

Suresh Chandra Satapathy
K. Srujan Raju
Jyotsna Kumar Mandal
Vikrant Bhateja *Editors*

Proceedings of the Second International Conference on Computer and Communication Technologies

IC3T 2015, Volume 1

Advances in Intelligent Systems and Computing

Volume 379

Series editor

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

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Rafael Bello, Universidad Central “Marta Abreu” de Las Villas, Santa Clara, Cuba
e-mail: rbellop@uclv.edu.cu

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

More information about this series at <http://www.springer.com/series/11156>

Suresh Chandra Satapathy · K. Srujan Raju
Jyotsna Kumar Mandal · Vikrant Bhateja
Editors

Proceedings of the Second International Conference on Computer and Communication Technologies

IC3T 2015, Volume 1

 Springer

Editors

Suresh Chandra Satapathy
Department of Computer Science
and Engineering
Anil Neerukonda Institute of Technology
and Sciences
Visakhapatnam
India

K. Srujan Raju
Department of Computer Science
and Engineering
CMR Technical Campus
Hyderabad
India

Jyotsna Kumar Mandal
Department of Computer Science
and Engineering
Kalyani University
Nadia, West Bengal
India

Vikrant Bhateja
Department of Electronics and
Communication Engineering
Shri Ramswaroop Memorial Group
of Professional Colleges
Lucknow, Uttar Pradesh
India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-81-322-2516-4 ISBN 978-81-322-2517-1 (eBook)
DOI 10.1007/978-81-322-2517-1

Library of Congress Control Number: 2015944452

Springer New Delhi Heidelberg New York Dordrecht London

© Springer India 2016

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

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

Preface

This volume contains 76 papers presented at the 2nd International Conference on Computer & Communication Technologies held during 24–26 July 2015 at Hyderabad, hosted by CMR Technical Campus in association with Division-V (Education and Research) CSI. It proved to be a great platform for researchers from across the world to report, deliberate, and review the latest progress in the cutting-edge research pertaining to intelligent computing and its applications to various engineering fields. The response to IC3T 2015 was overwhelming. It received a good number of submissions from different areas relating to intelligent computing and its applications in main tracks and three special sessions and after a rigorous peer-review process with the help of our program committee members and external reviewers, we finally accepted quality papers with an acceptance ratio of 0.35. We received submissions from seven overseas countries.

Dr. Vipin Tyagi, Jaypee University of Engineering & Technology, Guna, MP, conducted a Special Session on “Cyber Security and Digital Forensics,” Dr. K. Ashoka Reddy Principal, Kakatiya Institute of Technology & Science, Warangal, and Prof. Tara Sai Kumar, CMR Technical Campus, Hyderabad, conducted a Special Session on “Applications for Fuzzy Systems in Engineering” and Dr. Suma V. Dean, Research and Industry Incubation Centre (Recognized by Ministry of Science and Technology, Government of India), Dayananda Sagar Institutions, Bangalore, conducted a Special Session on “Software Engineering and Applications.”

We take this opportunity to thank all keynote speakers and special session chairs for their excellent support to make IC3T 2015 a grand success.

The quality of a referred volume depends mainly on the expertise and dedication of the reviewers. We are indebted to the program committee members and external reviewers who not only produced excellent reviews but also did them in short time frames. We would also like to thank CSI Hyderabad, CMR Group of Institutions, DRDO and JNTUH for coming forward to support us to organize this mega convention.

We express our heartfelt thanks to Mr. Ch. Gopal Reddy, Chairman of CMR Technical Campus, Smt. C. Vasanthalatha, Secretary of CMR Technical Campus,

and Dr. A. Raji Reddy Director of CMR Technical Campus, faculty and administrative staff for their continuous support during the course of the convention.

We would also like to thank the authors and participants of this convention, who have considered the convention above all hardships. Finally, we would like to thank all the volunteers who spent tireless efforts in meeting the deadlines and arranging every detail to make sure that the convention runs smoothly. All the efforts are worth and would please us all, if the readers of this proceedings and participants of this convention found the papers and event inspiring and enjoyable. We place our sincere thanks to the press, print, and electronic media for their excellent coverage of this convention.

July 2015

Suresh Chandra Satapathy
K. Srujan Raju
Jyotsna Kumar Mandal
Vikrant Bhateja

Team IC3T 2015

Chief Patrons

Sri. C. Gopal Reddy, Chairman
Smt. C. Vasanthalatha Reddy, Secretary
Dr. A. Raji Reddy, Director

Advisory Committee

Dr. A. Govardhan, SIT, JNTUH
Dr. V. Kamakshi Prasad, HOD-CSE, JNTUCEH
Prof. S.K. Udgata, HCU
Dr. Vasumathi, JNTUH
Dr. B. Padmaja Rani, JNTUH
Dr. O.B.V. Ramanaiah, JNTUH
Dr. B.N. Bhandari, JNTUH
Dr. Amit Acharya, JNTUH
Dr. D. Rajya Lakshmi, JNTUV
Dr. C. Srinivasa Kumar, VITSH(VU)
Dr. V. Kamaskshi Prasad, JNTUH
Dr. M.B.R. Murthy, CMRCET
Dr. M.V. Krishna Rao, CMRIT
Dr. M. Janga Reddy, CMRIT
Dr. L. Pratap Reddy, JNTUH
Dr. T. Anil Kumar, SRIITH
Dr. K. Srinivas Rao, CMRCET
Dr. Sahu Chatrapati, JNTUM
Dr. Vaka Murali Mohan, BRC

Program Chairs

Dr. A. Govardhan
Prof. G. Srikanth
Dr. K. Srujan Raju
Dr. Suresh Chandra Satapathy

Conveners

Prof. Dimlo. U. Fernandes
Mrs. K. Neeraja
Prof. T. Sai Kumar

Organizing Committee

Dr. K. Srujan Raju
Prof. G. Srikanth
Mr. N. Bhaskar
Mr. V. Naresh Kumar
Mr. A. Bharath Kumar, ECE
Mrs. B. Ragini Reddy, ECE
Mr. S. Venkatesh, ECE
Mrs. P. Satyavathi, CSE
Mr. N. Bhaskar, CSE
Prof. T. Sai Kumar
Prof. Dimlo. U. Fernandes
Mr. Md. Rafeeq
Mr. B. Ravinder
Mrs. J. Srividya, CSE
Mrs. Ch. Sudha Rani, ECE
Mrs. K. Mohana Lakshmi, ECE
Mr. J. Narasimha Rao, CSE
Mrs. B. Rama Devi, ECE, KITS

Program Committee

Ms. V. Swapna
Mr. Narasimha Rao
Mrs. Suvarna Gothane
Mrs. J. Srividya

Mr. K. Murali
Mr. Ch. Sudha Mani

Finance Committee

Prof. G. Srikanth
Mr. D. Venkateshwarlu

Publicity Committee

Dr. K. Srujan Raju
Mr. M. Ajay Kumar
Mr. P. Nagaraju
Mr. Anirban Paul

Exhibition Committee

Prof. T. Sai Kumar
Mrs. D. Anuradha
Mrs. A. Anusha
Mr. Ameen Uddin Md.

Transport Committee

Mr. R. Nagaraju
Mr. U. Yedukondalu
Mr. A. Bharath Kumar
Mr. V. Pradeep kumar

Hospitality Committee

Prof. K. Neeraja
Ms. Ch. Swapna
Mrs. K. Kiranmai
Mr. Md. Abdul Naqi

Sponsorship Committee

Mr. K. Bharath
Mr. P. Kranthi Rathan
Mr. E. Uma Shankar
Mr. Ch. Narendar

Marketing & PR Committee

Mr. Md. Shabeer
Mr. S. Satyanarayan Reddy
Mr. A. Vamshidhar Reddy
Mr. S. Madhu
Mr. S. Venkatesh
Mr. G. Kranthi Kiran

Registrations Committee

Mrs. P. Satyavathi
Mrs. K. Mohana Lakshmi
Mrs. K. Shrisha
Mrs. K. Jeji

Cultural Committee

Mrs. Shriya Kumari
Mrs. B. Ragini
Ms. B. Karuna Sree
Mr. M. Mahesh Babu

Web Portal Committee

Dr. K. Srujan Raju, CMRTC
Mr. T. Santhosh Kumar, CSI-Hyderabad Chapter
Mr. Chandra Mohan

International Advisory Committee/Technical Committee

Mr. Gautham Mahapatra, Sr. Scientist, DRDO, India.
Dr. A. Damodaram, Director, Academic & Planning, JNTUH
Dr. A. Govardhan, Director, SIT-JNTUH, India
Dr. Kun-Lin Hsieh, NTU, Taiwan
Dr. Ahamad J. Rusumdar, KIT, Germany
Dr. V.R. Chirumamilla, EUT, The Netherlands
Dr. Halis Altun, MU, Turkey
Dr. Vakka Murali Mohan, BRC, India
Dr. K. Ashoka Reddy, KITSW, India
Dr. Md. Zafar Ali Khan, IITH, India
Dr. S.K. Udagata, UOH, India
Mr. Anirban Pal, Tech Mahindra, India

External Reviewers Board

Ankur Singh Bist, KIET, Ghaziabad, India
Dac-Nhuong Le, VNU University, Hanoi, Vietnam
Sumit Ashok Khandelwal, MIT Academy of Engineering, Pune, India
Srinivas Sethi, Indira Gandhi Institute of Technology, India
Kavita Choudhary, ITM University, Gurgaon, India
Ashwini B. Abhale, D.Y. Patil College of Engineering, Akurdi, India
Sadhana J. Kamatkar, University of Mumbai, Mumbai, India
Musheer Ahmad, Jamia Millia Islamia, New Delhi, India
Mridu Sahu, NIT, Raipur, Chhattisgarh, India
Ranjan Tripathi, SRMGPC, Lucknow (Uttar Pradesh), India
Steven Lawrence Fernandes, Sahyadri College of Engineering & Management, Mangalore, India
G. Rosline Nesa Kumari, Saveetha School of Engineering, Saveetha University Chennai, India
Arshad Mohd. Khan, Innovative Technologies, Hyderabad, India
Nikhil Bhargava, CSI ADM, Ericsson, India
Chirag Arora, KIET, Ghaziabad, India
Jayashree H.V., PESIT, Bangalore, India
Ravi Tomar, University of Petroleum and Energy Studies, Dehradun, India
Sourav Samanta, University Institute of Technology, BU, India
Srinivas Aluvala, SR Engineering College, Warangal, India
Ritesh Maurya, ABVIITM, Gwalior, India
Abdul Wahid Ansari, University of Delhi, New Delhi, India
Gaikwad Bharatratna Pralhadrao, LPP Institute, Vivekanand College Campus, Aurangabad
A.N. Nagamani, PESIT, Bangalore, India

Balasaheb Deokate, Vidya Pratishthan's College of Engineering, Baramati, India
Satya Narayan Tazi, Government Engineering College, Ajmer, India
Sherin Zafar, Jamia Millia Islamia, New Delhi, India
Dileep Kumar Yadav, MRCE, Faridabad, India
Gustavo Fernandez, Austrian Institute of Technology, Vienna, Austria
Banani Saha, University of Calcutta, Kolkatta, India
Jagdish Chandra Patni, University of Petroleum and Energy Studies, Dehradun, India
Sayan Chakraborty, NIT, Durgapur, India
Kamble Vaibhav Venkatrao, Dr. Babasaheb Ambedkar Marathwada University Aurangabad, Maharashtra, India
Tushar V. Ratanpara, C.U. Shah University, Gujarat
Hem Kumar Gopal, Government College for Women, Mandya, India
Rupayan Das, University of Engineering & Management (UEM), Jaipur, Rajasthan
Maheswari Senthilkumar, Sambhram Institute of Technology, Bangalore, India
Hemprasad Y. Patil, LGNSCE, University of Pune, India
Angshuman Khan, University of Engineering & Management, Jaipur, India
Kamal Kant Sharma, Chandigarh University, Gharuan, Mohali, India
Sk. Md. Obaidullah, Aliah University, Kolkata, West Bengal, India
Nilanjan Dey, Bengal College of Engineering and Technology, Durgapur, India
Andhe Dharani, Mother Teresa Women's University, India
Sandip Das, University of Engineering and Management, Jaipur
Chayan Halder, West Bengal State University, Barasat, Kolkata, India
Vipin Khattri, SRMU, Lucknow-Deva Road, Uttar Pradesh
Alak Majumder, NIT, Arunachal Pradesh, India
Amartya Mukherjee, Bengal College of Engineering and Technology, Durgapur, India
Suvojit Acharjee, NIT, Agartala, India
Aarti Singh, MMITBM, M.M. University, Mullana, India
Ramesh Sunder Nayak, Canara Engineering College, Benjanapadavu, Mangalore, India
P.K. Gupta, Jaypee University of Engineering and Technology, Raghogarh, India
Shilpa Bahl, KIIT, Gurgaon, India
Sudhir Kumar Sharma, Ansal Institute of Technology, GGS Indraprastha University, Gurgaon, India
Bikesh Kumar Singh, NIT, Raipur, Chhattisgarh, India
Inderpreet Kaur, Chandigarh University, Gharuan, Mohali
Subuhi Khan, AMU, Aligarh, India
Shabana Urooj, GBU, Greater Noida, India
Mukul Misra, SRMU, Lucknow-Deva Road, Uttar Pradesh
Paras Jain, Jaypee University of Engineering and Technology, Raghogarh, India
Suresh Limkar, AISSMS IOIT, Pune, India
Pritee Parwekar, ANITS, Vishakhapatnam, India
Sri. N. Madhava Raja, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India

S. Ratan Kumar, ANITS, Visakhapatnam
S. Sridevi Sathya Priya, Karunya University, Coimbatore, Tamilnadu, India
Nishant Shrivastava, Jaypee University of Engineering and Technology, Raghogarh, India
Rajnikanth Venkatesan, St. Joseph's College of Engineering, Chennai, India
Sireesha Rodda, GITAM University, Visakhapatnam, Andhra Pradesh, India
Tanmoy Halder, University of Kalyani, West Bengal, India
Garima Singh, Jaypee University of Information Technology, Wagnaghat, Solan, Himachal Pradesh, India
A. Rajireddy, CMR Technical Campus, Hyderabad
Somnath Mukhopadhyay, University of Kalyani, West Bengal, India
Abhinav Krishna, SRMGPC, Lucknow (Uttar Pradesh), India
Himanshi Patel, SRMGPC, Lucknow (Uttar Pradesh), India
Arindam Sarkar, University of Kalyani, West Bengal, India
Srinivasa Murthy Y.V., NIT, Surathkal, India
Uttam Mondal, College of Engineering & Management, Kolaghat, India
Akanksha Sahu, SRMGPC, Lucknow (Uttar Pradesh), India
Tara Sai Kumar, CMR Technical Campus, Hyderabad
B.N. Biswal, BEC, Bhubaneswar
And many more.....

Contents

Medical Image Fusion in Curvelet Domain Employing PCA and Maximum Selection Rule.	1
Himanshi, Vikrant Bhateja, Abhinav Krishn and Akanksha Sahu	
Gigabit Network Intrusion Detection System Using Extended Bloom Filter in Reconfigurable Hardware	11
Akshay Eldho Jose and T. Gireeshkumar	
Hash-Based Rule Mining Algorithm in Data-Intensive Homogeneous Cloud Environment	21
Raghvendra Kumar, Prasant Kumar Pattnaik and Yogesh Sharma	
Privacy Preservation in Distributed Environment Using RSA-CRT	29
Raghvendra Kumar, Prasant Kumar Pattnaik and Yogesh Sharma	
Priority-Based Classification: An Automated Traffic Approach	35
Somasree Bhadra and Anirban Kundu	
A Taxonomy of Software’s Non-functional Requirements	47
Nida Afreen, Asma Khatoon and Mohd. Sadiq	
Segmentation of the Human Corpus Callosum Variability from T1 Weighted MRI of Brain	55
Shayak Sadhu, Sudipta Roy, Siddharth Sadhukhan and S.K. Bandyopadhyay	
Analysis of Organic Molecular Single-Electron Transistor Using C₄H₆B₂ with Different Metal Electrodes	67
E. Meher Abhinav, M. Chandra Mohan, A. Suresh Reddy, Vemana Chary and Maragani Thirupathi	

Analysis of Molecular Single-Electron Transistors Using Silicene, Graphene and Germanene	77
E. Meher Abhinav, Sai Naveen Kavuri, Thota Sandeep Kumar, Maragani Thirupathi, M. Chandra Mohan and A. Suresh Reddy	
Identification of the Plants Based on Leaf Shape Descriptors.	85
Pradip Salve, Milind Sardesai, Ramesh Manza and Pravin Yannawar	
Use of Global Positioning System to Track Movement of Mobile Node in Proxy Mobile Internet Protocol Version 6	103
Nitesh M. Tarbani and A.S. Alvi	
Is Mobile Cloud Computing Efficient for E-Learning?	109
Aftab Alam, Abdul Mateen Ansari, Mohammad Mujahid Barga and Ahmad Albhaisi	
A Comparative Analysis of Different Social Network Parameters Derived from Facebook Profiles	125
Paramita Dey and Sarbani Roy	
A Fast and Hardware-Efficient Visual Cryptography Scheme for Images.	133
Dipesh Vaya and Sarika Khandelwal	
Energy-Efficient Modified Bellman Ford Algorithm for Grid and Random Network Topologies.	143
Rama Devi Boddu, K. Kishan Rao and M. Asha Rani	
Text-Dependent Versus Text-Independent Speech Emotion Recognition	153
Biswajit Nayak and Manoj Kumar Pradhan	
A Novel Approach for Speaker Recognition by Using Wavelet Analysis and Support Vector Machines	163
Kanaka Durga Returi, Vaka Murali Mohan and Y. Radhika	
Real-Time Surveillance for Critical Activity Detection in ICUs	175
Dhakate Pankaj	
High Speed Network Intrusion Detection System Using FPGA.	187
S. Anuraj, P. Premalatha and T. Gireeshkumar	
Hawk Eye: A Plagiarism Detection System	195
Karuna Puri and Preeti Mulay	

A Novel Action Descriptor to Recognize Actions from Surveillance Videos 205
 T. Pradeepa, S. Abirami, M. Sivarathinabala and S. Murugappan

Protecting the Augmented Browser Extension from Mutation Cross-Site Scripting 215
 S. Remya and K. Praveen

Large-Scale Data Management System Using Data De-duplication System 225
 S. Abirami, Rashmi Vikraman and S. Murugappan

Movement Detection Using LabVIEW by Analysis of Real-Time Video 235
 D.V. Sumalatha and A.V. Ravi Kumar

Connectivity Model for Molecular Communication-Based Nanomachines Network in Normal and Sub-diffusive Regimes 245
 Prachi Raut and Nisha Sarwade

Application of Locally Weighted Regression for Predicting Faults Using Software Entropy Metrics 257
 Arvinder Kaur, Kamaldeep Kaur and Deepti Chopra

Effect of Defects on Current-Voltage Characteristics of a Silicene ZNR-Based Field Effect Transistor 267
 E. Meher Abhinav, M. Chandra Mohan, A. Suresh Reddy, Vemana Chary and Maragani Thirupathi

Performance Improvement of Read Operations in Distributed File System Through Anticipated Parallel Processing 275
 B. Rangaswamy, N. Geethanjali, T. Ragunathan and B. Sudheer Kumar

Steerable Texture Descriptor for an Effective Content-Based Medical Image Retrieval System Using PCA 289
 B. Jyothi, Y. MadhaveeLatha, P.G. Krishna Mohan and V.S.K. Reddy

A Novel Multi-view Similarity for Clustering Spatio-Temporal Data 299
 Vijaya Bhaskar Velpula and M.H.M. Krishna Prasad

WRSP-Miner Algorithm for Mining Weighted Sequential Patterns from Spatio-temporal Databases 309
 Gurram Sunitha and A. Rama Mohan Reddy

Performance Analysis in Free Space Optical Communication System Using Aperture Averaging 319
Saloni Bhatia, Monika Singh and Hemani Kaushal

Performance Evaluation of Free Space Optical Link Under Various Weather Conditions. 329
Monika Singh, Saloni Bhatia and Hemani Kaushal

Evolutionary Improved Swarm-Based Hybrid K-Means Algorithm for Cluster Analysis. 343
Janmenjoy Nayak, D.P. Kanungo, Bighnaraj Naik and H.S. Behera

Analysis of Energy Efficient, LEACH-Based Cooperative Wireless Sensor Network 353
V. Murali and Sindhu Hak Gupta

A Monotonic Digitally Controlled Delay Element-Based Programmable Trigger Pulse Generator 365
Amit Krishna Dwivedi, Manisha Guduri, Rishab Mehra and Aminul Islam

Layered-Encoding Video Multicasting in Mobile WiMAX Networks. 375
Jamil M. Hamodi and Ravindra C. Thool

Intuitionistic Fuzzy Similarity and Information Measures with Physical Education Teaching Quality Assessment 387
Arunodaya Raj Mishra, Divya Jain and D.S. Hooda

Molecular Level Insight into the Interactions of SoxC and SoxD from Epsilonproteobacteria *Sulfurimonas denitrificans*: A Biomolecular Computational Approach. 401
Sujay Ray, Arundhati Banerjee and Angshuman Bagchi

“Jugaad”—The Creativeness for Selection of Software Development Methodology Advisory System—Fuzzy Expert System. 411
Kardile Vilas Vasantrya and Chitra G. Desai

A Framework for Data Clustering of Large Datasets in a Distributed Environment. 425
Ch. Swetha Swapna, V. Vijaya Kumar and J.V.R. Murthy

Segmentation of Ancient and Historical Gilgit Manuscripts 443
 Pinjari Hameed, Rosemary Koikara and Chethan Sharma

Semantic-Based Approach for Automatic Annotation and Classification of Medical Services in Healthcare Ecosystem 449
 Vijayalakshmi Kakulapati, Rishi Sayal, Ravi Aavula and Sunitha Devi Bigul

Real-Time Graphs for Communication Networks: A Fuzzy Mathematical Model 463
 Siddhartha Sankar Biswas, Bashir Alam and M.N. Doja

PWLCM-Based Random Search for Strong Substitution-Box Design. 471
 Musheer Ahmad, Danish Raza Rizvi and Zishan Ahmad

Fuzzy C-Means and Fuzzy TLBO for Fuzzy Clustering. 479
 P. Gopala Krishna and D. Lalitha Bhaskari

A Mean-Based Thresholding Approach for Broken Character Segmentation from Printed Gujarati Documents 487
 Riddhi J. Shah and Tushar V. Ratanpara

Entity Resolution-Based Jaccard Similarity Coefficient for Heterogeneous Distributed Databases 497
 Ramesh Dharavath and Abhishek Kumar Singh

Energy Efficient Algorithms for Hot Spot Problem in Wireless Sensor Networks 509
 Srikanth Jannu and Prasanta K. Jana

Mining Community-Based Top-*k* Experts and Learners in Online Question Answering Systems. 519
 S. Rao Chintalapudi and M.H.M. Krishna Prasad

An Approach for Creating Framework for Automated Question Generation from Instructional Objective. 527
 Syaamantak Das, Rajeev Chatterjee and Jyotsna Kumar Mandal

Adaptive Output Feedback Control System Design for Low-Cost Electronic Knee Prosthesis 537
 Salman Shaikh and Akshay Malhotra

Low-Leakage, Low-Power, High-Stable SRAM Cell Design	549
Soumitra Pal, Y. Krishna Madan and Aminul Islam	
Optimizing Routes in Mobile Ad Hoc Networks Using Genetic Algorithm and Ant Colony Optimization	557
Pankaj Uttam Vidhate, R.S. Bichkar and Yogita Wankhade	
Concatenation Technique for Extracted Arabic Characters for Efficient Content-based Indexing and Searching	567
Abdul Khader Jilani Saudagar and Habeeb Vulla Mohammed	
Harmonic Elimination Using DSTATCOM	577
Shabana Urooj and Pikasha Sharma	
Fault Tolerance in Wireless Sensor Networks: Finding Primary Path	593
Pritee Parwekar and Sireesha Rodda	
PSO-Based Multiple-sink Placement Algorithm for Protracting the Lifetime of Wireless Sensor Networks	605
C. Srinivasa Rao P., Haider Banka and Prasanta K. Jana	
Denoising of GPS Positioning Data Using Wavelet-Based Hidden Markov Tree	617
Ch. Mahesh, K. Ravindra and V. Kamakshi Prasad	
Parent Selection Based on Link Quality Estimation in WSN	629
Siba Mitra, Sarbani Roy and Ajanta Das	
Enhanced Higher Order Orthogonal Iteration Algorithm for Student Performance Prediction	639
Prema Nedungadi and T.K. Smruthy	
A One-to-One Communication Model to Facilitate Conversation Between Differently-Abled People by a Portable and Handy Machine	651
Rajat Sharma and Swarnima Gupta	
Malicious File Hash Detection and Drive-by Download Attacks	661
Ibrahim Ghafir and Vaclav Prenosil	
Parameter Optimization of Support Vector Machine by Improved Ant Colony Optimization	671
Srujana Rongali and Radhika Yalavarthi	

KT3F: A Key-Based Two-Tier Trust Management Filtering Scheme for Intrusion Detection in Wireless Sensor Network 679
 Moutushi Singh, Rupayan Das, Mrinal Kanti Sarkar,
 Koushik Majumder and Subir Kumar Sarkar

Cascading De-noising Algorithm for Improving GPS Positional Accuracy. 691
 Ch. Mahesh, R. Pavan Kumar Reddy, K. Ravindra
 and V. Kamakshi Prasad

A Corpus of Word-Level Offline Handwritten Numeral Images from Official Indic Scripts 703
 Sk Md Obaidullah, Chayan Halder, Nibaran Das and Kaushik Roy

A Planar Ultra-Wide Band Antenna Design Using Circularly Truncated Corners and Notches. 713
 H.S. Mewara, M.M. Sharma, Mayank Sharma, Mukesh Gupta
 and Ajay Dadhich

Genetic Algorithm for k -Connected Relay Node Placement in Wireless Sensor Networks 721
 Suneet K. Gupta, Pratyay Kuila and Prasanta K. Jana

Segmentation of Cotton Bolls by Efficient Feature Selection Using Conventional Fuzzy C-Means Algorithm with Perception of Color. 731
 Sandeep Kumar, Manish Kashyap, Akashdeep Saluja
 and Mahua Bhattacharya

System Dynamics Modeling for Analyzing Recovery Rate of Diabetic Patients by Mapping Sugar Content in Ice Cream and Sugar Intake for the Day 743
 Suhas Machhindra Gaikwad, Rahul Raghvendra Joshi and Preeti Mulay

Enhanced Stable Period for Two Level and Multilevel Heterogeneous Model for Distant Base Station in Wireless Sensor Network 751
 Pawan Singh Mehra, M.N. Doja and Bashir Alam

Dynamic Timetable Generation Using Constraint Satisfaction Algorithm 761
 Urmila Kalshetti, Deepika Nahar, Ketan Deshpande, Sanket Gawas
 and Sujay Sudeep

Application of Genetic Algorithm for Evolution of Quantum Fourier Transform Circuits 773
Swanti Satsangi and C. Patvardhan

An Improved Polygon Clipping Algorithm Based on Affine Transformation 783
Mugdha Sharma and Jasmeen Kaur

Alternative Design Space Analysis for Electronic Commerce System 793
P. Rajarajeswari, D. Vasumathi and A. Ramamohanreddy

Author Index 805

About the Editors

Dr. Suresh Chandra Satapathy is currently working as Professor and Head, Department of Computer Science and Engineering, Anil Neerukonda Institute of Technology and Sciences (ANITS), Vishakhapatnam, Andhra Pradesh, India. He obtained his Ph.D. in Computer Science Engineering from JNTUH, Hyderabad and Master degree in Computer Science and Engineering from National Institute of Technology (NIT), Rourkela, Odisha. He has more than 27 years of teaching and research experience. His research interests include machine learning, data mining, swarm intelligence studies, and their applications to engineering. He has more than 98 publications to his credit in various reputed international journals and conference proceedings. He has edited many volumes from Springer AISC and LNCS in past and he is also the editorial board member in few international journals. He is a senior member of IEEE and Life Member of Computer society of India. Currently, he is the National Chairman of Division-V (Education and Research) of Computer Society of India.

