Proceedings of IEEE International Conference on Computing, Communication and Automation (ICCCA), pp. 149–154 (2016)
11. Tripathy, A., Agrawal, A., Rath, S.K.: Classification of sentimental reviews using machine learning techniques. In: Proceedings of International Conference on Recent Trends in Computing (ICRTC), vol. 57, pp. 821–829 (2015)
12. Le, B., Nguyen, H.: Twitter sentiment analysis using machine learning techniques. Adv. Intell. Syst. Comput. (AISC) **358**, 279–289 (2015)
13. Gautam, G., yadav, D.: Sentiment analysis of twitter data using machine learning approaches and semantic analysis. In: Proceedings of Conference on Contemporary Computing (IC3), pp. 437–442. IEEE (2014)

Author Index

A
Agarwal, Manisha, 187
Aggarwal, Gaurav, 399
Agrawal, Subham, 323
Ahuja, Jyoti, 475
Ahuja, Yash, 265
Ajith, K.A.M., 165
Ali, Layak, 571
Arjun, S.M., 165
Arora, Sandhya, 551
Arora, Srishti, 447
Athesh Kumar, S., 165
Aydav, Prem Shankar Singh, 465

B
Bala, Anju, 255
Balaji, V., 341, 351
Bendre, Varsha, 605
Bhalchandra, Parag, 81
Bhargava, Lava, 645
Bhatia, Jitendra, 495
Bhatnagar, Shaleen, 205, 245
Bhavani, Nallamilli P.G., 341, 351
Bhosale, Amarja Hanumant, 89
Bhuiyan, Touhid, 41
Bogale, Deepa, 81

C
Chakraborty, Aruna, 701
Chatterjee, Niladri, 149
Chaubey, V.K., 737
Chaudhari, Rupal, 591
Chavan, Pallavi, 1
Chaware, Gauri, 551
Chinchankar, Devangi, 551
Choudhury, B.B., 303, 323

D
Darkunde, Nitin, 509
Das, Satyabrata, 293
Dave, Meenu, 109
Dixit, Eesha, 551
Dominic, Vinitha, 245
Doshi, Nishant, 439
D'Sa, Ashwin Geet, 65
Dua, Rajat, 561

E
Elanangai, V., 535

F
Fatima, Ramsha, 177

G
Galiyawala, Hiren J., 591
Garg, Ritu, 157, 275
Gawhale, Priyanka, 625
Geetha, G., 419
Ghate, Pravin M., 615
Girija, P.N., 655
Giri, Nupur, 265
Godhani, Vijay, 265
Goel, Vikas, 763
Gorasia, Nikita, 439
Goyal, Garima, 245

H
Hasan Mahmud, S.M., 41
Hossain, Rifat, 41
Hossin, Md. Altab, 41

S. Fong et al. (eds.), *Information and Communication Technology for Competitive Strategies*, Lecture Notes in Networks and Systems 40,
https://doi.org/10.1007/978-981-13-0586-3

Hua, Yin Xiao, 41

I
Iyer, Nalini C., 485

J
Jailia, Manisha, 187
Jain, Bhavesh, 377
Jain, Sarika, 197
Jain, Shevi, 551
Jalla, Hanumantharao, 655
Jayalakshmi, S., 351
Jetawat, Ashok Kumar, 31
Jindal, Poonam, 223, 455
Joe Louis Paul, I., 165
Jyotiyana, Priya, 687

K
Kadiya, Tejas, 377
Kamath, Radhakrishna, 265
Kathpal, Chesta, 275
Kaur, Jasmeet, 255
Kaur, Kiranpreet, 519
Kaur, Sawroop, 419
Khairnar, Vaishali, 31
Khan, Aysha, 235
Khan, Hannan Ahmad, 727
Khan, Muneeb Hasan, 177, 235, 313
Kishore Kumar, K., 581
Kole, Dipak K., 701
Kolhe, Likhesh N., 31
Kore, Shweta, 485
KrishnaKumar, R., 341
Kulkarni, Govind, 81
Kulkarni, Supraja, 571
Kumari, Sushmita, 205
Kureshi, Abdul Kadir, 605
Kushwaha, Dharmender Singh, 99, 121