Dr. K. Srujan Raju is the Professor and Head, Department of Computer Science and Engineering, CMR Technical Campus. Professor Srujan earned his Ph.D. in the field of Network Security and his current research include computer networks, information security, data mining, image processing, intrusion detection, and cognitive radio networks. He has published several papers in referred international conferences and peer reviewed journals and also he was on the editorial board of CSI 2014 Springer AISC series 337 and 338 volumes. In addition to this, he has served as reviewer for many indexed journals. Professor Raju is also awarded with Significant Contributor, Active Member Awards by Computer Society of India (CSI) and currently he is the Hon. Secretary of CSI Hyderabad Chapter.

Dr. Jyotsna Kumar Mandal has M.Sc. in Physics from Jadavpur University in 1986, M.Tech. in Computer Science from University of Calcutta. He was awarded Ph.D. in Computer Science and Engineering by Jadavpur University in 2000. Presently, he is working as Professor of Computer Science and Engineering and former Dean, Faculty of Engineering, Technology and Management, Kalyani University, Kalyani, Nadia, West Bengal for two consecutive terms. He started his

career as lecturer at NERIST, Arunachal Pradesh in September, 1988. He has teaching and research experience of 28 years. His areas of research include coding theory, data and network security, remote sensing and GIS-based applications, data compression, error correction, visual cryptography, steganography, security in MANET, wireless networks, and unify computing. He has produced 11 Ph.D. degrees, three submitted (2015) and 8 are ongoing. He has supervised 3 M.Phil. and 30 M.Tech. theses. He is life member of Computer Society of India since 1992, CRSI since 2009, ACM since 2012, IEEE since 2013, and Fellow member of IETE since 2012, Executive member of CSI Kolkata Chapter. He has delivered invited lectures and acted as program chair of many international conferences and also edited 9 volumes of proceedings from Springer AISC series, CSI 2012 from McGraw-Hill, CIMTA 2013 from Procedia Technology, Elsevier. He is reviewer of various international journals and conferences. He has over 355 articles and 5 books published to his credit.

Prof. Vikrant Bhateja is Associate Professor, Department of Electronics and Communication Engineering, Shri. Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow and also the Head (Academics & Quality Control) in the same college. His areas of research include digital image and video processing, computer vision, medical imaging, machine learning, pattern analysis and recognition, neural networks, soft computing and bio-inspired computing techniques. He has more than 90 quality publications in various international journals and conference proceedings. Professor Vikrant has been on TPC and chaired various sessions from the above domain in international conferences of IEEE and Springer. He has been the track chair and served in the core-technical/editorial teams for international conferences: FICTA 2014, CSI 2014 and INDIA 2015 under Springer-ASIC Series and INDIACom-2015, ICACCI-2015 under IEEE. He is associate editor in International Journal of Convergence Computing (IJConvC) and also serving in the editorial board of International Journal of Image Mining (IJIM) under Inderscience Publishers. At present he is guest editor for two special issues floated in International Journal of Rough Sets and Data Analysis (IJRSDA) and International Journal of System Dynamics Applications (IJSDA) under IGI Global publication.

Medical Image Fusion in Curvelet Domain Employing PCA and Maximum Selection Rule

Himanshi, Vikrant Bhateja, Abhinav Krishn and Akanksha Sahu

Abstract Curvelet transform achieves a compact representation of edges and curved shapes in the image, which other techniques like wavelets and ridgelets are not able to represent. This property of curvelet transform facilitates the retrieval of complementary information from medical images for precise and efficient clinical diagnosis. This paper presents a combination of curvelet transform along with principal component analysis (PCA) and maximum selection rule as an improved fusion approach for MRI and CT-scan. The proposed fusion approach involves image decomposition using curvelet transform followed by application of PCA for dimensionality reduction and the selection of maximum matrix to select only the relevant information in the images. Fusion factor (FF) and structural similarity index (SSIM) are used for performance evaluation of the proposed approach. Simulation results demonstrate an improvement in visual quality of the fused image supported by higher values of fusion metrics.

Keywords Curvelet transform · Fusion factor · Max fusion rule · MRI · PCA

Himanshi (✉) · V. Bhateja · A. Krishn · A. Sahu
Department of Electronics and Communication Engineering, Shri Ramswaroop Memorial
Group of Professional Colleges, Lucknow 227105, U.P, India
e-mail: himanshi2593@gmail.com

V. Bhateja
e-mail: bhateja.vikrant@gmail.com

A. Krishn
e-mail: abhinavkrishn01@gmail.com

A. Sahu
e-mail: akankshasahu1708@gmail.com

1 Introduction

‘Medical image fusion’ is the process of combining and merging complementary information into a single image from two or more source images to maximize the information content of images and minimize the distortion and artifacts in the resulting image [1]. The complementary nature of medical imaging sensors of different modalities, (X-ray, magnetic resonance imaging (MRI), and computed tomography (CT)) has brought a great need of image fusion for the retrieval of relevant information from medical images. The significance of fusion process is important for multimodal images as images obtained from single modalities provide only specific information; thus it is not feasible to get all the requisite information from image generated by single modality [2, 3]. This highlights the need toward the development of multimodality medical imaging sensors for extracting clinical information to explore the possibility of data reduction along with better visual representation. In the past decades, several fusion algorithms varying from the traditional fusion algorithms like simple averaging and weighted averaging, maximum and minimum selection rule [4] have been proposed. With the advancement of research in this field, algorithms such as intensity–hue–saturation (IHS) [5] and Brovey transform (BT) [6] have been used to fuse medical images. In the recent years multiresolution approaches using Mallat [7], the à trous [8] transforms, laplacian pyramids [9], wavelet [10], complex wavelet [11], contourlet [12], and non-subsampled contourlet [13] have been proposed for image fusion. Luo et al. [14] used a combination of PCA with à trous wavelet transforms which focused on the spatial and spectral resolutions. But, the technique did not lay emphasis on edge or shape detection, which are fundamental structures in natural images and particularly relevant from a visual point of view. Petrovic et al. [15] proposed a ‘fuse then decompose’ technique which represented input image in the form of gradient maps at each resolution level. It has been observed that the said approach by authors did not yield satisfactory performance but in turn increased the computational complexity due to the involvement of gradient maps. Sadhasivam in their work [16] applied PCA along with the selection of maximum pixel intensity to perform fusion. The said combination produced an image which had less structural similarity with the source images along with low contrast and luminescence. Singh and Khare [4] proposed Daubechies complex wavelet transform for multimodal image fusion using maximum selection rule. DCxWT conquered the drawback of complex wavelet by eliminating shift sensitive property and also provided the phase information of the source images. But the resultant image lacked contrast and failed to detect the shapes and contours from the source images. Further, these fused images are also processed with denoising [17, 18], contrast, and edge enhancement [19–24] techniques to improve upon the visualization of diagnostic information. The proposed work in this paper presents a combination of curvelet transform and PCA as an improvement to the aforementioned limitations. The obtained results have been evaluated using FF and SSIM as fusion metrics; yielding satisfactory performance. The rest of the paper is organized as follows: The proposed fusion approach is

discussed in Sect. 2. Section 3 presents experimental results and finally the conclusions are drawn in Sect. 4.

2 Proposed Fusion Approach

2.1 Motivation

Digital images are generally described via orthogonal, non-redundant transforms like wavelet [10] or discrete cosine transform (DCT) for the purpose of multi-scale analysis. The wavelet transform portrays more emphasis on catching zero-dimensional singularities; however, the performance of wavelets in mono-dimensional domain is lost when they are applied to images using 2D separable basis. While ridgelet transform [25] poses to be a powerful instrument in catching and representing mono-dimensional singularities (in bidimensional space), it is unable to detect curved shapes in an image [26]. Therefore, an advanced version of ridgelet transform, i.e., curvelet transform has been used. Moreover, curvelet transform achieves very compact representation of linear singularities in images. Instrumental in the implementation of the curvelets is the Radon transform, which is preferred as a tool to extract lines in edge dominated images [27]. Thus, reliable detection and representation of edges as well as curved shapes in an image using curvelets forms the motivation to apply fusion process in this domain.

2.2 Proposed Fusion Algorithm

The first step in the proposed fusion approach involves the preprocessing of the MRI and CT-scan images, i.e., the conversion of image from RGB scale to gray-scale (RGB components of the image are converted into grayscale components). The next step is to decompose the source images using curvelet transform [32]. Two-dimensional curvelet transform can be viewed as a wavelet analysis in Radon domain; which itself is a tool of shape detection [28]. The curvelet coefficients of an image $f(x_1, x_2)$ are obtained by following steps [29, 30]:

- Sub-band Decomposition: The image $f(x_1, x_2)$ is filtered into sub-bands.

$$f \rightarrow (P_\alpha f, \Delta_1 f, \Delta_2 f, \dots) \quad (1)$$

- Smooth Partitioning: Sub-bands are smoothly windowed into squares.

$$\Delta_s f \rightarrow (W_Q \Delta_s f) Q \in Q_s \quad (2)$$

- Renormalization: Normalization of each resulting square to unit scale.

$$g_Q = 2^{-s}(T_Q)^{-1}(W_s \Delta_s f) \quad Q \in Q_s \quad (3)$$

- Ridgelet Analysis: Analysis of each square in Ridgelet domain [34, 35].

$$\alpha_\mu = \langle g_Q, P_\lambda \rangle \quad \mu = (Q, \lambda) \quad (4)$$

where α_μ represents the curvelet coefficients. The decomposition process stated above can be understood more clearly from Fig. 1.

Once the source images are decomposed using curvelet transform and curvelet coefficients are obtained; PCA is applied as a fusion rule to selectively combine the appropriate curvelet coefficients of input images [31]. PCA serves to transform/project the features from the original domain to a new domain (PCA domain) where they are arranged in their order of variance. Fusion process is achieved in PCA domain by retaining only those features that contain a significant amount of information. This can be achieved by retaining only those components that have a larger variance. The steps involved in the proposed PCA algorithm are outlined in Fig. 2. The next step involves the reconstruction of the processed coefficients (after PCA fusion rule) using inverse curvelet transform. This step is followed by the selection of maximum matrix from the reconstructed images using maximum selection rule to yield the fused image.

The final step in the proposed approach is its objective evaluation, which requires usage of HVS-based image quality assessment (IQA) approaches. In the present work, performance evaluation is carried out with metrics such as fusion

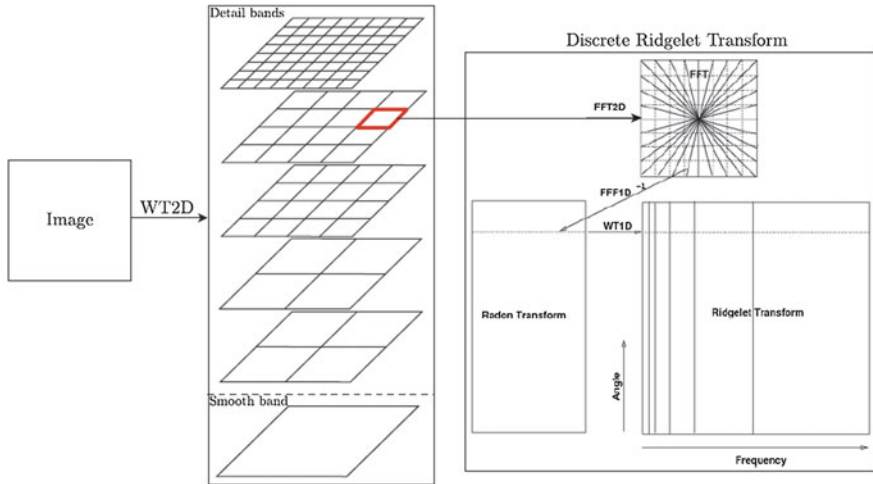


Fig. 1 Curvelet transform flowgraph showing all the steps required to obtain the curvelet coefficients [29, 30]

BEGIN

Step 1 : *Input*: Curvelet coefficients (of both MRI & CT).
 Step 2 : *Compute*: Column vectors from curvelet coefficients.
 Step 3 : *Compute*: Covariance matrix using these vectors.
 Step 4 : *Process*: Diagonal elements of the covariance vector.
 Step 5 : *Compute*: Eigen vectors and Eigen values of covariance matrix.
 Step 6 : *Process*: Column vector corresponding to large Eigen value (by dividing each element with the mean of Eigen vector).
 Step 7 : *Compute*: Multiplication of normalized Eigen vector values by each term of curvelet coefficient matrix.
 Step 8 : *Process*: Inverse curvelet transform of two scaled matrices calculated in Step 7.
 Step 9 : *Compute*: Maximum matrix from the images calculated in Step 8.
 Step10 : *Output*: Fused image.

END

Fig. 2 PCA algorithm in curvelet domain

factor (FF) [4] and structural similarity index (SSIM) [32, 33]. Higher values of fusion metrics indicate better fusion performance.

3 Experimental Results and Discussions

Simulations in the present work have been performed on images of two different modalities (CT and MRI). This section deals with the qualitative and quantitative analysis of the fused image obtained from the proposed approach. The results of fused images obtained using the proposed approach are given in Fig. 3; while the quantitative analysis of the same is shown in Table 1.

Fusion results in Fig. 3 for Sets 1 and 2 show that the fused image has a better visual characteristic from the diagnostic point of view. CT-scan images give information about the shape of the tumor which is helpful in determining the extent of disease, whereas MRI image gives soft tissue details. It can be clearly seen that fused image contains complementary information about both the images; i.e., soft tissue details as well as the shape of the tumor. This is further supported by high values of fusion metrics (FF and SSIM). Values of SSIM above 0.8 demonstrate a good structural, contrast, as well as luminance restoration in the fused image. In addition, high values of fusion factor show the effectiveness of the proposed fusion approach.

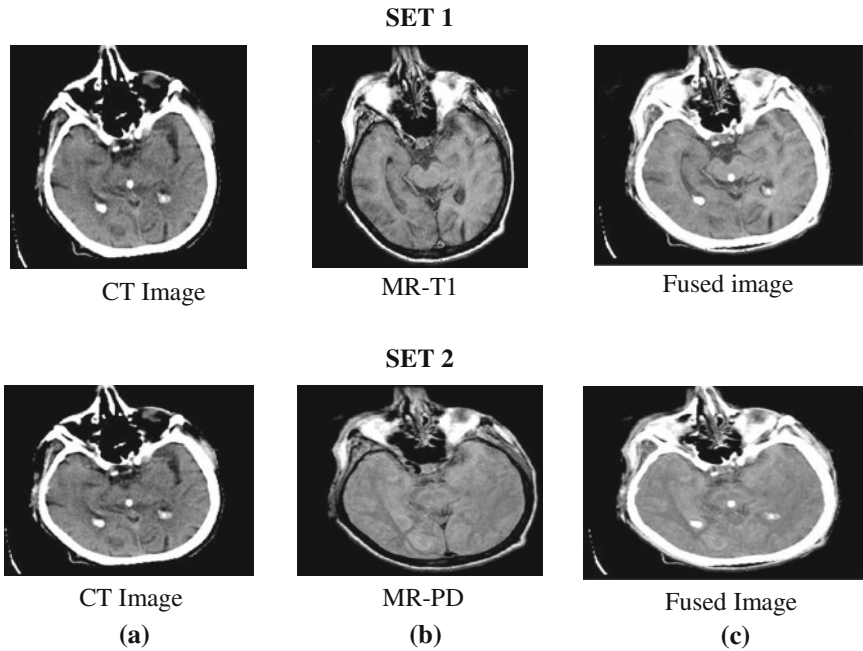


Fig. 3 Image fusion results for different sets of MRI and CT-scan images with the proposed approach. **a** Input CT image. **b** Input MRI image. **c** Fused image

Table 1 Quantitative analysis of proposed fusion approach

Set no.	FF	SSIM
1	5.7501	0.8188
2	5.6738	0.8390

3.1 Comparison with Other Fusion Approaches

The present approach has been compared with the DWT-based fusion approach proposed by Zheng et al. [34] and sparse representation approaches like simultaneous orthogonal matching pursuit (SOMP) and orthogonal matching pursuit (OMP) [35] for medical image fusion. The obtained result shows the effectiveness of the proposed approach in visual representation compared to DWT, SOMP, and OMP approaches. The fused image obtained from the proposed approach represents edges and curved shapes more effectively, as the image has better sharpness compared to images obtained from other approaches. Moreover, the higher values of the fusion metric shown in Table 2 validate that the proposed fusion approach has better diagnostic utility than the other approaches.

Table 2 Performance comparison of fused images

Fusion approach	SSIM
DWT [34]	0.58
OMP [35]	0.67
SOMP [35]	0.68
Proposed approach	0.85

4 Conclusion

An image fusion approach employing PCA and maximum selection rule in curvelet domain is presented in the paper. This approach is more appropriate for image fusion as it combines the feature enhancement property of PCA and edge detection property of curvelet. Moreover, the involvement of maximum selection rule facilitates the representation of only relevant information from the input images. The proposed fusion approach yielded a fused image that represented shapes and soft tissue details of tumor. Thus, it provides the details of two different modalities in one single image, justifying the purpose of medical image fusion. Significant results relevant from a visual point of view, as well as high values of the fusion metrics (FF/SSIM), have been obtained from the proposed fusion approach. Comparison results indicate improvement in restoration of structural, contrast, and luminance features in the obtained fused image, as depicted by high value of SSIM, in comparison to other fused images. Hence, the proposed fusion approach is more precise and can be used more effectively for medical diagnosis than other methods of fusion.

References

1. Dasarathy, B.V.: information fusion in the realm of medical applications—a bibliographic glimpse at its growing appeal. *Inf. Fus.* **13**(1), 1–9 (2012)
2. Schoder H., Yeung H.W., Gonen M., Kraus D., Larson S.M.: Head and neck cancer: clinical usefulness and accuracy of PET/CT image fusion. *Radiology* 65–72 (2004)
3. Nakamoto Y., Tamai K., Saga T., Higashi T., Hara T., Suga T., Koyama T., Togashi K.: Clinical value of image fusion from MR and PET in patients with head and neck cancer. *Mol. Imaging Biol.* 46–53 (2009)
4. Singh, R., Khare, A.: Fusion of multimodal images using daubechies complex wavelet transform—a multiresolution approach. *Info. Fusion.* **19**, 49–60 (2014)
5. Tu, T.M., Su, S.C., Shyu, H.C., Huang, P.S.: A new look at IHS-like image fusion methods. *Inf. Fusion* **2**(3), 177–186 (2001)
6. Gillespie, A.R., Kahle, A.B., Walker, R.E.: Color enhancement of highly correlated images—II. Channel ratio and ‘chromaticity’ transformation techniques. *Remote Sens. Environ.* **22**, 343–365 (1987)
7. Mallat S.: A theory for multi-resolution signal: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693, July (1989)
8. Shensa M.J.: The discrete wavelet transform: wedding the *à Trous* and Mallat algorithms. *IEEE Trans. Signal Process.* **40**(10), 2464–2482, Oct (1992)

9. Sahu A., Bhateja V., Krishn A., Himanshi: Medical image fusion with Laplacian pyramids. In: Proceedings of (IEEE) International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MEDCOM-2014), Gr. Noida, 448–453, Nov (2014)
10. Krishn A., Bhateja V., Himanshi, Sahu A.: medical image fusion using combination of PCA and wavelet analysis. In: Proceedings of 3rd (IEEE) International Conference on Advances in Computing, Communications and Informatics (ICACCI-2014), pp. 986–991. Gr. Noida, India, Sept (2014)
11. Himanshi, Bhateja V., Krishn A. and Sahu A.: An improved medical image fusion approach using PCA and complex wavelets. In: Proceedings of (IEEE) International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MEDCOM-2014), pp. 442–447. Gr. Noida, India, Nov (2014)
12. Aboubaker M. ALEjaily: Fusion of Remote Sensing Images Using Contourlet Transform. Innovations and Advanced Techniques in Systems, Computing Sciences and Software Engineering, Springer 213–218 (2008)
13. Cunha, L.D., Zhou, J.P.: The non-subsampled contourlet transform: theory, design, and applications. *IEEE Tran. Image Proc.* **15**(10), 3089–3101 (2006)
14. Luo, Y., Liu, R., Zhu, Y.Z.: Fusion of remote sensing image based on the PCA and atrous wavelet transform. *Int. Archives Photogram. Remote Sens. Spat. Info. Sci.* **XXXVII**(Part B7), 1155–1158 (2008)
15. Petrovic, V.S., Costas, S.X.: Gradient-based multiresolution image fusion. *IEEE Trans. Image Process.* **13**(2), 228–237 (2004)
16. Sadhasivam, S.K., Keerthivasan, M.K., Muttan, S.: Implementation of max principle with PCA in image fusion for surveillance and navigation application. *Electron. Lett. Comput. Vision Image Anal.* **10**(1), 1–10 (2011)
17. Jain, A., Singh, S., Bhateja, V.: A robust approach for denoising and enhancement of mammographic breast masses. *Int. J. Converg. Comput.* **1**(1), 38–49 (2013). Inderscience Publishers
18. Shrivastava, A., Alankrita, Raj, A., Bhateja, V.” Combination of wavelet transform and morphological filtering for enhancement of magnetic resonance images. In: Proceedings of International Conference on Digital Information Processing and Communications (ICDIPC 2011), CCIS-188, pp. 460–474, July 2011
19. Bhateja, V., Urooj, S., Pandey, A., Misra, M., Lay-Ekuakille, A.: A polynomial filtering model for enhancement of mammogram lesions. In: Proceedings of IEEE International Symposium on Medical Measurements and Applications, pp. 97–100 (2013)
20. Siddhartha, Gupta, R., Bhateja, V.: A log-ratio based unsharp masking (UM) approach for enhancement of digital mammograms. In: Proceedings of CUBE International Information Technical Conference and Exhibition, pp. 26–31 (2012)
21. Bhateja, V., Devi, S.: An improved non-linear transformation function for enhancement of mammographic Breast masses. In: Proceedings of 3rd International Conference on Electronics and Computer Technology, vol. 5, pp. 341–346 (2011)
22. Bhateja, V., Misra, M., Urooj, S., Lay-Ekuakille, A.: A robust polynomial filtering framework for mammographic image enhancement from biomedical sensors. *IEEE Sens. J.* **13**(11), 4147–4156 (2013)
23. Bhateja, V., Devi, S.: A novel framework for edge detection of microcalcifications using a non-linear enhancement operator and morphological filter. In: Proceedings of (IEEE) 3rd International Conference on Electronics and Computer Technology (ICECT-2011), Kanyakumari, vol. 5, pp. 419–424, April 2011
24. Raj, A., Alankrita, Shrivastava, A., Bhateja, V.: Computer aided detection of brain tumor in MR images. *Int. J. Eng. Technol. (IACSIT-IJET)* **3**, 523–532 (Oct 2011)
25. Granai L., Moschetti F., Vanderghenst, P.: Ridgelet transform applied to motion compensated images. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 561–564, 6–10 April (2003)
26. Krishn A., Bhateja V., Himanshi, Sahu, A.: PCA based medical image fusion in ridgelet domain. In: Proceedings of 3rd International Conference on Frontiers of Intelligent

- Computing: Theory and Applications (FICTA-2014), Advances in Intelligent Systems and Computing, vol. 328, pp. 475–482. Springer, Nov (2014)
27. Ali, F.E., El-Dokany, I.M., Saad, A.A., Abd El-Samie, F. E.: Curvelet fusion of MR and CT images. *Prog. Electromagn. Res.* **3**, 215–224 (2008)
 28. Stark J.L., Candes E.J. Donoho D.L.: The curvelet transform for denoising. *IEEE Trans. Image Proc* **11**(6), 670–684, June (2002)
 29. Candes E.J., Donoho D.L.: Curvelets—a surprisingly effective nonadaptive representation for objects with edges. In: Cohen, A., Rabut, C., Schumaker, L.L. (eds.) *Saint-Malo Proceedings*. Vanderbilt University Press, Nashville, pp. 1–10 (1999)
 30. Candes, E.J., Donoho, D.L.: Ridgelets—a key to higher dimensional intermittency. *Phil. Trans. R. Soc. Lond. A* **357**, 2495–2509 (1999)
 31. Naidu, V.P.S., Rao, J.R.: Pixel-level image fusion using wavelets and principal component analysis. *Def. Sci. J.* **58**(3), 338–352 (2008)
 32. Piella G., Heijmans, H.: A new quality metric for image fusion. In: *International Conference on Image Processing*, Barcelona, 14 Sept (2003)
 33. Bhateja, V., Srivastava, A., Kalsi, A.: Fast SSIM index for color images employing reduced-reference evaluation. In: *Proceedings of International Conference on Frontiers in Intelligent Computing Theory and Applications*, vol. 247, pp. 451–458, (2013)
 34. Zheng, Y., Essock, E.A., Hansen, B.C.: An Advanced image fusion algorithm based on wavelet transform –incorporation with PCA and morphological processing. *Proc. SPIE* **5298**, 177–187 (2004)
 35. Yang, B., Li, S.: Pixel-level image fusion with simultaneous orthogonal matching pursuit. *Inf. Fusion.* **13**, 10–19 (2012)

Gigabit Network Intrusion Detection System Using Extended Bloom Filter in Reconfigurable Hardware

Akshay Eldho Jose and T. Gireeshkumar

Abstract Network intrusion detection system collects information from network and identifies all the possible existing network security threats. Software based detection systems are common but are not good enough for the current network security requirements. Present day network intrusion detection needs wire-level data transfer to avoid the inefficiency in pattern matching process. Hardware based solutions like field programmable gate array which is known for its high processing capability can easily solve these issues. This paper implements a hardware based gigabit intrusion detection system using extended Bloom filter concepts. The paper presents a solution to reduce the high error rate of Bloom Filter by introducing a Reference Vector to the work and evaluates its performance. The reference vector verifies the Bloom filter output for any possible false positive results and reduces the error rate in the system.

Keywords Network intrusion detection · Field programmable gate array · Extended bloom filter · Reference vector

1 Introduction

Network security refers to the protection of various resources from all kind of malicious activities and ensures safety to network and data. It implements the security policies and analyzes various threats and stops it from entering the network. Network security consist of many layers such as firewall and network

A.E. Jose (✉) · T. Gireeshkumar
TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham,
Coimbatore, Tamil Nadu, India
e-mail: akshayeldhojose@gmail.com

T. Gireeshkumar
e-mail: gireeshkumart@gmail.com

intrusion detection systems. Firewalls are used to block the access between the networks but it does not study the traffic nor alert the administrator. An intrusion detection system is capable of studying the traffic patterns and compares it against the known attack patterns. It has the capability to inspect the packet contents deeply and protects against network threats.

Software based network intrusion detections are common which can be implemented easily. Snort is an example of open source light weight intrusion detection system [1] which uses signatures to compare thousands of attack patterns. But if the network falls in gigabit speed, it will be difficult for the software to support the system. Hardware solutions solve these issues by converting the rules into Hardware description language. Field programmable gate array is one such hardware which is designed to be configured by a Hardware description language. It consists of many high speed logic blocks capable of parallel processing to produce high performance gain. The general functional model consists of three sections: buffering, packet analyzer and rule matching section. The gigabit network is connected to hardware through Ethernet interfaces. The packets are queued in buffer section to balance between hardware and network. Packets are forwarded to the packet analyzer where it extracts the information from packet. These information are then compared against a set of rules in rule matcher. The alerting and logging mechanism works according to the output of rule matcher.

In this paper, a gigabit network intrusion detection system is designed based on Bloom Filter, to identify the attack patterns in the network. The paper efficiently designs Bloom filter algorithm for string matching engine (SME). Bloom filter test the participation of an element from a group of elements. In this approach, the elements in a group are hashed with multiple hash functions and are stored in the memory. This stored information can be used to check whether a given element belongs to the group or not. The major advantage in Bloom filter is the constant amount of memory needed to store the hashed values irrespective of length of the input element. Also, the amount of computation needed for hash functions and memory lookups are constant thereby making process faster. The designs are based on the concepts of spectral Bloom filters [2] which is an extended version of naive Bloom filter [3] and optimized version of counting Bloom filter [4]. They solve some of the problems such as multi-set query, insertion and deletion to Bloom in real-time which are not possible with original Bloom Filters. In spectral Bloom Filter [2], instead of bit vectors, an array of counters was implemented and is incremented or decremented according to corresponding insertions or deletions.

The rest of the paper is organized as follows: In Sect. 2 the background and related works are given; In Sect. 3, the design implementation of the work and description of the design; In Sect. 4, results and discussions of the experiment and Sect. 5 concludes the paper.

2 Related Works

The high speed Intrusion Detection System is an area of high opportunity, especially in hardware based NIDS. The increase in data speed has led to the requirement of dedicated hardware components and its improvement is necessary for a near perfect product. Roesch [1], developed a tool in 1999 called Snort, that can rapidly find the possible holes in network security. Sidhu and Prasanna [5] focused on methods that could convert the regular expression faster to hardware circuitry. They skipped the conversion to deterministic finite automaton (DFA), directly compiled regular expression to nondeterministic finite automaton (NFA). This implementation was extended in the works of Hutchings, Brad and Franklin [6] in 2002 which proposed a high speed Network Intrusion Detection system. The system extracted regular expression from snort with the help of java code which was then processed by the Xilinx software to create the FPGA input. They proposed an automated compiler that could convert regular expression automatically.

Bloom Filter from Bloom [3] was mainly used for checking the string and database applications. Broder and Mitzenmacher [7] uncovered the various applications of Bloom filters in networking, its modern variants, and the mathematical basis behind them. The main advantage of Bloom filter is that it takes only constant amount of memory to hash each of the elements irrespective of its length and also the computation involved in detecting an element is constant. Dharmapurikar and Krishnamurthy [8] in 2003 proposed a technique with Bloom filter to verify the membership of the queries. The focus was to implement the fast string matching in hardware based Intrusion Detection System. The design consist of bloom filters corresponding to each string length which ranges from a minimum value to size of window. In 2004 [9] they analyzed its performance against the finite automata solutions. Universal Hash function known as H3, is found to be suitable for implementation of hash table in hardware from the study conducted by Ramakrishna et al. [10].

Song and Lockwood [11] in 2005 suggested a method for long string matching to reduce the supported signature length. Three bit extended Bloom filter were chosen because of its scalability and fast incremental updating ability. Fan et al. [4] proposed an extension to bloom filter that could insert and delete from Bloom vector in real time. Cohen and Matias [2] optimized the work for multiset query and introduced two new algorithms. Pontarelli and Bianchi [12] proposed a system where instead of purely distributing the packets across the modules they grouped similar traffic packets and dispatched it to differently capable blocks. The design mainly used the header information to classify it into different categories and each module of hardware processes the disjoint rule sets. They followed a shift and compare architecture which was presented by Baker and Prasanna [13].

3 Design and Implementations

Let a string be processed by some multiple hash functions and they result in some values less than the size of the vector to which hash function map the string. Those values are set as bit positions in vectors. The query procedure is same as programming where the strings are checked for membership. The bit in corresponding values of vector is compared against the hash values of the queried string and if at least one value is found different then it is declared as a non member. The false positive probability f is given by [4]

$$f = \left(1 - \left(1 - \frac{1}{m} \right)^{nk} \right)^k \approx \left(1 - e^{-\frac{nk}{m}} \right)^k$$

where m is the size of the vector with n members and k hash functions. In optimal case, for a given value of m and n , we get number of hash functions with minimum false positivity,

$$k = \left(\frac{m}{n} \ln 2 \right).$$

The extended Bloom filter replaces the vector with array of counters. The counter corresponding to hash value increases when strings are inserted and decreased when deleted. When an item is queried or deleted it checks only the minimum counter value which should be greater than one if it existed in Bloom. When an item is inserted to Bloom filter, minimum counter is increased which is equivalent to increasing all the counters. The hash function implemented in this paper is the universal hash function, H3. This class of function are found suitable for hardware implementations [14] since they are bound to the limited memory resources. For any bit string X with bits represented as

$$X = \langle x_1, x_2, x_3, \dots, x_b \rangle$$

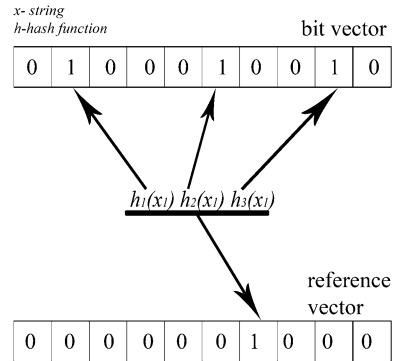
the i th hash function over X is calculated as,

$$h_i = d_{i1} \cdot x_1 \oplus d_{i2} \cdot x_2 \oplus d_{i3} \cdot x_3 \oplus \dots \oplus d_{ib} \cdot x_b$$

where ‘ \cdot ’ is a bitwise AND operator and ‘ \oplus ’ is a bitwise XOR operator. d_{ij} is a predetermined random number in the range $[0 \dots m - 1]$.

A second bloom vector of same size known as the reference vector is introduced in this paper. The resultant hash values of multiple hash functions from previous vector are hashed together to get a single value and is set in the reference bloom vector. The reference vector is considered only when there is a chance of false positivity. The reference vector is ignored when the first vector finds the item as a non member. This can further reduce the false positivity. The hash function uses simple XOR of previous hash function values to increase the lookup speed.

Fig. 1 Reference vector



The Fig. 1 shows the implementation of reference vector in Bloom filter and the hash function is given as

$$h_1 = h_1(x) \oplus h_2(x) \oplus \dots \oplus h_k(x)$$

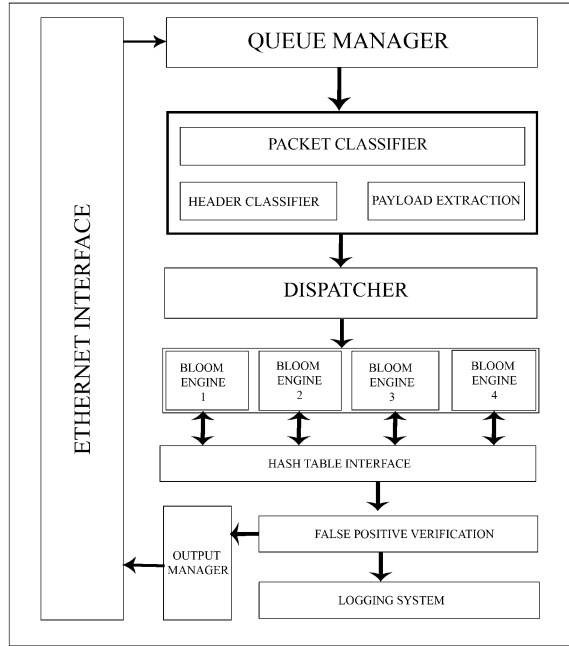
where x is the string and k is the number of hash functions in first vector.

3.1 Architectural Model

The architectural model of network intrusion detection based on reference Bloom Vector is shown in Fig. 2. It consists of an ethernet interface to which the network is connected. The queue manager record the packets and puts it in order. It allows synchronization between hardware and network. The packet classifier consists of header classifier and payload extractor. The header classifier extracts the header through the layers of protocol and the control is passed to the payload extractor. The information from payload extractor is given to the dispatcher unit which decides the distribution of the packet. It distributes the payload to one of the multiple bloom engines in such a way that the load is uniformly distributed. A Bloom engine consist of many Bloom filters, each having a different input window size. At each clock cycle, one byte is shifted in the window. The Bloom filter compares the resultant hash values of the input with the values stored in the hash table. The number of parallel bloom filters depends on the maximum length of the string to be compared and so window size is set from minimum to maximum length.

Multiple such Bloom engines are considered in order to increase the throughput of the system. These Bloom engines are connected to hash tables stored in the memory. Hash tables consist of stored hashes of attack patterns which are compared with hashed inputs in bloom filters. If the item is found to be a member, then the hash values are forwarded to False Positive Verifier. The reference vector is similar to a Bloom vector, except that the function intakes the Bloom filter hash values. The reference vector is already set with values corresponding to each string. If the

Fig. 2 Architectural model



comparison with the bloom vector gives a positive result with a certain error probability, the reference vector confirms the result, reducing the error rate. The alerting and logging systems are activated incase the output of reference vector is positive. The alerting mechanism notifies the administrator for the event and logging system saves the collected information to a file.

4 Results and Discussions

FPGA is the most feasible solution for wire-speed implementations. The design mainly focused on the minimum resource utilization in hardware. Different hardware implementable string matching algorithms were selected and analyzed, based on their complexities and features. Table 1 shows a brief comparison between

Table 1 Study of different String matching algorithms

Algorithm	Search method	Pre-processing	Search-time complexity
Brute force	Linear	No	$\mathcal{O}((n - m + 1)m)$
Rabin-Karp	Hash	Yes	$\mathcal{O}((n - m + 1)m)$
Morris Pratt	Heuristics based	Yes	$\mathcal{O}(m + n)$
Aho-Corasick	Automaton-based	Yes	$\mathcal{O}(m + z)$
Bloom filter	Hash	Yes	$\mathcal{O}(k)$

different hardware string matching algorithms. Bloom filter found to be the most compatible algorithm among them. It remembers only the flipping of bits in bit array and do not store any hashed keys. The selection of hash function, thus directly impacts the performance of the hardware. Five different non cryptographic hash functions were chosen for investigation. Universal Hash function, CRC 32, Pearson hash, Bit extraction hashing, XOR hashing were tested with uniformity test and avalanche test. Uniformity test analyze the distribution of hash values. Avalanche test checks whether a small change in input produce a large change in output. Considering the results from tests, H3 class of universal hash function was found better than the rest for hardware implementation.

The design is implemented in Xilinx Virtex II Pro with 4,176 Kbit block RAM [15]. Hash table were stored in SRAM with a total capacity of 4.5 Mbytes. Multiple engines with 20 distinct lengths can be scanned at a time. Despite the fact that the rate is kept at the same level as in other comparable works, the error rate is decreased to a small value. The system is designed in such a way that, the reference vector is inquired only after a positive outcome in bloom filter. The value $\frac{m}{n}$ gives the ratio between vector size and number of elements. The proper selection of m , n and k values can reduce the error probability.

The relationship between them clearly shows the size of bit vector (m) to be greater than the number of elements (n) and larger k values could reduce the errors. The Fig. 3 shows the comparison of false positive probability between old and new proposed system for different values of k with $\frac{m}{n} = 15$. The graph shows an enormous reduction in error probability making the system highly efficient.

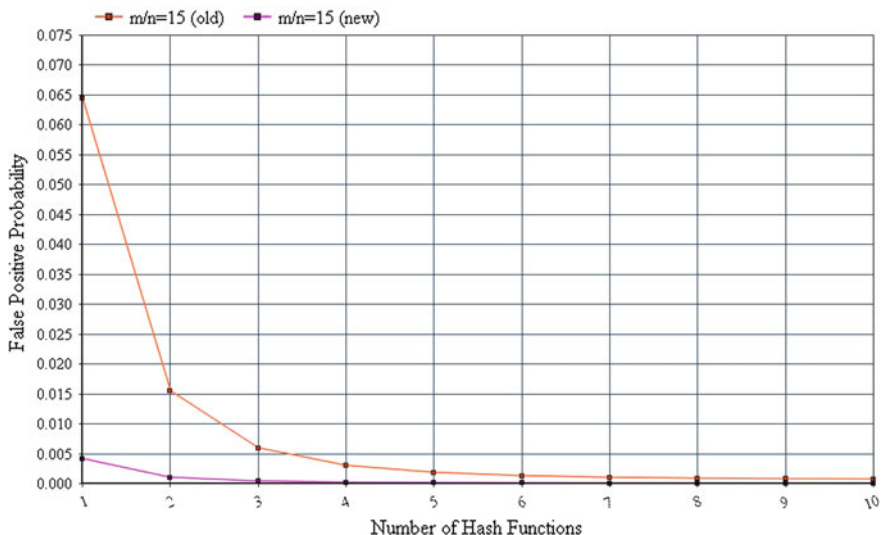


Fig. 3 Comparison of new error probability for various k values with $\frac{m}{n} = 15$

5 Conclusion

The paper proposed a network intrusion detection system in hardware platform to meet the current network requirements. The system adopted Bloom filter algorithm along with H3 hash function for the generation of bit vector. This paper presents an architecture for fast string matching with the help of multiple bloom engines. Further the design consists of a reference vector, which is meant to reduce the error rate in the system. An analysis of the trade-offs between number of hash functions and false positive probability has been presented. The FPGA based implementation is performed with the help of Xilinx Virtex Pro II FPGA board to support gigabit speed.

References

1. Roesch, Martin, et al.: Snort: lightweight intrusion detection for networks. *LISA* **99**, 229–238 (1999)
2. Cohen, S., Matias, Y.: Spectral bloom filters. In: Proceedings of the 2003 ACM SIGMOD international conference on Management of data, pp. 241–252. ACM (2003)
3. Bloom, B.H.: Space/time trade-offs in hash coding with allowable errors. *Commun. ACM* **13**(7), 422–426 (1970)
4. Fan, L., Cao, P., Almeida, J., Broder, A.Z.: Summary cache: a scalable wide-area web cache sharing protocol. In: *IEEE/ACM Transactions on Networking (TON)* **8**(3):281–293 (2000)
5. Sidhu, R., Prasanna, V.K.: Fast regular expression matching using FPGAs. In: The 9th Annual IEEE Symposium on Field-Programmable Custom Computing Machines, FCCM'01, pp. 227–238. IEEE (2001)
6. Hutchings, B.L., Franklin, R., Carver, D.: Assisting network intrusion detection with reconfigurable hardware. In: *Proceedings 10th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*. pp. 111–120. IEEE (2002)
7. Broder, A., Mitzenmacher, M.: Network applications of bloom filters: a survey. *Internet Math* **1**(4), 485–509 (2004)
8. Dharmapurikar, S., Krishnamurthy, P., Sproull, T., Lockwood, J.: Deep packet inspection using parallel bloom filters. In: *Proceedings of 11th Symposium on High Performance Interconnects*. pp. 44–51. IEEE (2003)
9. Dharmapurikar, S., Attig, M., Lockwood, J.: Design and implementation of a string matching system for network intrusion detection using FPGA-based bloom filters. In: IEEE Symposium on Field-Programmable Custom Computing Machines (FCCM04) (2004)
10. Ramakrishna, M., Fu, E., Bahcekapili, E.: A performance study of hashing functions for hardware applications. In: Proceedings of International Conference on Computing and Information, pp. 1621–1636 (1994)
11. Song, H., Lockwood, J.W.: Efficient packet classification for network intrusion detection using FPGA. In: *Proceedings of the 2005 ACM/SIGDA 13th international symposium on Field-programmable gate arrays*, pp. 238–245. ACM (2005)
12. Pontarelli, S., Bianchi, G., Teofili, S.: Traffic-aware design of a high-speed FPGA network intrusion detection system. *Trans. Comput. IEEE* **62**(11), 2322–2334 (2013)
13. Baker, Z.K., Prasanna, V.K.: Automatic synthesis of efficient intrusion detection systems on FPGAs. In: *Field Programmable Logic and Application*, pp. 311–321. Springer, Berlin (2004)

14. Hua, N., Norige, E., Kumar, S., Lynch, B.: Non-crypto hardware hash functions for high performance networking ASICs. In: *Proceedings of the 2011 ACM/IEEE Seventh Symposium on Architectures for Networking and Communications Systems*, pp. 156–166. IEEE Computer Society (2011)
15. Xilinx Inc. Virtex-II Pro and Virtex-II Pro X platform FPGAs: Complete data sheet (2004)

Hash-Based Rule Mining Algorithm in Data-Intensive Homogeneous Cloud Environment

Raghvendra Kumar, Prasant Kumar Pattnaik and Yogesh Sharma

Abstract Today Innovative Technology is used to analyze and manipulate huge amount of data in the cloud computing environment. It is very challenging task because the privacy and security are the main issue. Because the scenario of the cloud environment is given, then the distributed database comes in the picture as well as privacy. In this paper, we used the concept of pseudo random number, and for finding the strong Association rule in the database, we used the Inverted hashing and pruning as well as distributing the database into the different number of cloud nodes, and finding the global result, we used Distributed secure sum protocol in the homogenous cloud environments, where the number of attributes will be same, the number of transactions wearies from node to node.

1 Introduction

Data mining technology [1–3] is playing a very important role in analyzing the data in cloud environments [4–6] because the last few years cloud computing has received significant investments in the industry. And many cloud providers are participating in the market, forming competitive environments which are referred to as multiuser and multiproviders cloud market. Cloud computing (Internet) [7–9] is very emerging technology in today scenario. Because cloud computing reduces the both hardware and software cost, this provides the facility to the client as per their use.

R. Kumar (✉) · Y. Sharma
Faculty of Engineering and Technology, Jodhpur National University,
Jodhpur, Rajasthan, India
e-mail: raghvendraagrawal7@gmail.com

Y. Sharma
e-mail: yogeshsharma@gmail.com

P.K. Pattnaik
School of Computer Engineering, KIIT University, Bhubaneswar, India
e-mail: patnaikprasant@gmail.com

In simple word, if we say that the cloud computing is simply pay per use “if you want milk then you buy cow” no because if we want milk, then we simply purchase the milk and use it. So same way, we also define the cloud computing. Cloud computing Model divided into two types of models: deployment model and service model. Deployment model again divided into three models: public cloud, private cloud, and hybrid cloud. Cloud computing provides three main services: software as a service (SaaS), platform as a service (PaaS), and infrastructure as a service (IaaS), and in the distributed cloud computing environment, every node contains the resources that are physically distributed and logically connected with other resources to mining the result from the database; data mining has many different techniques like classification, association rule mining, clustering, etc. In this paper, we used one of the data mining techniques is association rule mining to find that the item is frequent or not. Association rule mining [10–12] is an important class of regularities within data which has been studied by the data mining experts. The problem of association rule [13–15] arises when we want to calculate the support and confidence of that particular item set in the distributed cloud computing environment, so in the proposed, we proposed an algorithm to solve that kind of problem and calculate the global support and confidence of the item or rule.