L
Lama, Vishal, 197

,M
Madan, Mamta, 109
Maheshwari, Saurabh, 687
Maji, Arnab Kumar, 665
Maji, Srabanti, 447
Malathy, E.M., 285
Mangla, Sachin Kumar, 409
Manjrekar, Amrita A., 89
Maryam, Amrah, 313
Minz, Sonajharia, 465
Mishra, Dillip K., 675
Mohanty, Asit, 675
Mondal, Sraboni, 701
Muley, Aniket, 81

N
Nasreen, Syeda Musfia, 665
Nayak, Chinmya Kumar, 293
Nijhawan, Rahul, 561
Nijhawan, Vani Kapoor, 109

O
Oza, Mayur, 439

P
Padole, Dinesh V., 21
Pahuja, Sonika, 455
Panigrahi, Tapas Kumar, 675
Patel, Nehal, 137
Patel, Radhika, 137
Patel, Sandip, 137
Patel, Sanjay, 429
Patil, Arunkumar, 509
Patil, Bharat, 625
Patil, Pravin P., 409
Patil, Yuvaraj S., 131
Pattanashetty, Vishal B., 485
Pavani, Movva, 581
Pinto, Ajay Prinston, 333
Prajapati, Ramesh T., 377
Prasada, Padma, 333
Prasad, B.G., 65

R
Radha, N., 285
Rahul Kumar, 223
Rajak, Biru, 121
RajeshKumar, 377
Rajeshwara Rao, R., 709
Rakshit, Somnath, 701
Ramakrishna, S., 11
Rani, Rama, 157
Ranjith, H.D., 333
Rao, Annavarapu Chandra Sekhara, 633
Rathee, Seema, 475
Ratnoo, Saroj, 475
Ray, Prakash K., 675
Rihan, Mohd., 727
Rishi, Megha, 561
Roberts, Michaelraj Kingston, 721
Rout, Jayprakash, 323

S
Sachdeva, Geeta, 53
Saha, Goutam, 665
Sahu, Mridu, 749

Sahu, Supriya, 303
Sambasiva Rao, N., 359
Sambhe, Vijay K., 625
Sandhya, M., 391
Sangam, Savita, 213
Sankpal, Swati V., 131
Sarkar, Apurba, 701
Sarosh Umar, M., 177, 235, 313
Sasirekha, S., 165
Sathisha, 333
Seema, 53
SenthilKumar, K., 341
Shah, Namra Bhadreshkumar, 495
Shah, Neel Dhananjaybhai, 495
Shankar, S., 391
Sharma, Anil, 519
Sharma, Gaurav, 645
Sharma, Pallabi, 665
Sharma, Yogesh Kumar, 409
Shettar, Rajashekar B., 11
Shinde, Subhash, 213
Shirbahadurkar, Suresh D., 615
Shirol, Suhas B., 11
Shukla, Sneha, 749
Siddiqui, Nadia, 177
Singh, Latika, 399
Singh, Rishija, 763
Somaiya, Havan, 265
Sreenivasulu, G., 359
Sridevi, M., 709
Sujatha, K., 341, 351
Sunny, Elizabeth, 721
Surywanshi, Jaya R., 21
Syed Thouheed Ahmed, S., 391

T
Tarun Kumar, 99
Tiwari, Amit, 561
Trivedi, Harshal, 495
Tyagi, Dinesh Kumar, 737

U
Uma Maheswari, S., 165
Usmani, Atiya, 313

V
Viswanadha Raju, S., 359
Viswanath, H.L., 485
Vyas, Krutika, 429

W
Wahid, Abdul, 633
Wahlang, Imayanmosha, 665
Wasule, Archana, 625
Waykole, Saurabh, 605

Y
Yadav, Nidhika, 149

Z
Zaheer, Zainab, 235
Zaid, Mohammad, 727