2 Proposed Work

In this proposed work, we assume that the data is distributed in the different numbers of cloud center, and every cloud center has more than one cloud node, and every cloud node has their own database that is horizontally partitioned database, and every cloud node wants to calculate their global result without revealing their private information to all the other cloud nodes that presents in the cloud center, and same for the cloud center, and for managing the high privacy to the distributed database used the concept of hash value and pseudo random number. Algorithm 1 shows the steps to how to work in this cloud environment, Fig. 1 shows the distribution of the cloud nodes and cloud center in the cloud environments, in which data is physically distributed but it is logically connected to all the other cloud center or cloud nodes present in the environments, and in this paper, RGPV University database is distributed in the different distributed database, the number of cloud center is three, and every cloud center contains only a single cloud node and the online database has been taken for cloud center 1 in cloud center 1, cloud node 2 in cloud center 2 and cloud node 3 in cloud center 3 with their hash value. In this paper, we consider that the threshold of support and support count are 40 % and 9 separately for all the cloud nodes presented in the distributed environments.

Algorithm 1

Input: Horizontal-Partitioned database $\{P_1, P_2, \dots P_n\}$, Pseudo Random Number $\{R_1, R_2, \dots R_n\}$ for all nodes presented in cloud computing environments when the number of nodes ($N \geq 3$)

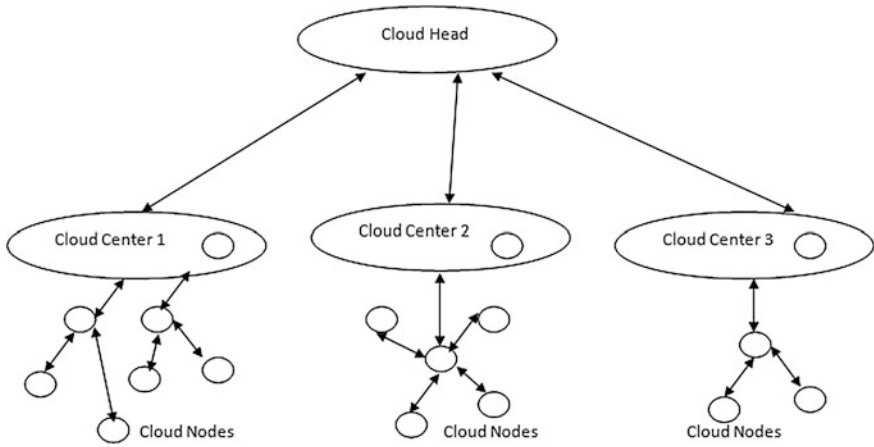


Fig. 1 The distribution of the cloud nodes and cloud center in the cloud environments

Output: Secure Global Support value for each frequent item sets

- Step 1: Partitioned the centralized database into the distributed database in cloud environment.
- Step 2: Now the each cloud center is again divided into the distributed database as cloud nodes.
- Step 3: Each cloud node contains the horizontally partitioned database $\{P_1, P_2, \dots, P_n\}$.
- Step 4: Each cloud node and cloud center select their own pseudo random number $R_{n1}, R_{n2}, \dots, R_{nm}$.
- Step 5: Arrange the entire cloud center in physically distributed but logically connected distributed database into the cloud distributed database environments.
- Step 6: Each cloud center has more than two cloud nodes (used the concept of secure multiparty computation).
- Step 7: Now the each cloud node contains the horizontally partitioned distributed database.
- Step 8: Select one of the cloud centers as the cloud head, that contains all the information of the rest of the cloud center and cloud nodes present in the cloud environment.
- Step 9: Now the cloud head sends the hello message to the rest of the cloud center presented.
- Step 10: After receiving the hello message, the cloud center sends the hello message to the cloud nodes presented in every cloud center.
- Step 11: Now all the cloud nodes send the reply message to their cloud center after adding the own pseudo random. Reply message + Pseudo Random Number (Cloud Nodes)

- Step 12: After that all the cloud centers send the activation message to cloud head after adding the own pseudo random. Reply message + Pseudo Random Number (Cloud Nodes) + Pseudo Random Number (Cloud Center)
- Step 13: Every Cloud node will calculate their hash value ($\text{Hash Value} = M \bmod N$) to convert Transaction table into the Hash table.
- Step 14: Now the converted hash table will calculate their own attribute value by considering their hash value.
- Step 15: Now consider only those attributes that have value greater than the minimum support.
- Step 16: Now the every cloud node will calculate their frequent item sets by using the Apriori Algorithm from their original horizontally partitioned database.
- Step 17: Every cloud node will calculate their own Partial Support (PS) by using $\text{PS}_{ij} = X. \text{Support} - \text{Threshold Support} * |\text{DB}|$
 for $I = 1$ to n
 {
 for $J = 1$ to n
 {
- Step 18: Each cloud node selects own pseudo random number.
- Step 19: After calculating partial support, all the cloud nodes will send the Partial support
 $\text{PS}_{ij} = X. \text{Support} - \text{Threshold Support} * |\text{DB}| + \text{Pseudo random number}$
- Step 20: Now the entire cloud center receives the partial support from their different cloud nodes and adds their own pseudo random number by using the following formula, $\text{PS}_{ij} = X. \text{Support} - \text{Minimum Support} * |\text{Size of the database}| + \text{Pseudo random number (Cloud Nodes)} + \text{Pseudo random number (Cloud Center)}$
- Step 21: The cloud head receives the encrypted partial support values from all other cloud center.
- Step 22: Then the cloud head decrypted
 $\text{PS}_{ij} = X. \text{Support} - \text{Threshold Support} * |\text{DB}| + \text{Pseudo random number (Cloud Nodes)} + \text{Pseudo random number (Cloud Center)} - \text{Pseudo random number (Cloud Center)}$
 Now again decrypted
 $\text{PS}_{ij} = X. \text{Support} - \text{Threshold Support} * |\text{DB}| + \text{Pseudo random number (Cloud Nodes)} - \text{Pseudo random number (Cloud Nodes)}$
- Step 23: Now cloud head calculates the global support count (GSC) by using the following formula
 $\text{GSC} = \text{Sum of all the PS} - \text{Threshold Support} * |\text{DB}|$
- Step 24: Cloud head broadcasts the global support count value to the other entire cloud center present in the cloud environments.
- Step 25: Now every cloud center broadcasts the global support value to all the cloud nodes present under in the cloud center. } }
- Step 26: End of Process

Now calculate the frequent item sets in cloud node 1 for cloud center 1 = {Attr1, Attr2, Attr3}

Consider that the item is {Attr1} for calculating their global support values from all the cloud environments that may be frequent locally or globally Partial support that is calculated for cloud node 1 $PS_1 = X$. Support – Threshold Support * |DB| = $12 - 4 * 16 = 5.6$

Calculate the number of frequent item sets in cloud node 2 for cloud center 2 = {Attr1, Attr2, Attr3, Attr4}, for cloud node 2 $PS_2 = 11 - 0.4 * 17 = 4.2$, Frequent item sets at cloud node 3 = {Attr1, Attr2, Attr3, Attr4}, $PS_3 = 11 - 0.4 * 14 = 5.4$, $PS_1 = 5.6$, $PS_2 = 4.2$, $PS_3 = 5.4$

If consider the first number for generation of pseudo random is 50, then square it, get a number is 2500, select two middle numbers from the squared number, the next value is again 50, then again square it, and select the two middle number after squaring it. Like that find the all pseudo random number depending upon the user requirements. So the $R_1 = 50$ for node 1, $R_2 = 50$ for node 2, $R_3 = 50$ for node 3. And every cloud center contains number of different cloud nodes but in this paper, we have consider that every cloud center contains only a single cloud node. After calculating the local or partial support valve, every cloud node adds their own private secure pseudo random number to send the cloud center for encrypting the data values, by using the above formula for cloud node 1: the value of partial support $PS_1 = 12 - 0.4 * 16 + 50 = 55.6$ For cloud node 2, the value of partial support $PS_2 = 11 - 0.4 * 17 + 50 = 54.2$ For cloud node 3, the value of partial support $PS_3 = 11 - 0.4 * 14 + 50 = 55.4$.

When the cloud nodes receive the encrypted partial support value from their different cloud nodes, the cloud center added their own value to encrypt the data again, so reduce the data leakage to other cloud center presents in the environment. Consider the pseudo random number is 60 for cloud center 1, 2, and 3. Now the partial support value for the cloud center 1 $PS_1(\text{Cloud Center 1}) = 12 - 0.4 * 16 + 50 + 60 = 115.6$. For cloud center 2, $PS_2 = 11 - 0.4 * 17 + 50 + 60 = 114.2$. For cloud center 3, $PS_3 = 11 - 0.4 * 14 + 50 + 60 = 115.4$. Now all the cloud center sends their encrypted partial support value to the cloud head, when the cloud head received that encrypted value from their different cloud center, the cloud center now decrypted that value by using the following formula $PS_{ij} = X$. Support – Minimum Support * |Size of the database| + Pseudo random number (Cloud Nodes) + Pseudo random number(Cloud Center) – Pseudo random number(Cloud Nodes)

Now the partial support value for the cloud center 1 in cloud head PS_1 (Cloud Center 1) = $12 - 0.4 * 16 - 50 - 60 = 5.6$, for cloud center 2, decrypted values $PS_2 = 11 - 0.4 * 17 - 50 - 60 = 4.2$, and for cloud center 3, decrypted values $PS_3 = 11 - 0.4 * 14 - 50 - 60 = 5.4$

After decrypting the partial support value, then the cloud head calculating total $PS = PS_1 + PS_2 + PS_3 - \text{Threshold Support} * |DB| = 5.6 + 4.2 + 5.4 - 0.4 * (16 + 17 + 14) = 15.2 - 18.8 = -3.6$

Global support count value is -3.6 that is less than zero so it is globally infrequent item sets; it may be frequent in the local cloud center or cloud center.

After calculating the global support value, the head node broadcasts that value to all other cloud centers present in the cloud environments, like that the entire cloud centers broadcast the global value to all other cloud nodes present in the cloud environments.

3 Conclusion

Most of the people preserve the privacy in the original dataset in the cloud environment using the concept of pseudo random number that is nonreproducible that is very difficult to machine to produce that number again for decrypting the original datasets. But in this paper, we used the concept of the pseudo random number that is reproducible and also for providing the highest privacy that the middle node never knows the result of the neighbor node so we changed the position of every node by using the distributed secure sum protocol. This paper implements the new algorithm and implementation to overcome the privacy to the data in the cloud computing environment. And this paper is also to reduce the complexity of finding association rule in the huge amount of data because we used the multi-path inverted hashing and pruning concept; the complexity of this algorithm is very less as compare to the well-recognized Apriori algorithm, and also the data leakage is zero percent.

References

1. Yao, A.C.: Protocol for secure computations. In: Proceedings of the 23rd annual IEEE symposium on foundation of computer science, pp. 160–164. IEEE Press, Chicago, USA, 1982
2. Yao, A.C.C.: How to generate and exchange secrets (extended abstract). In: Proceedings of the 27th IEEE Symposium on Foundations of Computer Science (FOCS). IEEE Press USA (1986)
3. Agrawal, R. et al.: Mining association rules between sets of items in large database. In: Proceedings of ACM SIGMOD'93, pp. 207–216. D.C. ACM Press, Washington (1993)
4. Li, D.Y., Li, D.R.: Mining association rule with Linuistic cloud model [J]. *J. Softw.* **2**, 143–158 (2000)
5. Clifton, C., Lin, D.: Tool for privacy preserving distributed data mining [J]. *SIGKDD Explorations* **2**, 28–34 (2002)
6. Lindell, Y.: Privacy preserving data mining [J]. *J. Cryptog.* **3**, 177–206 (2002)
7. Chen, X., Orłowska, M.: A new framework for privacy preserving data sharing. In: Proceedings of the 4th IEEE ICDM Workshop: Privacy and Security Aspects of Data Mining, pp. 47–56. IEEE Computer Society (2004)
8. Mielikainen, T.: On inverse frequent set mining. In: Proceedings 3rd IEEE ICDM Workshop on Privacy Preserving Data Mining, pp. 18–23. IEEE Computer Society (2003)
9. Du, W., Atallah, M.: Secure Multi-party Computation: A Review and Open Problems. CERIAS Tech. Report 2001-51, Purdue University (2001)
10. Srikant, R., Agrawal, R.: Mining generalized association rules. In: VLDB'95, pp. 479–488 Zurich, Switzerland, 1994
11. Agrawal, R., Srikant, R.: Privacy-preserving data mining. In: Proceedings of the 2000 ACM SIGMOD on management of data, pp. 439–450. ACM Press, Dallas, TX USA (2000)

12. Lindell, Y., Pinkas, B.: Privacy preserving data mining. In: Proceedings of 20th Annual International Cryptology Conference (CRYPTO), Santa Barbara, California, USA (2000)
13. Clifton, C., Kantarcioglou, M., Xiadong, L., Michael, Y.Z.: Tools for privacy preserving distributed data mining. *SIGKDD Explorations* **4**(2), 43–48 (2002)
14. Vaidya, J., Clifton, C.: Privacy-Preserving Decision Trees over vertically partitioned data. *Lecture Notes in Computer Science*, vol. 3654 (2005)
15. Ioannidis, I., Grama, A., Atallah, M.: A secure protocol for computing dot-products in clustered and distributed environments. In: Proceedings of International Conference on Parallel Processing, pp. 379–384, 18–21 Aug 2002

Privacy Preservation in Distributed Environment Using RSA-CRT

Raghvendra Kumar, Prasant Kumar Pattnaik and Yogesh Sharma

Abstract Most data mining applications are based on information sharing and additional challenges, when we are dealing with data that are containing sensitive or private information. There is no common data mining technique available that deals with private information without any leakage. Therefore, the knowledge extracted from such data may disclose the pattern with sensitive/private information. This may put privacy on the individual/group of parties. In the past few years, privacy preserving data mining has attracted the research interest and potential for a wide area of applications. There are many techniques for privacy preservation like cryptography, anonymity, and randomization, etc., experimented for privacy preservation in data mining. In this paper, information system-based approach is considered, so some of the attributes required higher privacy compared to the other attributes. This paper explores the use of cryptography techniques, namely RSA with Chinese remainder theorem (CRT) to encrypt and decrypt the database, when all the parties are trying to find their global result in presence of trusted third party or absence of trusted third party, without disclosing their private information to each other.

1 Introduction

Due to the tremendous growth of data and Internet, [1, 2], collection and analysis of huge amount of data is a big question in today's era because data is used everywhere like business strategy, scientific calculation, and research in the field of

R. Kumar (✉) · Y. Sharma
Faculty of Engineering and Technology,
Jodhpur National University, Jodhpur, Rajasthan, India
e-mail: raghvendraagrawal7@gmail.com

Y. Sharma
e-mail: yogeshsharma@gmail.com

P.K. Pattnaik
School of Computer Engineering, KIIT University, Bhubaneswar, India
e-mail: patnaikprasant@gmail.com

medical [3], etc. However, when data is distributed in the distributed environment then the main problem is privacy and security. Some useful information such as customer buying habits, electrical records, medical record, research records, and bank details, etc., are leaked due to a centralized database. So, the main problem is privacy in the distributed database; some of the methods are described below are divided into four different types

1. Privacy preservation in the centralized database [4]
2. Rule hiding algorithm [5]
3. Privacy preservation in the distributed database [6, 7]
4. Privacy preservation in the rule mining algorithm [8]

In this paper, we used the combination of RSA [7–9], CRT [10, 11], and SMC [6] to provide high privacy in the distributed data without disclosing their private information to the other party present in the environment.

2 Proposed Work

In this paper, an honest and semi-honest model for adversary is used where each party follows correctly the protocol for secure computing, but they do not want to disclose their personal information to all the parties presented. The proposed algorithms presented a method for privately computing data mining process from distributed data sources disclosing any information about the data sources or their data except those that revealed final global results [6, 12, 13]. In this paper, we proposed four algorithms for considering the two different scenarios. Algorithm 1 gives the concept when the absence of trusted third party and number of parties is two, but they try to find the common function or common result for both of them without disclosing their private information to each other. Algorithm 2 gives the concept when the trusted third party is presented and the number of parties is two, but they try to find the common function or common result for both of them without disclosing their private information to each other, so that trusted third party calculates their common result or global result and sends to both the parties. Algorithm 3 gives the concept when the absence of trusted third party and number of parties is more than three, but they try to find the common function or common result for both without disclosing their private information to each other, so in this scenario party P1 is considered as a trusted party for calculating the global result or global information. Algorithm 4 gives the concept when the presence of trusted third party and number of parties is more than three, but they try to find the common function or common result for both without disclosing their private information to each other, so in this scenario, trusted third party is considered as trusted party for calculating the global result or global information.

Algorithm 1: Absence of trusted third party (TTP)

Input: $M = 2$ Parties, y_i class values, L attribute values, x_p query instance $\{x_1, x_2 \dots x_n\}$

- Step 1: Let us consider the scenario when the number of parties is two and the parties (P_1 and P_2) have their own database DB_1 and DB_2 , respectively.
- Step 2: Now parties P_1 and P_2 , consider two large prime number P_i and Q_i , respectively, and calculate $N_i = P_i \times Q_i$
- Step 3: $\{e_i, d_i\}$ keys of RSA algorithm that are generated by EDMS [10].
- Step 4: Now parties, P_1 and P_2 calculate the number of frequent and infrequent item sets [5] from their datasets with considering the threshold value.
- Step 5: After that parties P_1 and P_2 deliberate the PS
for $i = 1$ to 2 do
 $PS_i = X_i \cdot \text{Support} - \text{Minimum Support} * |DB_i|$
- Step 6: Encrypt PS_i to get homo orphic encryption $E_i(PS_i)$ (RSA-CRT Key generation) for parties P_1 and P_2
- Step 7: RSA encrypts P_i to $C_i = E_i(PS_i)$ and transport C_i to the EDMS
- Step 8: Now party P_2 sends C_i to party P_1
- Step 9: When party P_1 received the encrypted result from party P_2 , party P_1 Decrypt $E_i(PS_i)$ (RSA-CRT Decryption)
- Step 10: After that, calculate the common result from both parties P_1 and P_2 , which is represented as global result. (Global Result/Support = Sum of original values/ N = Total number of parties)

Algorithm 2: Presence of trusted third party (TTP)

Input: $M = 2$ parties, y_i class values, L attribute values, $x_p \{x_1, x_2 \dots x_n\}$ query instance

- Step 1: Let us consider the scenario when the number of parties is two and the parties (P_1 and P_2) have their own database DB_1 and DB_2 , respectively.
- Step 2: Now parties P_1 and P_2 , consider two large prime numbers P_i and Q_i , respectively, and calculate $N_i = P_i \times Q_i$
- Step 3: $\{e_i, d_i\}$ keys of RSA algorithm that are generated by EDMS [10].
- Step 4: Now the parties P_1 and P_2 calculated the number of frequent and infrequent item sets [5] from their datasets with considering the threshold value.
- Step 5: Trusted third party sends hello message to all the parties presented in the environment.
- Step 6: When all the parties receive the hello message they send the reply message to the third party.
- Step 7: After that party, P_1 and P_2 calculate the PS,
for $i = 1$ to 2 do
 $PS_i = X_i \cdot \text{Support} - \text{Minimum Support} * |DB_i|$
- Step 8: Encrypt PS_i to get homo orphic encryption $E_i(PS_i)$ (RSA-CRT Key generation) for parties P_1 and P_2
- Step 9: RSA encrypts P_i to $C_i = E_i(PS_i)$ and transport C_i to the EDMS
- Step 10: Now parties P_1 and P_2 send the C_i to TTP

- Step 11: When the TTP is received, the encrypted value from P_1 and P_2 , the trusted third party decrypt E_i (PS_i) (RSA-CRT Decryption)
- Step 12: After that, trusted third party calculates the global result from both parties P_1 and P_2 , broadcast the results to both the parties presented in the environment. (Global Result/Support = Sum of original values/
 N = Total number of parties)

Algorithm 3: Absence of trusted third party (TTP) and number of parties is more than three

Input: M Parties, y_i Class values, L Attribute values, x_p $\{x_1, x_2 \dots x_n\}$ query instance

- Step 1: Let us consider the scenario when the number of parties $m \geq 3$ (P_1, P_2, \dots, P_m) have their own database DB_i for $I = 1$ to m .
- Step 2: P_i and Q_i are large security prime numbers and calculate $N_i = P_i \times Q_i$
- Step 3: $\{e_i, d_i\}$ keys of RSA algorithm that are generated by EDMS [10].
- Step 4: Now the parties ($P_1, P_2 \dots P_m$) calculated the number of frequent and infrequent item sets [5] from their datasets considering the threshold value and consider P_1 as protocol initiator party.
- Step 5: After the parties ($P_1, P_2 \dots P_m$) calculate PS value
for $i = 1$ to m do
 $PS_i = X_i$. Support-Minimum Support * $|DB_i|$
- Step 6: Encrypt PS_i to get homo orphic encryption E_i (PS_i) (RSA-CRT Key generation) for the parties ($P_1, P_2 \dots P_m$)
- Step 7: RSA encrypts P_i to $C_i = E_i$ (PS_i) and transport C_i to the EDMS, for $i = 1$ to m do
- Step 8: Now all the parties for $i = 2$ to m send the PS_i to the P_1
- Step 9: When the party P_1 received the encrypted result from the parties ($P_2, P_3 \dots P_m$), P_1 decrypt E_i (PS_i) (RSA-CRT Decryption) for $i = 2$ to m do
- Step 10: After that, party P_1 , calculates the global support value and broadcast global support value to all the parties presented. These values are locally frequent for some parties and infrequent for some other parties. (Global Result/Support = Sum of original values/ N = Total number of parties)

Algorithm 4: Presence of trusted third party and number of parties is more than three

Require: M parties, y_i Class values, L Attribute values, x_p $\{x_1, x_2 \dots x_n\}$ query instance

- Step 1: Let us consider the scenario when the number of parties $m \geq 3$ ($P_1, P_2 \dots P_m$) have their own database DB_i for $I = 1$ to m .
- Step 2: P_i and Q_i are large security prime numbers and calculate $N_i = P_i \times Q_i$
- Step 3: $\{e_i, d_i\}$ keys of RSA algorithm that are generated by EDMS [10].

- Step 4: Now the Parties $\{P_1, P_2 \dots P_m\}$ calculated the number of frequent and infrequent item sets [5] from their datasets considering the threshold value and consider trusted third party as protocol initiator party.
- Step 5: Trusted third party sends the hello message to all the parties present in the environment.
- Step 6: When all the parties receive the hello message they send the reply message to the third party.
- Step 7: After the parties $\{P_1, P_2 \dots P_m\}$ calculate PS,
for $i = 1$ to m do
 $PS_i = X_i \cdot \text{Support} - \text{Minimum Support} * |DB_i|$
- Step 8: Encrypt PS_i to get homo orphic encryption $E_i (PS_i)$ (RSA-CRT Key generation) for the parties $\{P_1, P_2 \dots P_m\}$
- Step 9: RSA encrypts P_i to $C_i = E_i (PS_i)$ and transport C_i to the EDMS, for $i = 1$ to m do
- Step 10: Now all the parties for $i = 1$ to m send the PS_i to the trusted third party
- Step 11: When the TTP received the encrypted result from parties $\{P_1, P_2 \dots P_m\}$, trusted third party decrypt $E_i (PS_i)$ (RSA-CRT Decryption) for $i = 1$ to m do
- Step 12: After that, trusted third party calculates the global support value and broadcasts global support value to all the parties presented. These values are locally frequent for some parties and infrequent for some other parties. (Global Result/Support = Sum of original values/N = Total number of parties)

3 Conclusion

The use of the RSA and CRT algorithms preprocess the required privacy preservation proposed and discussed in this paper. The data are encrypted before being communicated to the trusted third party or protocol initiator party that applies the data mining function/global results. This paper proposes novel algorithms to secure the sensitive and private information in distributed database environments when the number of parties is two or more than two. In the future, this work will be modified using the different privacy preservation algorithms in distributed database environment.

References

1. Apostol, T.: Introduction to Analytical Number Theory, Student edn. Springer International (1989)
2. Boneh, D., Shacham, H.: Fast variants of RSA. Crypto Bytes, Springer **5**(1), 1–9 (2002)
3. Sun, H.M., Wu, M.E.: An approach towards rebalanced RSA-CRT with short public exponent. Cryptology ePrint Archive: Report 2005/053 (2005)

4. Agrawal, S., Krishnan, V., Haritsa, J.R.: On addressing efficiency concerns in privacy-preserving mining. In: Proceedings of the 9th International Conference on Database Systems for Advanced Applications. LNCS 2973, pp. 113–124. Springer-Verlag, Jeju Island
5. Peng, Z., Yun-Hai, T., Shi-Wei, T., Dong-Qing, Y., Xiu-Li, M.: An effective method for privacy preserving association rule mining. *J. Softw.* **17**(8), 1764–1773 (2006)
6. Linedell, Y., Pinkas, B.: Secure multiparty computation for privacy-preserving data mining. *J. Privacy Confidentiality* **1**(1) 59–98 ((2009))
7. Paillier, P.: Public-key cryptosystems based on composite degree residuosity classes. *Adv. Cryptog. EUROCRYPT '99*, pp. 223–238. Prague, Czech Republic (1999)
8. Rivest, R., Adleman, L., Dertouzos, M.: On data banks and privacy homomorphism's. In: R. A. De Milloetal (ed.) *Foundations of Secure Computation*, pp. 169–179. Academic Press (2000)
9. Qiong, G., Xiao-hui, C.: A privacy preserving distributed for mining association rules. *Int. Conf. Artif. Intell. Comput. Intell.* 294–297 (2009)
10. Koblitz, N.: *A Course in Number Theory and Cryptography*, 2nd edn. Springer (1994)
11. Stallings, W.: *Cryptography and network security*, 3rd edn. Pearson Education
12. Yao, A.C.: Protocol for secure computations. In: Proceedings of the 23rd Annual IEEE Symposium on Foundation of Computer Science, pp. 160–164 (1982)
13. Gold Reich, O., Micali, S., Wigderson, A.: How to play any mental game. In: Proceedings of the 19th Annual ACM Symposium on Theory of Computation, pp. 218–229 (1987)

Priority-Based Classification: An Automated Traffic Approach

Somasree Bhadra and Anirban Kundu

Abstract In this paper, an advanced methodology is proposed to automate the traffic system by categorizing the incoming vehicles. Vehicles are classified as 'Public' and 'Private' transport. 'Public' transport is considered to carry large number of people. It is considered in this paper that avoidance of traffic congestion and man power wastage are achieved by releasing public vehicles with highest priority. Lanes are categorized as high priority lane (L_P), normal lane (L_N), and idle lane (L_I). Probability of waiting in the queue for an incoming vehicle is measured based on the Erlang distribution method. Avoidance of traffic congestion and manpower wastage due to indefinite waiting time in traffic is handled efficiently by our proposed approach. It is presumed that efficiency and productivity of human resource are increased by providing efficient and smooth congestion-free transport system. Minimum production time is expected from the human resources; hence usage time of electronic resource is minimized. Power and energy consumption are minimized as an indirect effect of efficient traffic system.

Keywords Traffic automation · Energy efficient system · Smart automation · Erlang

S. Bhadra (✉)

Swami Vivekananda Institute of Science and Technology, Sonarpur,
Kolkata, West Bengal, India
e-mail: cse.soma@gmail.com

A. Kundu

Netaji Subhash Engineering College, Garia, Kolkata, West Bengal, India
e-mail: anik76in@gmail.com

S. Bhadra · A. Kundu

Innovation Research Lab, West Bengal, India

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_5

1 Introduction

It is observed that manpower wastage is increased day by day due to the ever increasing traffic. Lanes are controlled and released by predefined algorithm irrespective of the lane status. A dynamic lane release algorithm is required to control the lanes based on the run time data. Maximum usage of man power is achieved minimizing the transport time. It is evident that pollution rate and number of vehicles on the road are directly proportional. Hence, pollution rate is decreased by minimizing the number of active vehicle on the road. People are encouraged to use public transport by offering minimum traffic for public vehicle.

Vehicular ad hoc network is considered as one of the most prospective fields of research in recent times due to its obvious impact in transportation system [1]. An efficient, opportunistic strategy is introduced in [2] to enhance reliable vehicle-to-vehicle communication by exploiting the infrastructure through access points [2]. In [3], an efficient approach is proposed to collect and disseminate real-time congestion information through GPS to identify vehicles' location and speed [3]. Feasibility of information dissemination in vehicular networks is studied in [4].

Erlang distribution method is widely used in telecommunication sector to avoid call waiting. Number of available attainer, incoming call rate, and other parameters are considered to calculate the probabilistic value of call waiting. Incoming calls are diverted accordingly to achieve maximum resource utilization.

The proposed approach is considered as power and energy saving by implementing Erlang method due to the simple operational activity and optimum time complexity. In this paper, minimum number of parameters is considered and simple method is used to automate the traffic system. Time complexity is made optimum as the number of operand is optimum.

The rest of the paper is organized as follows: related works have been discussed in Sect. 2; proposed system framework is discussed in Sect. 3. Experimental results have been described in Sect. 4; conclusion is depicted in Sect. 5.

2 Related Works

Traffic congestion is increased and the traffic control efficiency is reduced by the ever increasing traffic in recent years. Many approaches are introduced by the researchers to enhance the performance of traffic controlling system. Intelligent traffic system using multi-agent technology is proposed in [5]. Lanes are categorized dynamically by implementing fuzzy set theory [5]. Image processing concept is implemented to monitor and evaluate the real-time traffic system in [6]. Instead of tracking entire vehicles, sub-features are tracked to increase the robustness of the system [7]. Fuzzy logic is used to determine continuous and six discrete levels of congestion, respectively [8, 9].

Different automated road traffic systems are proposed by researchers in recent times. In [10], five traffic statuses are estimated from video images using hidden Markov models. Different versions of Erlang method are widely used to distribute, analyze, and calibrate the traffic system [11–13]. Different energy saving methods are implemented to reduce energy consumption in domestic environment [14–16]. Different green services in corporate environment are proposed in [17].

An automatic road traffic control and monitoring system for day time sequences is presented in [18]. Mean speed, dimension, and vehicles counting are considered as the decisive parameters. These features are collected through computerized vision detection method. A frame-differencing algorithm and texture information based on gray scale intensity are used to extract data from moving objects [18]. Inductive loops are placed in a roadbed to examine vehicles. Speed, length, and weight of vehicles and the distance between them are inspected in more sophisticated approaches [19, 20]. In [21], researchers have introduced a new method to monitor congested freeways. A feature-based tracking approach is constructed for the task of tracking vehicles under congestion. Instead of tracking entire vehicles, vehicle sub-features are tracked to make the system robust to partial occlusion [21]. A new framework is designed in [22] toward a real-time automated recognition of traffic accident based on the Histogram of Flow Gradient (HFG) and statistical logistic regression analysis [22].

In [23], traffic management is controlled by implementing video monitoring and surveillance system. Required information is gathered through surveillance and required measure is taken to provide better traffic flow. Traffic density of the lanes is measured by this approach. Priority-based signaling system is implemented on the particular lane to avoid traffic congestion. Release sequence of the lanes is also controlled by this approach.

A Web-based traffic management system is proposed in [24]. Master–Slave architecture is used to implement remote-controlled traffic automation system. Slave nodes are placed in different geographical locations. Entire data are controlled and analyzed by the Master node. Traffic system is controlled through the Internet [24].

Automated traffic management system is proposed in [25] to prioritize Ambulance services. The proposed system consists of an “Intelligent Transport System (ITS).” Ambulance location is tracked by intelligent sensor system. Nearest hospital location and shortest path to reach the hospital are calculated by the central unit. Traffic signals are controlled accordingly to provide fastest and congestion-free traffic to the Ambulance [25].

Automated vehicle detection in traffic is important in many aspects. Tracking a specific vehicle automatically is important in traffic surveillance. Vehicle detection is typically used to track vehicle, analyze road traffic situation, and security purpose [26]. Vehicle tracking is implemented by using differential morphological closing profile [27].

3 Proposed Work

Different automated traffic management systems are proposed by the researchers to provide better traffic flow. Traffic congestion and vehicle density are considered in existing approaches. In our paper, man power wastage is considered as the major constraint. Hence, an intelligent traffic system approach is proposed to minimize the man power and energy wastage.

Lanes are categorized based on some predefined assumption in typical automated traffic system. Ever increasing traffic is not controlled efficiently by this predefined method. A dynamic energy saving traffic automation method is introduced in this paper. Lanes are classified as high priority lane (L_P), normal lane (L_N), and idle lane (L_I) based on “Car Density” and “Waiting Time.” “Car Density” is defined as the number of cars passing per unit time. Lane status is measured as a combination of “Car Density” and “Waiting Time” as shown in Table 1.

High priority lane (L_P) is released first at any time instance, and then followed by normal lane (L_N) and idle lane (L_I). A lane is designated as L_P based on the importance of that lane at that time instance.

It is assumed that a lane designated as ‘ L_N ’ has the high probability to be transformed into ‘ L_P ’ lane. Hence incoming vehicles on ‘ L_N ’ are classified and sent to by lanes to avoid probable traffic congestion.

Incoming vehicles are selected based on the probability that the vehicle has to wait in the lane. The probability is calculated by existing Erlang C formula as shown in Eq. (1) [28].

$$P_w = \frac{\frac{A^N}{N!} \frac{N}{N-A}}{\sum_{i=0}^{N-1} \frac{A^i}{i!} + \frac{A^N}{N!} \frac{N}{N-A}}, \quad (1)$$

where

- P_w Probability that a vehicle has to wait on the lane;
- A Total traffic offered in the units of Erlangs; and
- N Stoppage time on the lane.

In our paper, total offered traffic in Erlang [28] is defined as follows:

$$A = \sum_{i=0}^N T_i \quad (2)$$

Table 1 Lane status measurement table

	Low car density	Moderate car density	High car density
Minimum waiting time	L_I	L_I	L_P
Medium waiting time	L_I	L_N	L_P
Maximum waiting time	L_N	L_N	L_P

where

N Number of cars waiting on the lane; and

T_i Waiting time of a car on the particular lane.

It is presumed that lanes are subdivided as ‘Public Lane (L_{PU})’ and ‘Private Lane (L_{PR}).’ Incoming vehicles are categorized as ‘Public’ and ‘Private.’ Categorized vehicles are sent to appropriate lane as shown in Table 2.

Incoming vehicles on ‘ L_N ’ are categorized and sent to the newly created by lanes to avoid heavy traffic. Hence, status of the lane remains unchanged.

High traffic congestion on L_P lane is avoided by implementing similar categorization process on incoming vehicles on this lane as depicted in Algorithm 1. A threshold value (T_H) is considered to measure the frequency of incoming vehicles (F_R). T_H is calculated based on Eq. (3).

$$T_H = \frac{\text{Number of vehicles arriving on the lane}}{\text{Time}} \tag{3}$$

Algorithm 1: Busy_Lane_incoming_vehicles_classification

Input: Frequency of incoming vehicles (F_R).

Output: Categorization and diversion of incoming vehicles.

Table 2 Vehicle categorization table

Probabilistic value	Vehicle type	By Lane
≤30 %	Public	Lane remains unchanged
	Private	
>30 %	Public	L_{PU}
	Private	L_{PR}

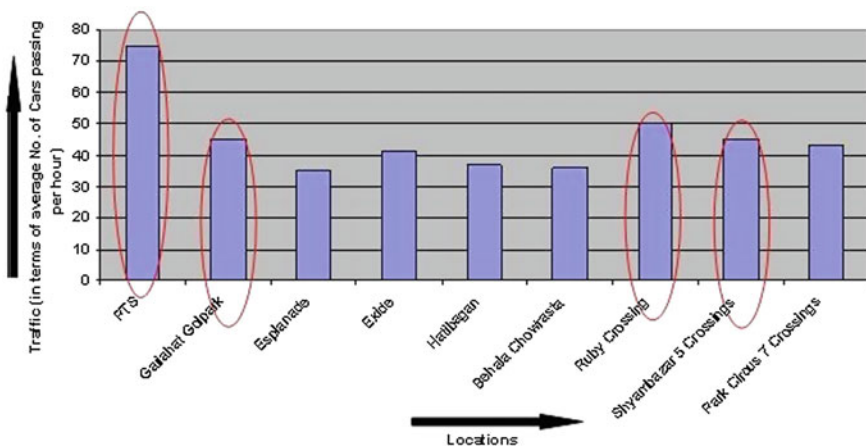


Fig. 1 Traffic data of busy crossings of Kolkata, West Bengal

- Step 1: Start.
- Step 2: If $T_H \leq F_R$, then go to Step 6.
Otherwise, go to Step 3.
- Step 3: If $T_H > F_R$, then, go to Step 4.
- Step 4: Divide the lane as L_{PU} and L_{PR} .
- Step 5: If the vehicle is ‘Public,’ then send it to L_{PU} .
Otherwise, send it to L_{PR} .
- Step 6: Repeat Step 2 to Step 5 for next incoming vehicle.

It is presumed that maximum people is accumulated in ‘Public’ vehicle. Hence, a huge man power is wasted by the typical release algorithm. It is considered that importance of ‘Public’ vehicle in terms of man power wastage is very high. Hence, more efficient and feasible outcome is achieved by introducing priority-based release concept. ‘Public’ vehicles are considered as the highest priority vehicle to be released.

It is also considered that if a lane is not released for a long-time starvation problem may arise. Hence, each lane is checked in every time interval and the lane that is not released for a long time is prioritized and released. It is presumed that

Table 3 Real-time traffic data for selected points

Location	Number of intersecting lanes	Number of cars passing per day
PTS crossing	4	75,000
Golpark more	4	45,000
Ruby more	5	50,000
Shyambazar 5 points	5	45,000

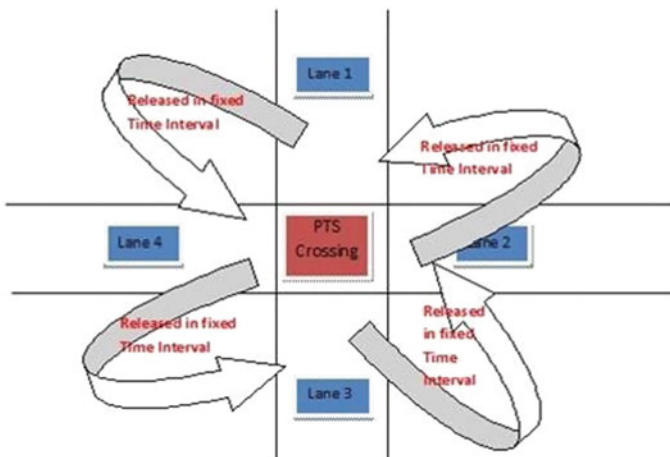


Fig. 2 Lane release scenario in typical traffic system

Table 4 Real-time traffic scenario of “PTS Crossing” on a particular day and time

Lane	6.00 pm		6.02 pm		6.04 pm		6.06 pm		Total waiting time
	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	
Lane 1	00	Running	15	Waiting	20	Waiting	23	Waiting	4 h 40 min
Lane 2	10	Waiting	00	Running	10	Waiting	12	Waiting	
Lane 3	05	Waiting	07	Waiting	00	Running	05	Waiting	
Lane 4	10	Waiting	13	Waiting	15	Waiting	00	Running	

total waiting time in traffic is directly proportional to the wastage of human resources in terms of time. Hence wastage of human resource is minimized by our proposed automated road traffic system.

4 Experimental Results

Real-time traffic data from different busy crossings have been collected and are shown in Fig. 1. Four busiest lanes are marked in Fig. 1.

Detailed traffic data of four lanes are shown in Table 3.

It is observed that traffic of these important road crossings are controlled by predefined automated traffic system. Vehicle release algorithms are predefined. Specific lanes are released in a cyclic manner as shown in Fig. 2. Real-time traffic scenario at “PTS Crossing” is shown in Table 4.

Consider, total ‘ n ’ numbers of cars are waiting for ‘ m ’ minutes on lane ‘ x ’ at time ‘ $t1$ ’. Then, total waiting time on lane ‘ x ’ at time ‘ $t1$ ’ is calculated as “ $m * n$ ”.

Energy wastage in terms of human resource in existing traffic approach is depicted in Fig. 3.

Lane 3 is designated as ‘ L_p ’ implementing our proposed approach based on inputs “Car Density” and “Waiting Time.” Lane 2 and Lane 4 are designated as ‘ L_N ’ by our proposed approach. Vehicles in Lane 3 are categorized and sent to appropriate by lanes. Probability that incoming vehicles towards Lane 2 and Lane 4 have to wait is examined using Erlang C formula and required action as depicted in this paper is executed to avoid traffic congestion. Traffic scenario of “PTS Crossing” after implementing our proposed approach on same data is shown in Table 5.

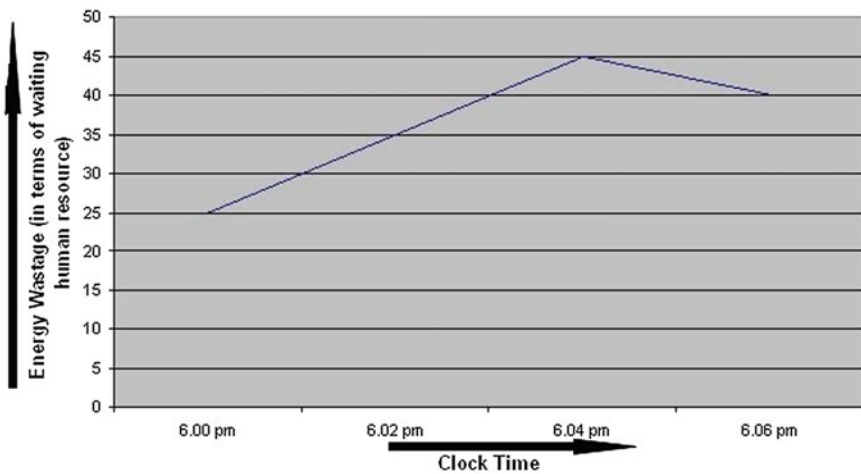


Fig. 3 Energy wastage in existing traffic approach

Table 5 Traffic scenario of “PTS Crossing” intersection on a particular day using proposed approach

Lane	6.00 pm		6.02 pm		6.04 pm		6.06 pm		Total Waiting time 3 h 24 min
	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	Number of cars in waiting mode	Status	
Lane 1	00	Running	00	Running	00	Running	15	Waiting	
Lane 2	10	Waiting	00	Waiting	10	Waiting	12	Waiting	
Lane 3	05	Waiting	07	Waiting	00	Waiting	05	Waiting	
Lane 3	10	Waiting	13	Waiting	15	Waiting	00	Running	

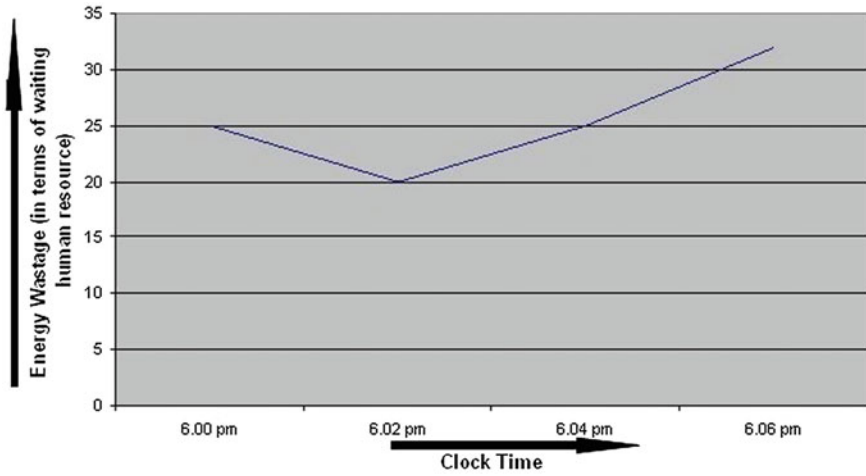


Fig. 4 Energy wastage by our proposed traffic approach

Table 6 Experimental result from ruby more, park circus 7 points, shyambazar 5 points at a particular time

Intersection (crossing) points	Total waiting time using typical method	Total waiting time using proposed method
Golpark more	6 h 27 min	54 h 13 min
Ruby more	5 h 46 min	44 h 27 min
Shyambazar 5 points	4 h 35 min	34 h 12 min

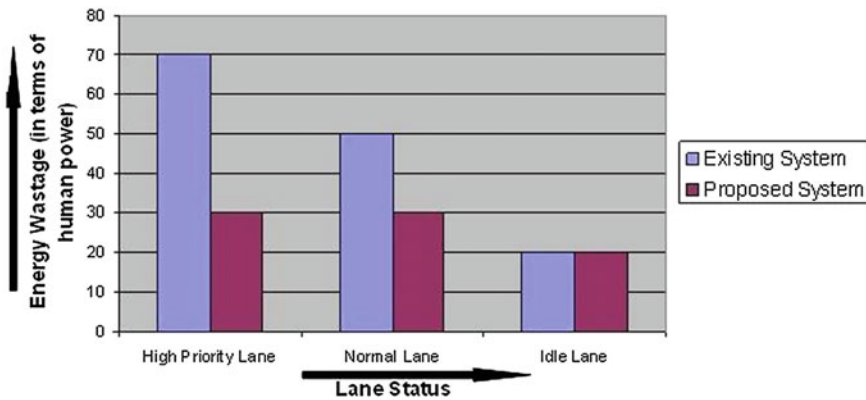


Fig. 5 Comparative energy wastage analysis between existing system and proposed system

Energy wastage in terms of human resource by our proposed approach is depicted in Fig. 4.

Experimental data are also being collected from other three intersections (crossings) implementing our approach as shown in Table 6.

It is observed from the above experimental data that more efficient and energy saving automated traffic system is achieved by implementing our proposed approach. Comparative efficiency analysis between existing approach and our proposed approach is shown in Fig. 5.

5 Conclusion

An advanced energy saving automated traffic system is proposed in this paper. Lanes are categorized as 'L_P,' 'L_N,' and L_I based on "Car Density" and "Waiting Time." Waiting vehicles in 'L_P' are classified as 'Public' and 'Private' and sent to appropriate by lanes. Similarly treatment is implemented on incoming vehicles towards 'L_N' lane to avoid traffic congestion based on the probabilistic value calculated using Erlang C formula. Traffic congestion is minimized by adopting this approach. Wastage of human resource in terms of time is also reduced. System productivity is increased due to minimum transport time. Hence power and energy usage is decreased.

References

1. Moreno, M.T.: Inter-Vehicle communications: assessing information dissemination under safety constraints. In: 4th Annual IEEE/IFIP Conference on Wireless on Demand Network Systems and Services (WONS), Jan 2007
2. Marfia, G., Pau, G., Sena, E.D., Giordano, E., Gerla, M.: Evaluating Vehicle Network Strategies for Downtown Portland: Opportunistic Infrastructure and the Importance of Realistic Mobility Models. MobiOpp07, San Juan (2007)
3. Dombush, S., Joshi, A.: Street Smart Traffic: Discovering and Disseminating Automobile Congestion Using VANET'. In: Vehicular Technology Conference, April 22, 2007
4. Lochert, C., Scheuermann, B., Caliskan, M., Mauve, M.: The feasibility of information dissemination in vehicular ad-hoc networks. In: 4th Annual Conference on Wireless on Demand Network Systems and Services, pp. 92–99. Obergurgl, Austria, (2007)
5. Somasree, B., Sutirtha Kumar, G., Anirban, K.: Erlang based traffic automation using multi-agent. Int. J. Adv. Comput. Technol. (IJACT) 5(12) (2013)
6. Panita, P., Poj, T., Kanokchai, N.: Evaluation of road traffic congestion using fuzzy techniques. In: IEEE TENCON 2011, pp. 1–4, Nov 2011
7. Lu, J., Cao, L.: Congestion evaluation from traffic flow information based on fuzzy logic. IEEE Intell. Transp. Syst. 1(2003), 50–53 (2003)
8. Krause, B., von Altrock, C.: Intelligent high way by fuzzy logic: Congestion detection and traffic control on multi-lane roads with variable road signs. In: 5th International Conference on Fuzzy Systems, vol. 3, pp. 1832–1837, Sept 1996

9. Porikli, F., Li, X.: Traffic congestion estimation using HMM models without vehicle tracking. *IEEE Intelligent Vehicles Symposium*, pp. 188–193, June 2004
10. Hoßfeld, T., Leibnitz, K., Remiche, M.-A.: Modeling of an online TV Recording Service. In: *ACM SIGMETRICS Performance Evaluation Review, Special Issue on the MAMA 2007 Workshop*, vol. 35, no. 2, pp. 15–17 Sept 2007
11. Jouini, O., Aksin, Z., Dallery, Y.: Queueing models for full-flexible multi-class call centres with real-time anticipated delays. *Int. J. Prod. Econ.* **120**, 389–399 (2009)
12. Shabayek, A.A., Yeung, W.W.: A simulation model for the Kwai Chung container terminals in Hong Kong. *Eur. J. Oper. Res.* **140**(1), 1–11 (2002)
13. Gyalistras, D., Gwerder, M., Oldewurtel, F., Jones, C.N., Morari, M., Lehmann, B., Wirth, K., Stauch, V.: Analysis of energy savings potentials for integrated room automation. *Clima—RHEVA World Congress*, Antalya, Turkey (2010)
14. Imaiya, P., Purnima, K., Gayathri, N., Sneha, D.: Self-Automated efficient energy saver in domestic environments. In: 2012 international conference on industrial and intelligent information (ICII 2012), *IPCSIT*, vol. 31, Singapore (2012)
15. Daniel, K., Tobias, T., Marek, K., Tim, N.: Development of a general virtualization approach for energy efficient building automation. *Int. J. Energy Eng. (IJEE)* **3**, 1–6 (2013)
16. Costanzo, G.T., Kheir, J., Guchuan, Z.: Peak-load shaving in smart house via online scheduling. In: *International Symposium on Industrial Electronics (ISIE)*, Gdansk, Poland, pp. 1347–1352, June 2011
17. Somasree, B., Anirban K., Sutirtha Kumar G.: Services in green computing. In: *1st International Science & Technology Congress*, pp. 357–361. Elsevier Publication, Kolkata, India, 28–29 Aug 2014
18. Alcantarilla, P.F., Sotelo, M.A.: Automatic daytime road traffic control and monitoring system. In: *11th International IEEE Conference on Intelligent Transportation Systems*, pp. 944–949. Beijing, China, 12–15 Oct 2008
19. Krause, B., von Altrock, C., Pozybill, M.: Intelligent Highway by Fuzzy Logic: Congestion Detection and Traffic Control on Multi-Lane Roads with Variable Road Signs. *EUFIT'96*. Aachen, Germany (1996)
20. Cherrett, T., Waterson, B., McDonald, M.: Remote automatic incident detection using inductive loops. *Inst. Civil Eng. Transport* **158**(3), 149–155 (2005)
21. Beymer, D., McLauchlan, P., Coifman, B., Malik, J.: A real-time computer vision system for measuring traffic parameters. In: *Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 495–501. San Juan, Puerto Rico, 17–19 June 1997
22. Sadek, S., Al-Hamadi, A., Michaelis, B., Sayed, U.: A statistical framework for real-time traffic accident recognition. *J. Signal Inform. Process.* **1**(1) 2010
23. Shaikh, F., Chandak, M.B.: An approach towards traffic management system using density calculation and emergency vehicle alert. *IOSR J. Comput. Sc. (IOSR-JCE)* **4**, 24–27 (2014)
24. Lakshmi Pallavi, P., Hussain, M.A.: Intelligent traffic control system using embedded web technology. *IRACST—Int. J. Comput. Sc. Inform. Technol. Secur. (IJCSITS)* **4**(2), 46–49 (2014)
25. Anand, J., Arul Flora, T.G.: Emergency traffic management for ambulance using wireless communication. *IPASJ Int. J. Electron. Commun. (IJEC)* **2**(7), 1–4 (2014)
26. Hadi, R.A., Sulong, G., Edwar George, L.: Vehicle detection and tracking techniques: a concise review. *Signal Image Process.: An Int. J. (SIPIJ)* **5**(1), 1–12 (2014)
27. Sharma, B., Katiyar, V.K., Gupta, A.K., Singh, A.: The automated vehicle detection of highway traffic images by differential morphological profile. *J. Transp. Technol.* **4**, 150–156 (2014)
28. http://en.wikipedia.org/wiki/Erlang_B

A Taxonomy of Software's Non-functional Requirements

Nida Afreen, Asma Khatoon and Mohd. Sadiq

Abstract Software requirements are divided into two parts, FRs and NFRs. FRs determine the functionality, while, NFRs determine how a system is supposed to be. In the literature, we have identified that most of the work is related to FRs. NFRs have received less attention by goal-oriented requirements engineering community. The aim of this paper is to present taxonomy of non-functional requirements so that the requirements analyst can easily identify different types of NFRs according to their needs in the early phase of requirements engineering.

Keywords Non-functional requirements · Types of non-functional requirements · Classification

1 Introduction

Non-functional requirements (NFRs) are accepted as exceptionally imperative features in the accomplishment of software development [1–3]. NFRs are significant as they deal with the critical concern of software quality [4]. In spite of their significance, NFRs have received less attention by the requirements engineering community [5]. In the software architecture literature, NFRs are considered as quality attribute. The IEEE standard [6] defines quality as “amount to which software possesses following attributes like “trustworthiness”, “interoperability” etc.”. Glinz [7] defined NFR as: (i) “a property, or quality that the product must

N. Afreen (✉) · A. Khatoon

Department of Computer Science and Engineering, Faculty of Engineering and Technology, Al-Falah University, Dhauj, Faridabad, Haryana, India
e-mail: nida.afreen10@gmail.com

Mohd. Sadiq

Computer Engineering Section, UPFET, Jamia Millia Islamia (A Central University), New Delhi 110025, India
e-mail: sadiq.jmi@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_6

47

have following attributes such as (a) appearance, (b) speed (d) accuracy” and (ii) “the behavioural aspect of a system which have following quality functions (a) performance, and (b) usability etc.”

Paying no attention to NFRs is a severe crisis. For example, the London ambulance system failed because of the lack of NFRs [8, 9]. Classification of NFRs plays an important role to present the detailed picture of different types of NFRs in front of requirements analysts and customers. In the literature, we identify different classification schemes of NFRs. For example, Boehm [10] in 1978 defined software quality tree which tries to explain the key type of NFRs and relationship between them. They classify NFRs on the basis of general utility. General utility is further subdivided into portability, as-is-utility, and maintainability. In another classification scheme, proposed by Roman [11] in 1985, NFRs were divided into six sub-categories like (i) “interface requirements”, (ii) “performance requirements”, (iii) “operating requirements”, (iv) “lifecycle requirements”, (v) “economic requirements”, and (vi) “political requirements”. In 1987, Grady [12] presents software quality attributes as FURPS model, i.e. (i) “functionality”, (ii) “usability”, (iii) “reliability”, (iv) “performance”, and (v) “supportability”. ISO/IEC 9126 [13] also presents the classification schemes on the basis of different quality levels, i.e. (i) “quality in use”, (ii) “external quality”, (iii) “internal quality” and (iv) “process quality”. On the basis of our literature review, we identify that these classification schemes do not classify NFRs “on the basis of commonly used NFRs, on the basis of definition and attributes of NFRs, and on the basis of conflicted NFRs”. Therefore, it motivates us to present the taxonomy for NFRs on the basis of the above criteria.

This paper is organized as follows: Section 2 presents the proposed classification scheme of NFRs. In Sect. 3, we explain how the proposed classification is helpful for requirements analysts at the time of requirements elicitation process. Conclusions and the future work are given in Sect. 4.

2 Proposed Classification

In this section, we present our classification scheme of NFRs. In our scheme, we classify NFRs into three parts: classification on the basis of commonly used NFRs, on the basis of meaning and attributes of NFRs, and on the basis of conflicted NFRs (see Fig. 1).

(A) Commonly used NFRs

In this criteria, we have considered those NFRs that are useful for the successful development of different types of systems like information system, web based system, etc. On the basis of our literature review, we identified the following NFRs employed in different systems [4] like “performance”, “reliability”, “usability”, “security”, and “maintainability”.

Brief descriptions of these NFRs are given below:

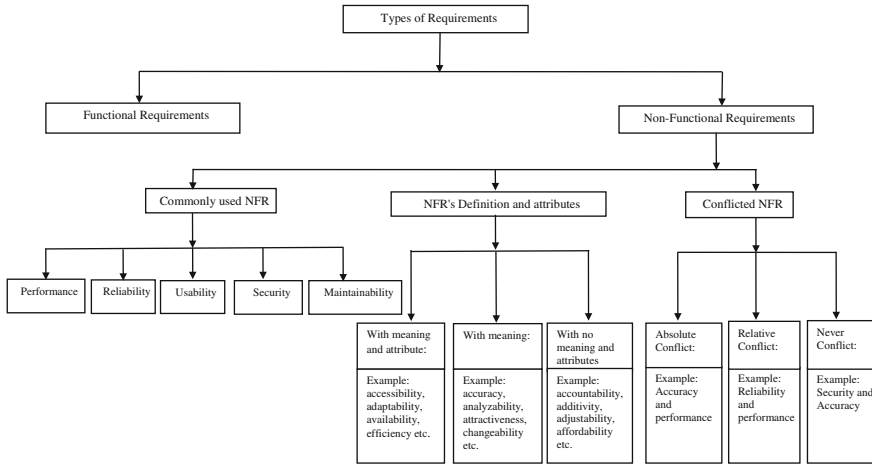


Fig. 1 Classification of non-functional requirements

Performance requirement describe the ability of software artefact “to make suitable performance comparative to the quantity of resources required to implement complete functionality in fixed condition [14]”. Some of the following attributes of performance requirement are: (i) “response time”, (ii) “capacity”, (iii) “latency”, (iv) “throughput”, (v) “computation”, (vi) “execution speed”, (vii) “transit delay”, (viii) “workload”, (ix) “resource utilisation”, (x) “memory usage”, and (xi) “accuracy” [4, 14].

Reliability requirements define the capacity of software artifact to operate with no failure and retain a specified phase of performance when used in particular normal circumstances during a certain time period [4]. Some of the attributes of reliability requirements are : (i) “completeness”, (ii) “accuracy”, (iii) “consistency”, (iv) “availability”, (v) “integrity”, (vi) “correctness”, (vii) “maturity”, (viii) “fault tolerance”, (ix) “recoverability”, (x) “reliability compliance”, and (xi) “failure rate critical” [15,16].

Usability requirements specify ability of the user communication with the system; also the attempt required to study, operate, form input, and understand the output of the system [4]. Some of the attributes of usability requirements are: (i) “learnability”, (ii) “understandability”, (iii) “operability”, (iv) “attractiveness”, (v) “usability compliance”, (vi) “ease of use”, (vii) “human engineering”, (viii) “user friendliness”, (ix) “memorability”, (x) “efficiency”, (xi) “user productivity”, (xii) “usefulness”, (xiii) “likeability”, and (xiv) “user reaction time” [4, 14].

Security requirements specify the concern to avoid illegal access to the system, program, and information [4]. It includes the following attributes like: (i) “confidentiality”, (ii) “integrity”, (iii) “availability”, (iv) “access control”, and (v) “authentication” [4, 14].

Maintainability requirements explain the ability of the software artifact to be tailored that may contain correcting a defect or change in the software [4]. Some of the

following attributes of maintainability requirements are: (i) “testability”, (ii) “understandability”, (iii) “modifiability”, (iv) “analyzability”, (v) “changeability”, (vi) “stability”, and (vii) “maintainability compliance” [14].

(B) NFRs on the basis of Definition and Attributes

In this section, we classify NFR on the basis of definition and attribute. The objective of these criteria is to identify those requirements that are clearly defined in the literature, like performance, security, usability, etc., and also identify those requirements that are supported by their attributes like accessibility, adaptability, availability, efficiency, etc. Therefore, as a result, we have identified that there are 252 types of NFRs but only 114 types of NFRs have been clearly defined in [4, 14]. In our work, we classify NFRs on the basis of the following: (i) With meaning and attributes, (ii) With meaning, (iii) With no meaning and attributes. A brief description of these classifications is given below:

With meaning and Attributes

These types of NFRs specify that each NFR has meaning and the attributes associated with them [4]. In these criteria only 23 types of NFRs have definitions, for example accessibility, adaptability, availability, efficiency, etc. *Accessibility* defines the quality of being at hand when needed. *Adaptability* defines ability to change (or be changed) to fit changed circumstances. *Availability* defines the quality of being in hand when needed. *Efficiency* defines the ratio of the output to the input of any system.

With Meaning

In this criterion, 30 types of NFRs have meaning but have no attributes, for example accuracy, analyzability, attractiveness, affordability, etc. [4, 14]. A brief description of these NFRs is given below: *Accuracy* defines the ability of a measurement to match the actual value of the quantity being measured. *Analyzability* defines to examine carefully and in detail so as to identify causes, key factors, possible result, etc. *Attractiveness* refers to a quality that causes an interest or desire in something or someone. *Changeability* defines liable to change or to be changed.

With No Meaning and Attributes

In this criterion, we identify 63 types of NFRs, for which there is no meaning and attributes [4, 14]. For example, accountability, additivity, adjustability, affordability, etc. A brief description of the NFRs is given below: *Accountability* is the answerability to blameworthiness, liability, and the expectation of account-giving or the ability to determine the responsible source for activity. *Additivity* is the property to add in small amounts to something else to improve, strengthen, or otherwise alter it. *Adjustability* defines to settle or arrange; to free from differences or discrepancies. *Affordability* refers the ability of software to maintain its entire cost within the range of affordability of an organization or customer. It is essentially about the capability to ‘pay as you grow’ and it has many scope to it [15].”

(C) Conflicted NFRs

At the time of requirements elicitation process [16–25] there are some NFRs that may conflict with other NFRs [4, 14], for example, “accuracy” and “performance”

[4, 14]. Conflicted NFRs further classify into three parts, i.e. absolute conflicted NFRs, relative conflicted NFRs and NFRs that are never conflicted. NFRs that may conflict with other NFRs known as absolute NFRs; example of absolute NFR is “performance” due to its conflict with others it is the most conflicted NFR in the literature, and it is conflicted with “accuracy”, “availability”, “confidentiality” etc. The second type of conflicted NFRs is relative conflicts that are conflicted sometimes. This type of relative conflicted NFR are declared to be in conflict in some definite cases, but they are not conflicted in some other cases [4, 14], for example, “reliability” and “performance”. The third type of NFR is never conflicted NFR. These types of NFRs show that it will never conflict with any other NFRs, for example, “security” and “accuracy”.

3 Case Study

In this section, we have shown how to identify NFRs from the proposed classification. We considered an example of an “institute automation system (IAS)”. While developing a system, the first thing that comes to our mind is who the end-users are, and what are the FR and NFRs of these end-users? In this case study, we emphasize only on NFRs. The student will login the IAS by using their login ID and password. In order to verify that only the enrolled students are using IAS, there must be some security requirements in the system. So, the login ID is the security NFR. How the students access and use the IAS is another type of NFR, i.e. accessibility and usability. In our classification, we considered security NFRs as “Commonly used NFRs” (see Fig. 1). The faculty uses the Login Id to access the IAS; therefore, there is a need for security NFRs again. How much IAS is reliable is also an NFR. If salary is transferred to the account then there must be some privacy. Therefore, privacy is also another NFR. In the literature, we identified some NFRs that have meaning and attributes, with no meaning and attributes. This classification criterion will help requirements analysts to identify the NFR along with their attributes, with meaning and attributes, and with meaning, with no meaning and attributes. Suppose we want accuracy in our system, then this NFR has meaning and there is no attribute for this NFR. Therefore, at the time of decomposing and refining this NFR, requirements analysts will not further decompose it because it cannot be further divided into sub-NFRs.

It should be kept in mind while developing the system that conflicted NFRs should never come together, and it should be avoided (see Fig. 1). In our classification, we have another criterion for NFR, i.e. “Conflicted NFRs”. Confliction among NFRs may be of three types, i.e. absolute conflict, relative conflict, and never conflict. For example, a customer wants to develop a system that should be accurate; and at the same time their performance should be good. Therefore, these two NFRs, i.e. accuracy and performance are conflicted requirements. These two NFRs cannot be implemented simultaneously at the same time.

4 Conclusion

In the requirement elicitation process, NFRs are considered as the criteria for selection and prioritization of requirements. Despite its importance, it is least considered and understood by the requirements engineering community. In the literature, there is no understanding among the requirements engineering community about the classification of NFRs. In this paper, we have classified NFRs on the basis of commonly used NFRs, NFRs having definitions and attributes, and conflicted NFRs. Such type of classification would be useful to elicit the NFRs at the time of requirements elicitation process. In the future, we will try to work on the following issues:

1. To apply the proposed classification for the elicitation of NFR of Institute Examination System.
2. Study of softgoal interdependency graph (SIG) tool.
3. Analysis of NFRs using SIG.

References

1. Chung, L., Nixon, A., Yu, E., Mylopoulos, J. et al.: *Non-Functional Requirements in Software Engineering*. Kluwer Academic Publishers, Massachusetts (2000)
2. Ebert, C.: Putting requirements managements into praxis: dealing with non-functional requirements. *Inform. Softw. Technol.* **40**, 175–185 (1998)
3. Firesmith, D.: Using quality models to engineer quality requirements. *J. Object Technol.* **2**, 67–75 (2003)
4. Mairiza, D., Zowghi, D.: Constructing a catalogue of conflicts among non-functional requirements. In: Maciaszek, L.A., Loucopoulos, P. (eds.) *ENASE 2010, CCIS 230*, pp. 31–44, Springer-Verlag, Berlin Heidelberg (2011)
5. Chung, L., Nixon, B., Yu, E., Mylopoulos, J.: *Non-functional requirements in software engineering*. Kluwer Publishers, London (1999)
6. IEEE Standard 1061-1992: *Standard for a Software Quality Metrics Methodology*. Institute of Electrical and Electronics Engineers, New York (1992)
7. Glinz, M.: On non-functional requirements. In: *15th IEEE International Requirements Engineering Conference*, pp. 21–26 (2007)
8. Brietman, K.K., Leite, J.C., Finkelstein, S.P.: The World's Stage A Survey on Requirements Engineering Using a Real-Life Case Study. *Braz. Comput. Soc.* **6**, 13–38 (1999)
9. Finkelstein, A., Dowell, J.: A comedy of errors: the London service case study. In: *Eight International Workshop Software Specification and Design*, pp. 2–5 (1996)
10. Boehm, B.W., Brown, J.R., Kaspar, H., Lipow, M., MacLeod, G.J., Merritt, M.J.: *Characteristics of Software Quality*. North-Holland, Amsterdam (1978)
11. Roman, G.-C.: A taxonomy of current issues in requirements engineering. *IEEE Computer*, pp. 14–21 April 1985
12. Grady, R., Caswell, D.: *Software Metrics: Establishing a Company-wide Program*. Prentice-Hall, Englewood Cliffs (1987)
13. ISO/IEC 9126-1:2001(E): *Software Engineering—Product Quality—Part 1: Quality Model* (2001)

14. Mairiza, D., Zowghi, D., Nurmuliani, N.: An investigation into the notion on non-functional requirements. SAC10, pp. 22–26, Sierre Switzerland, March 2010
15. <http://www.advoss.com/software-quality-attributes.html>
16. Sadiq, M., Jain, S.K.: A fuzzy based approach for the selection in goal oriented requirements elicitation process. International Journal of System Assurance Engineering and Management. Springer (2014)
17. Sadiq, M., Jain, S.K.: An Insight into Requirements Engineering Processes. In: International Conference of Advance in Communication, Networking and Computing (CNC), LNCSIT, pp. 313–318. Springer, India (2012)
18. Sadiq, M., Jain, S.K.: Applying fuzzy preference relation for requirements prioritization in goal oriented requirements elicitation process. Int. J. Syst. Assur. Eng. Manage. Springer 5(4), 711–723 (2014)
19. Sadiq, M., Jain, S.K.: A fuzzy based approach for requirements prioritization in goal oriented requirements elicitation process. In: International Conference of Software Engineering and Knowledge Engineering (SEKE), pp. 54–58. USA (2013)
20. Sadiq, M., Jain, S.K.: Stakeholders identification methods in goal oriented requirements elicitation process. In: Fifth IEEE International Workshop on Requirements Prioritization and Communication at IEEE International Requirements Engineering Conference (RE), pp. 25–33, Karlskrona, Sweden
21. Shahnawaz, A., Sadiq, M.: Goal oriented mind map generation method for requirements elicitation process. In: IEEE International Conference on Futuristic Trends in Computational analysis and Knowledge management, (A-BLAZE), organized by Amity University, Greater Noida, India 2015 (accepted for publication)
22. Sadiq, M., Shahnawaz, A., Arif, M.: Intertwining of Recommender Systems with GOMMG method for Requirements Elicitation. In: International Conference on Innovative Trends in Computing Technology and Mathematica-2015, Organized by Delhi Institute of Technology and management, Sonapat, Haryana, India
23. Boehm, B., In, H.: Identifying Quality-Requirement Conflicts. IEEE Software, pp. 25–35 (1996)
24. Chung, L., Sampaio do Prado Liete, J.C.: On non-functional requirements in software engineering. In: Borgida, A.T. et al. (eds.) Mylopoulos Festschrif, LNCS 5600, pp. 363–379, 2009. Springer-Verlag, Berlin Heidelberg (2009)
25. Chung, L.: Representing and Using Non-Functional Requirements: A Process Oriented Approach. Ph.D. Thesis, Dept. of Comp. Science. University of Toronto, June 1993. Also tech. Rep. DKBS-TR-91-1

Segmentation of the Human Corpus Callosum Variability from T1 Weighted MRI of Brain

**Shayak Sadhu, Sudipta Roy, Siddharth Sadhukhan
and S.K. Bandyopadhyay**

Abstract Corpus Callosum is an important part of the brain which works as major neural pathway that connects homologous cortical areas of the two cerebral hemispheres. The size of Corpus Callosum is affected by age, sex, neurodegenerative diseases and various lateralized behaviour in people. Here T1 weighted Magnetic Resonance Imaging (MRI) of brain, usually the sagittal sections is taken which is then followed by the automated segmentation of the MRI slide. This segmentation has an important application in neurology as the shape as the thickness, size and orientation of Corpus Callosum depends on the various characteristics of the person. Lobar connectivity based percolations of the corpus callosum can be computed by our proposed method which is very accurate segmentation.

Keywords Corpus callosum · Automated segmentation · MRI of brain · Quantification · Sagittal region · Bending angle · Midpoint

S. Sadhu (✉) · S. Roy
Department of Computer Science and Engineering, Academy of Technology,
Adisaptagram, West Bengal, India
e-mail: shayakchemistry@gmail.com

S. Roy
e-mail: sudiptaroy01@yahoo.com

S. Sadhukhan
Master in Computer Application, Academy of Technology, Adisaptagram
West Bengal, India
e-mail: siddhartha.sadhukhan.2014@gmail.com

S.K. Bandyopadhyay
Department of Computer Science and Engineering, University of Calcutta,
92 A.P.C. Road, Kolkata 700009, India
e-mail: skb1@vsnl.com

1 Introduction

The human nervous system is spread throughout the human body. This system controls all the voluntary and involuntary actions in the body. It transmits signals to and from the different organs present in the body. The human brain is one of the complex structures present in the body which is responsible for the normal working of the body. If we consider the anatomy of the brain we see that the brain is majorly divided into two parts namely left and right lobes of cerebrum. These two parts are joined by Corpus Callosum (CC) through which two hemispheres communicate with each other. CC also known as the colossal commisure is a part of the brain located in mid sagittal region. It is the largest neural pathway which connects the two hemispheres of the brain. It consists of between 200 and 800 axon fibres beneath the cortex in the ethereal brain at the longitudinal fissure. In some cases CC might be partially or completely absent in a person. This condition is known as agenesis of the CC. The nature and function of CC has been a long interest to researchers as its alteration in structure has resulted in psychiatric and developmental disorders. Magnetic Resonance Imaging (MRI) of brain is a painless test where radio waves and magnetic field is used for getting a detailed picture of brain and its associated structure. The MRI scan does not use radiation which differentiates it from CAT scan, also called a CT scan or a computed axial tomography scan. The MRI scans rather than from giving us a clearer picture also gives us the advantage of easier identification of different abnormalities of brain. In neuro-imaging, the segmentation of different parts of brain is considered to play an important role in several medical applications. This field of study has attracted much interest from the medical community and gives us crucial information that might significantly impact clinical management and practice. In the process of manual segmentation of brain like using photo editing software's is still considered the most accurate, but more time-consuming method for segmentation which includes much laborious work including detection and then extracting those parts. It is traditionally time-consuming and dependent rather than an experience. The designs of algorithms for automatic segmentation of brain are being developed to ease this task of manual segmentation. But it is usually hard to segment some parts such as the ones located at sub-cortical level. In this work, the automated segmentation was achieved by image processing algorithms and the processed images are the more refined and then compared with the manually segmented image to get accuracy using different metric algorithms. Also rather than using standard algorithms we have used morphological operation on the images to get a more refined picture of the obtained segmentation. The proposed method includes improved detection and measurement of area of CC. The proposed method includes the following phases: (1) Input of T1 weighted image and refining of the image to reduce noise. (2) Segmentation of CC using area selection and binary conversion of image. (3) Colouring of detected part in original image. (4) Detection of mid-point

and end points of segmented CC. (5) Measurement of bending angle of CC. The verification of proposed method's accuracy, reliability, robustness, and multisite consistency has been done by making comparisons with manual segmented image using randomly selected slides from the database. This proposed method has been tested on 10 data sets selected in random and has been seen that automatic segmentation gives us the most approximate result with minimum error percentage.

The preference for automated segmentation is the increasing popularization in the neuro-anatomical research area. Some of the volumetric measurements which are acquired from MRI are used to examine and quantify the impact of some brain diseases and disorders on the human central nervous system [1, 2]. In the normalization approach using voxel-based morphometry (VBM) allows us to compare different brain MRI slides on a voxel-by-voxel basis [3]. In accordance to Marchetti et al. [4], it has been seen that the time spent on manual segmentation of the hippocampus might be taking the time of 75 min. per patient for each exam. This takes up a lot of time in producing the segmented image. In [5] it was seen that using T1 slides of MRI of brain, the CC was extracted from mid sagittal section of brain. Then the threshold was used in segmentation of CC. The goal of this work is to use image-processing and using statistical models of automated segmented MRI slides as ground truth image to produce accurate segmented image with respect to ground truth image. The normalization approach used voxel-based morphometry (VBM). The VBM technique allows us to compare dissimilar brain composition stand on voxel-by-voxel examination once deformation fields have been employed to spatially standardize the images [6]. The CC is a well-known white matter bundle which is simply identifiable on MR imaging which is linking between two cerebral hemispheres in a homotopic group with orientation to the cortex illustration [7]. Besides the CC plays significant role on instructive processes of inter-hemispheric messages and specialization [8], it is of enormous significant due to its limitation to ecological poisons, white matter diseases (such as multiple sclerosis) and schizophrenia [9, 10]. The size of the CC was also investigated by neuro-biologists working in many specialized fields, including handedness [11], musical ability [12], schizophrenia [13], autism [14] and Alzheimer's disease [15]. The rest of this article is organized in three more sections. Section 2 describes the process of automatic segmentation of brain tissue. Section 3 tells us about our practical experiments and presents the CC surface area using our proposed method, and finally in Sect. 4 we conclude our methods.

2 Proposed Method

In the event of segmentation of corpus callosum we use 10 datasets each containing around 12–15 mid sagittal MRI slices of the brain [16, 17]. The images are entered in a bundle so as to get a 3D representation of the obtained segmented corpus

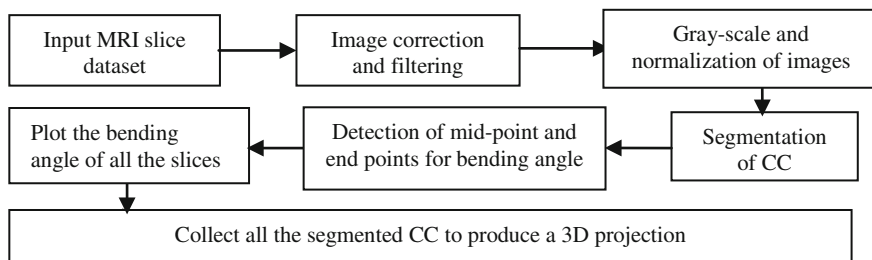


Fig. 1 Workflow of the proposed methodology

callosum. While considering each dataset we get individual result as parameters from each slice which is then plotted to analyse the result variation in each individual dataset. The work flow of our proposed methodology can be represented by Fig. 1.

2.1 Pre-processing of Slices

The T1 weighted MRI slices are chosen as input as soft tissues are clearly visible in T1 weighted MRI images. After taking the input the slices might contain many distorted element such as artefact, indistinct intensity values and uneven intensity distortion. In order to correct these defects in the slice the slice is processed through artefact removal [6] and image correction algorithm where intensity variation and other distortions are corrected. This pre-processing counts as an important step as these defects in the image gravely affects the segmentation part of our proposed methodology. The obtained image is usually a RGB image which can be represented as a 3D matrix where each 2D matrix represents Red, Green and Blue portion of the image. We convert it into gray-scale using the standard RGB to gray-scale conversion formula in which each pixel represent the intensity of images within the scale 0–255 which can be represented 8 bits of binary digit.

2.2 Segmentation of the Corpus Callosum (CC)

The segmentation of CC takes place through a series of steps. After the input and pre- processing step, we get a gray-scale image which can be represented as a 2D matrix where the image is made of intensity function done by variation of intensity in each pixel. Let us take that intensity function is X for the image I where each pixel can be given by $I[m, n]$ in which $1 \leq m \leq$ maximum number of rows (M) and

$1 \leq n \leq$ maximum number of columns (N). We need to find the threshold intensity of the grayscale image for its binary conversion. This is done by determining the threshold intensity and then putting all the pixels below the level as “0” and above it as “1”. For the conversion this calculation of threshold intensity has become an important step as a slight change might even put the region of CC as “0” which is unexpected. This is done by calculating the expectation intensity $E(I)$ and then taking the expectation value in our region of interest. This expectation value is calculated by considering the principles of discrete random variable when $E(I)$ can be given by: $(I) = I[m, n] \times P(I[m, n])$, where $P(X)$ is the probability of value X appearing in a corresponding pixel. This expectation gives us the most optimum threshold intensity to be selected and in this calculation the region of skull is not considered. Now after obtaining the threshold intensity the image is now to be sectioned to get a portion where probability of availability of CC is high. This probability is now calculated using feature extraction of CC which is present in our region of interest so that other region does not affect our extraction. Let us take rectangular distribution of probability of detection of CC in the image as it is confined in certain regions of the image. Let us divide the image into N equal divisions of rows and columns. So total number of divisions = $N * N = N^2$. Therefore the distribution can be given as:

$$P(I(m, n)) = f(x) = \frac{1}{b - a}, \text{ where } 1 \leq a, b \leq N^2$$

This probability is used to determine the region of interest. The image is then resized to our findings. After resizing, the extracted part is then converted into binary matrix in accordance to our previously found threshold intensity. After binary conversion of the slice, it is then detected for maximum connected component. For calculation of maximum connected component the whole slice is broken up into different components where each component is connected. This means that each component is a bunch “1” grouped together. Now each of the components is tested for its area. This can be given by the formula

$$A = \sum_{i=m}^M \sum_{j=n}^N I(i, j),$$

where the component I ranges from $[m:M, n:N]$. This summation gives us the effective area in pixel. Using this data we find out the maximum area present in the slice and extract only the component with maximum area. CC has been seen to be a connected component with maximum area which naturally segments out from slice, thus maximum area giving us only the CC in the slice. This image might contain noise as a result of previous operation on the image. This noise is corrected using morphological operation on this image. Thus we get our required segmented part.

After this the segmented part is brought back to the original size so that they can be compared with the original MRI slide to colour the detected part.

2.3 Detection of Mid-Point and Two End Points for Calculating Bending Angle

The segmented binary image after being brought back to original size is taken as input in this step. This image is then cropped to such a size such that it satisfies the following conditions:

- The image should contain only the segmented CC.
- At least one pixel value in leftmost, rightmost, topmost and bottommost extreme pixel should compulsorily contain "1".

The input binary image is brought to the specified condition as stated by running nested loops by detecting all the pixels in the image and finding the extreme positions of the segmented CC. Using this figure we find the topmost which is considered as mid-point and two extreme end point of CC by running nested loops. After obtaining the end-points we use this to calculate the bending angle. Let us assume that (a_1, b_1) is the position of the pixel considered as mid-point, (a_2, b_2) as the bottom leftmost position of the pixel and (a_3, b_3) as the bottom rightmost position of the pixel which is demonstrated in Fig. 2. Then the bending angle B can be calculated as

$$m_1 = \frac{b_2 - b_1}{a_2 - a_1}, \text{ and } m_2 = \frac{b_3 - b_1}{a_3 - a_1}$$

$$\text{deg1} = \tan^{-1}(m_1), \text{ and } \text{deg2} = \tan^{-1}(m_2)$$

$$B = \text{abs}(\text{deg2} - \text{deg1}),$$

where $\text{abs}(x)$ gives us the absolute value of x . This formula of derived from the concept of the equation of a line where the slope is calculated and inverse tangent if the slope gives us the angle of inclination. Using this angle of inclination we find the difference between the two slopes which gives us the bending angle. The bending angle obtained by the stated methodology gives us an error as we had

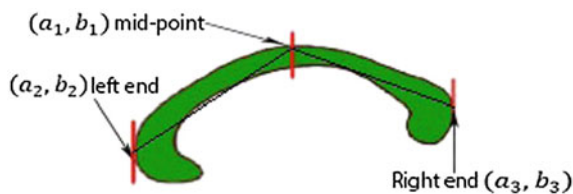
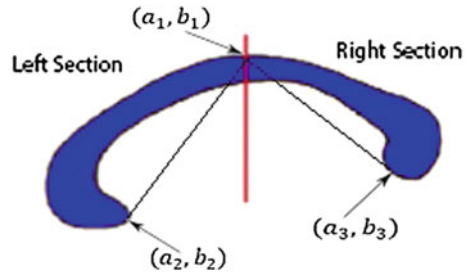


Fig. 2 Detection of mid-point and end point of CC

Fig. 3 Calculation of actual bending angle



aimed towards calculating bending angle between the two end points of CC. This is done by dividing the image into two parts and finding the bottom most points by applying reverse loops.

As illustrated in Fig. 3 this segmented part is divided into two parts namely left and the right section. Reverse loops are executed to find the positions of the tips from each sections. These reverse loops are useful in finding the tips using the reverse direction from the mid-point.

2.4 3D Projection of Corpus Callosum (CC)

After the segmentation and calculation of bending angle all the original size binary segmented image, the segmented image is collected. Each of the images is a 2D image with pixel containing values either “0” or “1”. The images containing the CC are only considered. The 3D figure is derived considering the features of stack data structure. This structure is useful in adding another dimension for a n -D structure into $(n + 1)$ -D structure. Let us consider an n -D structure. If this is pushed into the stack the total structure considering the stack becomes $(n + 1)$ -D structure as stack inserts length or size of the stack as another dimension. Thus we can get a 3D figure from a 2D image using the stated proposition. We know that in MRI scan the slices are taken in an interval of some distance in millimetre. So the slides in between two consecutive slides are not taken. So the absent slides are calculated and then stored. The number of absent slice = $\text{sep} \times 3.779527559$, where “ N ” is the number of absent slides and “sep” is the separation between two slides in mm. The nearest whole number is taken as the number of missing slides. The absent slides are calculated by modifying the size of an image generated by combining two consecutive slides. This modification is done by getting the values of linear interpolation of area of images. The variation is done with respect to the two areas of the consecutive slides. After getting all the sets of images the images are pushed to a top-down stack which gives us the correct orientation of the figure. This stack is then projected to give us the corresponding 3D figure of CC.

3 Results and Discussion

A number of theories have been proposed about detection and segmentation of CC such as using tensor imaging and watershed-based segmentation. But the proposed methodology uses a different approach to this problem by using statistics and feature extraction. The collected datasets each consisting sagittal plane MRI slides on which the proposed methodology has been tested obtaining different variations of output. Now as input of the algorithm we use T1 weighted MRI image as shown in Fig. 4a. Then the resizing of the image is done by the extraction algorithm to get the image as in Fig. 4b. Then the image is converted into binary image using threshold intensity as the reference intensity level as in Fig. 4c. Then extraction of maximum component executed on the resized figure to get the detected portion as in Fig. 4d. Then the extracted CC is then brought back to the original size which is the actual output as in Fig. 4e. Then this obtained output is compared with original MRI slide to get colouring on the output image showing detected CC.

The accuracy measures used to evaluate the performance [18, 19] of the proposed methods are the Relative area Error (RE) [18, 19], Kappa Index (Ki), Jacard Index (Ji), correct detection ratio (Cd) and false detection ration (Fd) has been described below. An important difficulty we have to face in developing segmentation methods is the requirement of a good benchmark for their evaluation. Although physical or digital phantasm can afford a level of known “reference or ground truth image”, they are still not capable to reproduce the full range of imaging characteristics, normal and abnormal anatomical inconsistency observed in medical data. Let AV and MV denote [17, 20] the region of by design and by hand segmented region, and TP is the joint section between AV and MV. The FP can be calculated by subtracting TP from AV and FN can be calculated by subtracting TP from MV, FP and FN denote to the “true positive” and “false negative” respectively. The kappa index is determine by two multiplied by intersection of MV and AV divided by sum of AV and MV. Jacard index can be found from intersection of MV and AV divided by sum of TP, FN and FP. This metric is more susceptible to differences since both denominator and numerator varies with rising or falling overlap. Correct detection ratio or sensitivity is defined by the intersection of AV and MV divided by MV. The Relative Error [21] (RE) can be calculated by AV subtracted by MV divided by MV (Table 1).



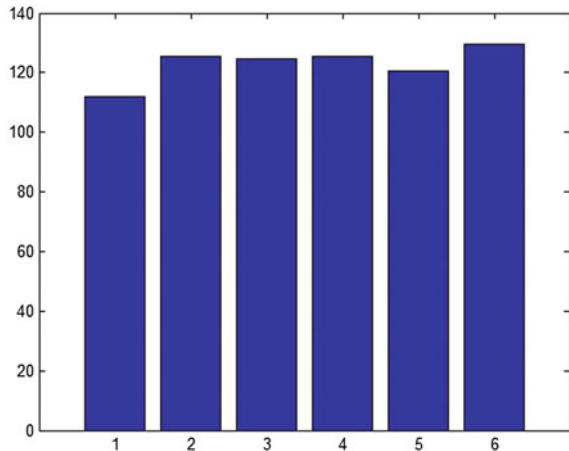
Fig. 4 a Original image, b resized image, c resized binary image, d detected CC, e expanding to original form, f actual result

Table 1 Quantification and accuracy estimation

Image name	AV	MV	RE in %	TP	FP	FN	AV +MV	Kappa index (%)	(TP +FN +FP)	Jacard index (%)
068	868	850	2.11	845	23	05	1738	97.23	873	96.79
069	920	910	1.09	901	19	09	1830	98.46	929	96.98
070	853	877	2.50	850	03	27	1730	98.26	880	96.59
071	920	893	2.93	887	33	06	1813	97.84	926	95.78
072	859	841	2.14	834	25	07	1700	98.11	866	96.30
073	868	853	1.75	848	20	05	1721	98.54	873	97.13
074	1570	1592	1.38	1563	17	29	3162	98.86	1609	97.14
075	920	908	1.32	901	19	07	1828	98.57	927	97.19
076	2418	2453	1.42	2406	12	47	2871	97.94	2465	97.60
077	2723	2749	0.945	2715	08	34	5472	99.23	2757	98.47

The above table illustrates the result of 10 T1 weighted MRI images taken as input from image database of a same person. In this table we have observed that the relative area error RE is having a mean of 1.7585 which is relatively low. The RE increases if it is difficult to distinguish between CC and its surrounding tissues. This possesses a threat to the given method as AV might differ greatly compared to MV as in image 071 with RE = 2.93 which is relatively high compared to other images. After the testing of segmentation of CC we get a set of binary images which is tested for bending angle as given by the plot in Fig. 5.

In Fig. 5 we have shown the variation of the bending angle in a dataset with 6 slides where CC has been correctly detected. As we see in the figure the variation observed is $\pm 0.5^\circ - 20^\circ$ in the taken Fig. 5 and generally stays within this range for all the tested datasets. The calculated angle gives us the correct variation. In the event if 3D representation of CC it have been observed that there is a little variation

Fig. 5 Plot of the bending angle of 6 CC detected slides

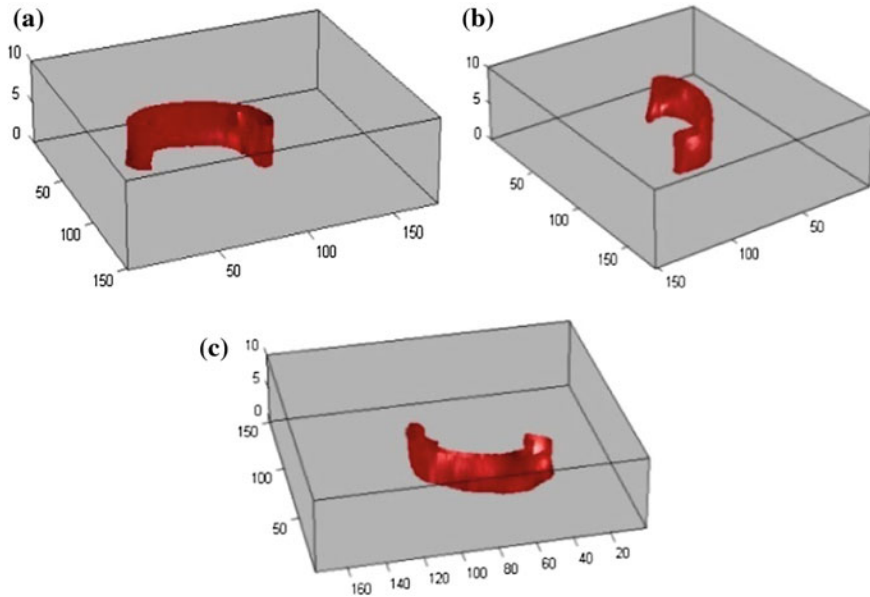


Fig. 6 Different views of the 3D figure

to consecutive slices, so the figure formed (Fig. 5) appears as a solid figure with smooth surface.

Figure 6a–c gives us the different views of the CC viewed at different angles. Using this figure we can view the actual shape of CC present in the brain. This 3D projection is observed by projecting the 3D stack of 2D images. The cracks and fissures present in the figure can be observed clearly by varying the output image.

4 Conclusion

The proposed algorithm usually works on MRI slides taken from mid sagittal section of brain. In this algorithm it is sometimes difficult to identify CC in some slides. To remove artefact we need morphological operation on binary images for correction of the result. This proposed methodology gives us the most suitable result as per T1 weighted MRI images are involved. In this proposed methodology we just highlight the detected portion of CC and keeping rest of the image untouched. This ensures us to detect other tissues in the slide with respect to CC. In the above proposed algorithm we see that the results obtained so far is promising but we need to improve on the removal of surrounding tissues which sometimes interfere with detection and extraction part of the algorithm.

References

1. Chaim, T.M., Duran, F.L.S., Uchida, R.R., Périco, C.A.M., de Castro Geraldo C.C.: Volumetric reduction of the corpus callosum in Alzheimer's disease in vivo as assessed with voxel-based morphometry. *Psychiatry Research, Neuroimaging* (2006)
2. Bakshi, R., Dandamudi, V.S.R., Neema, M., De Bermel, C.R.A.: Measurement of brain and spinal cord atrophy by magnetic resonance imaging as a tool to monitor multiple sclerosis. *J. Neuroimag.* **15.4**, 30 (2005)
3. Wright, I.C., McGuire, P.K., Poline, J.B.: A voxel-based method for statistical analysis of gray and whiter matter density applied to schizophrenia. *Neuroimage* **2**, 244–252 (1995)
4. Marchetti, R.L., de Campos Bottino C.M., NagahishiSuely, M., Campi, C.C.: Confiabilidade de Medidas Volumétricas de Estruturas Temporais Mesiais. *Arquivos de Neuro-Psiquiatria*, v. 60, n.2B (2002)
5. Nigro, S., Cerasa1, A., Zito, G., Perrotta, P., Chiaravalloti, F., Donzuso, G., Fera, F., Bilotta, E., Pantano, P., Quattrone, A.: The Alzheimer's Disease Neuroimaging Initiative: Fully Automated Segmentation of the Pons and Midbrain using Human T1 MR Brain Images, vol. 9, Issue 1, Jan 2014
6. Wright, I.C., McGuire, P.K., Poline, J.B.: A voxel-based method for statistical analysis of gray and whitematter density applied to schizophrenia. *Neuroimage* **2**, 244–252 (1995)
7. Pandya, D.N., Seltzer, B.: The topography of commissural fibers two hemispheres-one brain: functions of the corpus callosum, pp. 47–74. Alan R Liss Inc, New York (1986)
8. Trevarthen, C.: Integrative functions of the cerebral commissures. In: *Handbook of Neuropsychology*, 4, pp. 49–83. Elsevier Scientific Publishers, Amsterdam (1990)
9. Simon, J.H., Jabocs, L.D., Campion, M.K., Rudick, R.A., Cookfair, D.L., Herndon, R.M., Richert, J.R., Salazar, A.M., Fischer, J.S., Goodkin, D.E., Simonian, N., Lajaunie, M., Miller, D.E., Wende, K., Martens-Davidson, A., Kinkel, R.P., Munschauer, F.E.: A longitudinal study of brain atrophy in relapsing multiple sclerosis. *Am. Acad. Neurol.* **53**(1), 139–148 (1999)
10. Barkhof, F.J., Elton, M., Lindeboom, J., Tas, M.W., Schmidt, W.F., Hommes, O.R., Polman, C.H., Kok, A., Valk, J.: Functional correlates of callosal atrophy in relapsing remitting multiple sclerosis patients: a preliminary MRI study. *J. Neurol.* **245**, 153–158 (1998)
11. Habib, M., Gayraud, D., Oliva, A., Regis, J., Salamon, G., Khalil, R.: Effects of handedness and sex on the morphology of the corpus callosum: a study with brain magnetic resonance imaging. *Brain Cogn.* **16**, 41–61 (1991)
12. Lee, D.J., Chen, Y., Schlaug, G.: Corpus callosum: musician and gender effects. *Neuro Report* **14**, 205–209 (2001)
13. Woodruff, P.W.R., Manus, I.C., David, A.S.: Meta analysis of corpus callosum size in schizophrenia. *J. Neurol. Neurosurg. Psychiatry* **58**, 457–461 (1995)
14. Egaas, B., Courchesne, E., Saitoh, O.: Reduced size of corpus callosum in autism. *Arch. Neurol.* **52**, 794–801 (1995)
15. Salat, D., Ward, A., Kaye, J.A., Janawsky, J.S.: Sex differences in the corpus callosum with aging. *Neurobiology* **18**, 191–197 (1997)
16. <http://www.med.harvard.edu/aanlib/home.html>. May 2013
17. <http://brainweb.bic.mni.mcgill.ca/brainweb/>. May 2013
18. Roy, S., Bandyopadhyay, S.K.: Abnormal regions detection and quantification with accuracy estimation from MRI of Brain, pp. 611–615. In: *Proceedings 2nd IMSNA IEEE*, (2013)
19. Roy, S., Chatterjee, K., Bandyopadhyay, S.K.: Segmentation of acute brain stroke from MRI of brain image using power law transformation with accuracy estimation. In: *ICACNI-2014, Kolkata, Proceedings*, vol. 27, pp. 453–461. Springer (2014)
20. Roy, S., Chatterjee, K., Maitra, I.K., Bandyopadhyay, S.K.: Artefact Removal from MRI of Brain Image. *Int. Ref. J. Eng. Sci. (IRJES)*, **2**(3), 24–30 (2013)
21. Roy, S., Ghosh, P., Bandyopadhyay, S.K.: Segmentation and contour extraction of cerebral hemorrhage from MRI of brain by gamma transformation approach. In: *Proceedings Springer, AISC 328, FICTA, Bhubaneswar, Orissa* (2014)

Analysis of Organic Molecular Single-Electron Transistor Using $C_4H_6B_2$ with Different Metal Electrodes

E. Meher Abhinav, M. Chandra Mohan, A. Suresh Reddy, Vemana Chary and Maragani Thirupathi

Abstract An organic molecule-based single-electron transistor (SET) is analysed by ab initio method using Density Functional Theory (DFT). Initially, benzene molecule is taken; two carbon atoms from benzene are replaced by boron atoms, and the structure of the molecule is optimized. The optimized structure $C_4H_6B_2$ is kept above the gate dielectric in the island for weak coupling. The charge energies of device are calculated in both isolated and SET environment. We have done analysis by using different electrodes with gold (work function = 5.28 eV), osmium (work function = 5.93 eV) and caesium (work function = 2.14 eV) in SET environment. By charge stability diagrams, the conductance dependence of SET on gate voltage and bias potential are verified.

Keywords Single-electron transistor (SET) · $C_4H_6B_2$ · Ab initio · Density functional theory (DFT) · Non-equilibrium greens function (NEGF) · Coulomb blockage

1 Introduction

The number of transistors integrated on chip doubles approximately in every two years. Scaling cannot last for ever because below gate length of 10 nm many parasitics and second-order effects create problem. For continuation of Moore's law, different methods came to existence like spintronics and tunnel junctions. When transistors enter the nanometer regime, they suffer from many second-order effects and leakage current issues. Limitations of transistor scalings such as drain leakage have reached peak level, and hence different approaches depending on

E.M. Abhinav (✉) · M.C. Mohan · A.S. Reddy · V. Chary · M. Thirupathi
Department of Electronic and Communications, Malla Reddy College of Engineering,
Dhulapally, Hyderabad 500014, Telangana, India
e-mail: abhi4abhi09@gmail.com

© Springer India 2016
S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_8

electron spin transfer, tunnel junction, and confinement of channel material are used.

Enormous attention is given by industries and academic researchers on active electronic unit in single molecules. Single molecules have several unique properties which can be used in electronic unit. Fabricating such small scale devices will be advantageous because of self-assembly and their diversity and functionality.

For smaller and faster switches, single-electron transistors (SETs) have become an alternative. SET allows more number of transistors to be integrated on chip. In present years, the main focus is on using organic molecules as island in SET [1–5].

2 Single-Electron Transistor

SET involves adding a gate control to a coulomb blockade structure as shown in Fig. 1. SET exploits quantum effect of tunnelling to control and measure the movement of single electron. In SET, charge passes through the island in quantized manner. For electron to tunnel from source to island the energy must be equal to $e^2/2c$ (coulomb energy) as shown in Fig. 2, where C is self-capacitance of island.

$$C = e^2 \Delta E, \quad (1)$$

where ΔE is the separation between the energy levels of island.

If the energy is below coulomb energy, electron does not hop and current does not flow.

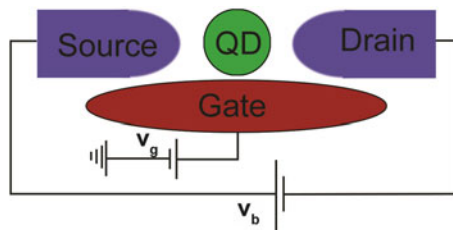
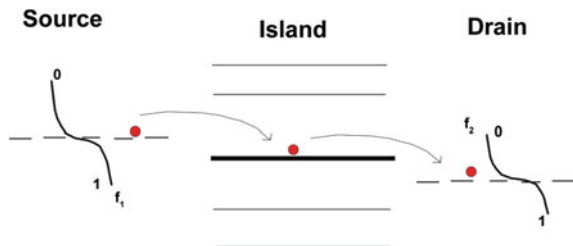


Fig. 1 Single-electron transistor

Fig. 2 Electron tunnelling



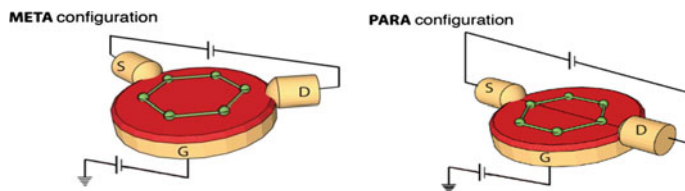


Fig. 3 Meta configuration and para configuration

When the bias voltage between drain and source is increased, the electron is passed through island when energy reaches coulomb energy.

In Meta configuration contact atoms are next nearest neighbours, and in Para configuration the atoms on opposite ends of molecule are coupled to leads, as shown in Fig. 3 [6–10].

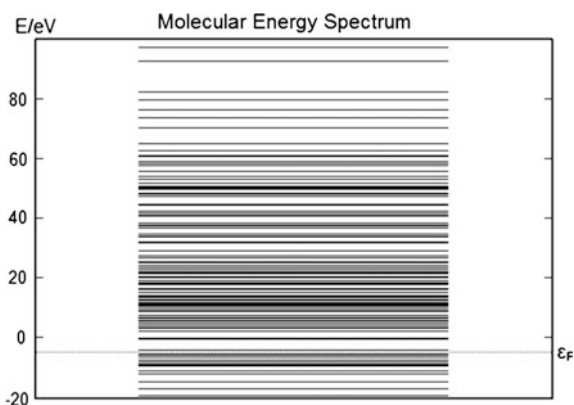
3 Theoretical Approaches

There are different experimental techniques and many theoretical approaches to describe transport in molecular devices. Combination of Density Functional Theory (DFT) with non-equilibrium greens function (NEGF) method is a standard way to study transport in nano-scale. The molecular energy spectrum consists of molecular energy levels as shown in Fig. 4, where Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO).

Basically Fermi-level can be seen as the energy level that lies between occupied and unoccupied states. It can be anywhere in HOMO-LUMO gap.

We define the ionization energy (E^I) as the energy required for removal of single electron from the molecule. It is equal to $-E_{\text{HOMO}}$.

Fig. 4 Molecular energy spectrum of $\text{C}_4\text{H}_6\text{B}_2$ with para configuration



The affinity energy (E^A) is defined to be the energy gained by adding a single electron to the molecule and it is equal to $-E_{\text{LUMO}}$.

In weak coupling regime, the transport is described by sequential tunnelling.

For movement of electron from source to island, island must have lower energy compared to source.

$$E^{\text{source}}(M) + E^{\text{Island}}(N) \geq E^{\text{source}}(M - 1) + E^{\text{Island}}(N + 1). \quad (2)$$

For movement of electron from island to drain, the drain must have lower energy compared to island.

$$E^{\text{drain}}(K) + E^{\text{Island}}(N + 1) \geq E^{\text{drain}}(K + 1) + E^{\text{Island}}(N). \quad (3)$$

Charging energy of island is

$$\Delta E^{\text{Island}}(N) = E^{\text{Island}}(N + 1) - E^{\text{Island}}(N). \quad (4)$$

Form the Eqs. (2, 3 and 4), the following variables are given as follows:

- N is the number of initial electrons on island;
- M is the number of initial electrons on source electrode;
- K is the number of initial electrons on drain electrode
- $E^{\text{Island}}(N)$ It gives energy of island as function of number of electrons on island;
- $E^{\text{source}}(N)$ It gives energy of source as function of number of electrons on source;
- and
- $E^{\text{drain}}(N)$ It gives energy of drain as function of number of electrons on drain.

4 Computational Procedure

By using electrostatic gate, the charging energy of the molecule is modified. Tuning the gate voltage, the energy levels can be moved inside and outside of bias window. Dependency of SET conductance on gate voltage and bias potential is emphasized by charge stability diagram. We also analysed total energy as function of gate voltage in SET environment.

We analysed the total energy by changing the electrodes (work function of gold is 5.28 eV, osmium is 5.93 eV and caesium is 2.14 eV) and plotted the graphs. Total energies and charging energies of molecule for various charge states (2, 1, 0, -1, -2) are calculated using ATK 12.8 Tool kit. These are based on non-equilibrium greens function and density functional theory (DFT).

$\text{C}_4\text{H}_6\text{B}_2$ molecule is used as an island in SET placed between metallic drain and source electrodes above dielectric slab with Para configuration. The dielectric slab of dielectric constant $10\epsilon_0$ and 3.8 Å thickness lies on metallic back gate. In order to

keep the perpendicular component of electric field zero at boundaries, we used Newman boundary condition. Energy zero is kept at absolute zero [11–17].

5 Results and Discussion

Initially, benzene molecule is taken and the charging energies and total energy are calculated. Figure 5 shows the charge stability diagram of benzene. Then the benzene molecule is taken and two carbon atoms are replaced by boron, and structure is optimized and placed in SET environment and even top view of set environment is seen in Fig. 6. Charging energy of $C_4H_6B_2$ in Para configuration is

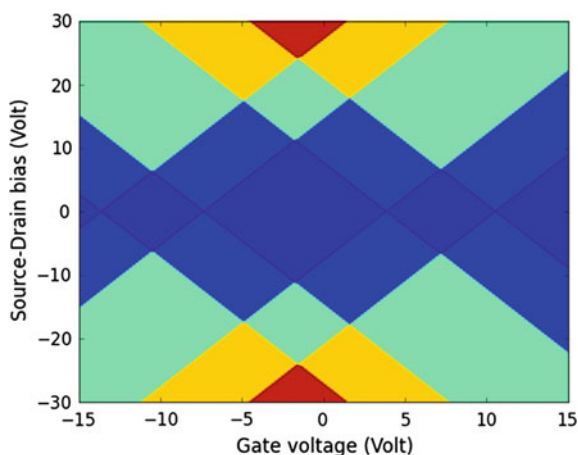


Fig. 5 Charge stability diagram of benzene, the colours represent: (0) blue, (1) light blue, (2) green, (3) orange and (4) red (Color figure online)

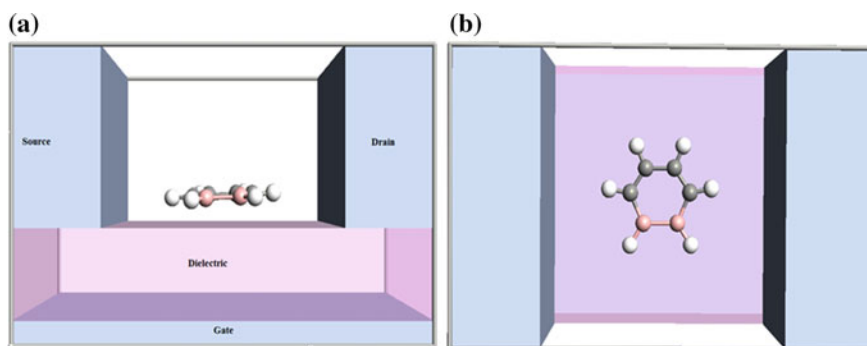


Fig. 6 **a** $C_4H_6B_2$ molecule in SET Environment. **b** Top view of $C_4H_6B_2$ molecule in SET

computed by the total energies of island for different charge states of molecule (-2 , -1 , 0 , 1 , 2) in both SET and isolated phases.

The Total Energy of Benzene molecule we obtained is -1040 eV.

The Total Energy of $C_4H_6B_2$ molecule is -880.41 eV.

As the Total Energy of $C_4H_6B_2$ is negative and more compared to Benzene, the coupling of $C_4H_6B_2$ molecule with the source and drain will be weak. Due to weak coupling of $C_4H_6B_2$ molecule with the source and drain, when the electron hops from source to island the electron gets more time to localize. So electron loses all information about its original state in the source electrode and it operates in incoherent transport regime.

Highest Occupied Molecular Orbital (HOMO), $E_{\text{HOMO}} = -5.94$ eV.

Lowest Unoccupied Molecular Orbital (LUMO) $E_{\text{LUMO}} = -0.747$ eV.

Table 1 defines the charging energies of $C_4H_6B_2$ molecule in both isolated and SET environments with different charge states. We obtain $E^1 - E^A = -10.49$ eV in the isolated phase and $E^1 - E^A = -10.74$ eV in SET environment.

The affinity energy (E^A) and the ionization energy (E^I) of $C_4H_6B_2$ are 1.001 eV and -9.05 eV in isolated phase and E_A and E_I of $C_4H_6B_2$ is obtained as 3.46 eV and -9.05 eV in SET environment, respectively. In isolated condition, charging energy of $C_4H_6B_2$ at zero state has highest charging energy of 1.001 eV and at -2 charge state it has lowest charging energy of -24.1 eV. In SET environment, at -1 state it has highest charging energy, where at negative potential the positive charges are stabilized and at -2 charge state it has highest charging energy of -17.14 eV. In SET environment, positive charges are stabilized.

We calculated total energy of different charge states of isolated and SET environment as function of gate potential. From Figs. 7 and 8, we can see the results of Total energy versus gate voltage.

Charging energy in SET phase is reduced due to stabilization of charge on island by electrostatic surrounding. Stabilization occurred since the $C_4H_6B_2$ molecule was

Table 1 Energies of $C_4H_6B_2$ molecule in isolated and SET environments

Environment	Energy	State	Charging energy (eV)
Energies of $C_4H_6B_2$ molecule in isolated condition	E^{I+1}	2	-24.1
	E^I	1	-9.05
	E^A	0	1.001
	E^{A-1}	-1	-3.35
Energies of $C_4H_6B_2$ molecule in SET environment	E^{I+1}	2	-17.14
	E^I	1	-7.28
	E^A	0	3.46
	E^{A-1}	-1	5.53

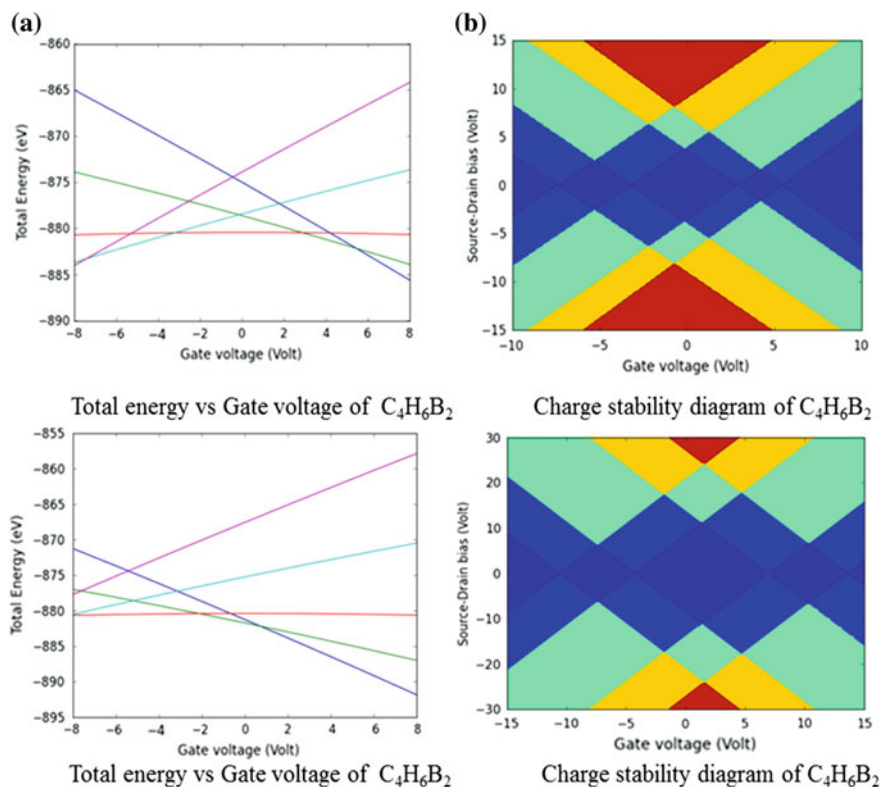


Fig. 7 **a** Total energy as the function of gate voltage of C₄H₆B₂ molecule with different electrodes [Gold ($W = 5.28$ eV) and caesium ($w = 2.14$ eV)]. Curve represents the charge state: violet [2], turquoise [1], red [0], green [-1] and blue [-2]. **b** Charge stability diagram of C₄H₆B₂ molecule with different electrodes. Gold ($W = 5.28$ eV) and caesium ($w = 2.14$ eV). The colours represent: (0) blue, (1) light blue, (2) green, (3) orange and (4) red (Colour figure online)

kept flat on dielectric. By tuning the gate voltage in SET, the charging energy of island can be modified. It moves the energy levels of SET in and out of the bias window. As it can be illustrated with charge stability diagram, number of energy levels in bias window is shown by colour [codes (0) blue, (1) light blue, (2) green, (3) orange and (4) red]. The work function of electrodes is changed and simulated total energies versus gate voltage are plotted as shown in Figs. 7 and 8.

Number of energy levels in bias window is directly related to conductance. With different electrodes (caesium, gold and osmium) separately in SET environment, the charge stability diagrams are plotted. It is seen in Figs. 7 and 8.

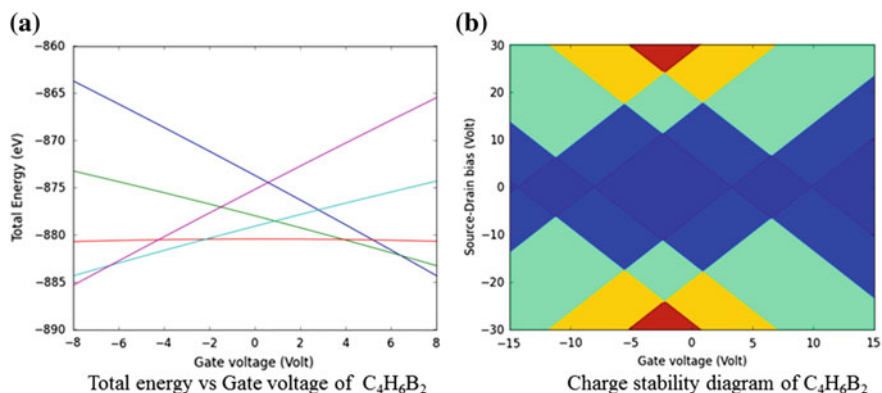


Fig. 8 **a** Total energy as the function of gate voltage of $C_4H_6B_2$ molecule in para configuration with osmium electrodes ($w = 5.93$ eV). Curve represents the charge state: violet [2], turquoise [1], red [0], green [-1] and blue [-2]. **b** Charge stability diagram of $C_4H_6B_2$ molecule with osmium electrodes ($w = 5.93$ eV). The colours represent: (0) blue, (1) light blue, (2) green, (3) orange and (4) red (Colour figure online)

6 Conclusions

In this work, we have taken benzene molecule and replaced the two carbon atoms with boron and optimized the structure. We used that molecule as island, and with different charge states we have calculated the Charging Energy, Total Energy, HOMO and LUMO. By this organic molecule, we got highest negative total energy, so it causes weak coupling of molecule with the source and drain. When the electron hops from source to island, the electron gets more time to localize and hence the electron loses all information about its original state in the source electrode and it operates in incoherent transport regime.

Acknowledgments The authors thank the department of science and technology of the government of India for partially funding this work.

References

- 1965—Moore's Law: Predicts the Future of Integrated Circuits. Computer History Museum. Retrieved 2009-03-19 (2007)
- Yoon, Y., Ganapathi, K., Salahuddin, S.: Nano Lett. **11**, 3768 (2011)
- Mak, K.F., Lee, C., Hone, J., Shan, J., Henix, T.F.: Phys. Rev. Lett. **105**, 136805 (2010)
- Abhinav, E., Chary, M., Vemana, D.: Strain-induced on germanene monolayer 6nm short channel FET from first-principle study. In: 2014 International Conference on Circuits, Communication, Control and Computing (I4C), pp. 1, 4, 21–22 Nov 2014. doi:[10.1109/CIMCA.2014.7057743](https://doi.org/10.1109/CIMCA.2014.7057743)

5. Brandbyge, M., Mozos, J.L., Ordejon, P., Taylor, J., Stokbro, K.: Density-functional method for nonequilibrium electron transport. *Phys. Rev. B* **65**, 165401 (2002)
6. Josephson, B.D.: Possible new effects in superconductive tunnelling. *Phys. Lett.* **1**, 251 (1962). doi:[10.1016/0031-9163\(62\)91369-0](https://doi.org/10.1016/0031-9163(62)91369-0)
7. Julliere, M.: Tunneling between ferromagnetic films. *Phys. Lett. A* **54**(3), 225–226 (1975). doi:[10.1016/0375-9601\(75\)90174-7](https://doi.org/10.1016/0375-9601(75)90174-7)
8. Grabert, H., Devoret, M.H. (eds.): *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*. Plenum Press, New York (1992)
9. Fulton, T.A., Dolan, G.J.: Observation of single-electron charging effects in small tunnel junctions. *Phys. Rev. Lett.* **59**, 109–112 (1987). doi:[10.1103/PhysRevLett.59.109](https://doi.org/10.1103/PhysRevLett.59.109)
10. Yano, K., Ishii, T., Hashimoto, T., Kobayashi, T., Murai, F., Seki, K.: Technical Digest of the 1993 International Electron Device Meeting, p. 541 (1993)
11. Atomistix Toolkit Version 12.8: Quantum wise A/S. <http://quantumwise.com>
12. Ghorbani-asl, M., Borini, S., Kuc, A., Heine, T.: *Phys. Rev. B* **23**, 235434 (2013)
13. Etymology for ab initio. etymonline.com. Retrieved 7 Feb 2010
14. Erdogan, E., Popov, I.H., Enyashin, A.N., Seifert, G.: *Eur. Phys. J. B* **85**, 33 (2012)
15. Tour, J.M., et al.: Recent advances in molecular scale electronics. *Ann. New York Acad. Sci.* **852**, 197–204 (1998)
16. Splendiani, A., Sun, L., Zhang, Y., Li, T., Kim, J., Chim, C.Y., Galli, G., Wang, F.: *Nano Lett.* **10**, 1271 (2010)
17. Popov, I., Seifert, G., Tománek, D.: *Phys. Rev. Lett.* **108**, 156802 (2012)

Analysis of Molecular Single-Electron Transistors Using Silicene, Graphene and Germanene

E. Meher Abhinav, Sai Naveen Kavuri, Thota Sandeep Kumar, Maragani Thirupathi, M. Chandra Mohan and A. Suresh Reddy

Abstract By using Ab initio approach, we have analysed Silicene-, Germanene- and Graphene-based molecular single-electron transistors. It is based on non-equilibrium greens function (NGEF) and density functional theory (DFT). Three different fullerene molecules are taken and optimization is done. In Coulomb blockade regime, silicene, germanene and graphene are kept above gate dielectric between drain and source for weak coupling. We have taken gold electrodes for SET environment. Gold is widely used as metal electrode in nanoscale devices. We have calculated the HOMO and LUMO values and total energy versus gate voltage. Charge stability diagrams are obtained by calculating charging energy as function of external gate potential. By these calculations, the analysis of three different molecular single-electron transistors is done. The total energies of these molecules are highly negative (very low) compared to other molecules.

Keywords Single-electron transistor (SET) · Graphene · Silicene · Germanene · Ab initio · Density functional theory (DFT) · Non-equilibrium greens function (NEGF) and GGA-PBE

1 Introduction

Transistor is considered to be one of the greatest inventions of twentieth century [1]. Modern digital IC, computer chips and logic gates are based on CMOS transistor technology.

E.M. Abhinav (✉) · S.N. Kavuri · T.S. Kumar · M. Thirupathi · M.C. Mohan · A.S. Reddy
Department of Electronic and Communications, Malla Reddy College of Engineering,
Dhulapally, Hyderabad 500014, Telangana, India
e-mail: abhi4abhi09@gmail.com

Moore's Law states that the number of transistors on a single silicon chip doubles about every two years [2]. As technology is scaled down to nanometres, second-order effects and parasitics are added. Hence tunnel junctions, SET and negative resistance devices are being used as alternative to CMOS.

The negative differential resistance concept has been demonstrated by quantum well structures. By using negative differential resistance, circuit applications like Analogue–digital converters, high frequency oscillators, logic gates and memories can be manufactured [3–5].

By using scanning tunnelling microscope, researchers studied self-assembled silicene sheets and nanoribbons deposited onto silver crystal. Synthesis of the first 2D material graphene was done by K.S. Novoselov and Andre Geim in 2004 [6]. Development of the emerging field of nanotechnology and nanoscience has been boosted by invention of scanning tunnelling microscope by Ch. Gerber, H. Rohrer and E. Weibel in 1981 [7] and atomic friction force microscopes [8].

Carbon has central focus both in physical science and in life due to small core and four valence electrons. Carbon can form different types of bonding and so it results in materials with different bondings. By sp^2 hybridization, carbon atoms form fullerene, nanotubes, graphite and many organic molecules. By sp^3 hybridization, carbon atom forms the tetrahedral bonds which build diamond structure. Despite the presence of fullerene (0D), nanotubes (1D) and diamond (3D) structures, 2D counterparts of carbon were absent till 2004 [9]. Till then, no 2D material was synthesized and many serious doubts were there on the stability of 2D material.

Graphene has unusual electronic properties with its honeycomb structure. The charge carriers are massless (Dirac fermions) [10]. The synthesis of graphene has demonstrated its stability. Novel properties of graphene have been elucidated in [11–14]. Besides size and fundamental properties of graphene, geometry-dependent magnetic and electronic properties of 1D quasi nanoribbon have been revealed [15–18]. While research interest on graphene ribbons is increasing rapidly, the question is whether other group IV elements such as Ge and Si have stable 2D structures.

Graphene is one of the highly investigated materials in past decade because of its huge advantages compared to conventional materials in semiconductor industry.

Ab initio based minimization of the total energy shows that honeycomb Si and Ge can have stable buckled structure [19–21]. Based on optimized atomic structure, the low-buckled honey comb structures of Ge and Si are found to be stable. Faster, novel devices with graphene have been proposed to replace their silicon counterparts. Major problems include band gap incompatibility with silicon-based technology. Recent discovery of silicene promises to be the solution to above problems. Silicene and germanene share same important electronic properties and it is easily incorporated in present technology with huge advantage and better natural band gap. So silicene and germanene can replace graphene.

2 Modelling of Silicene and Germanene

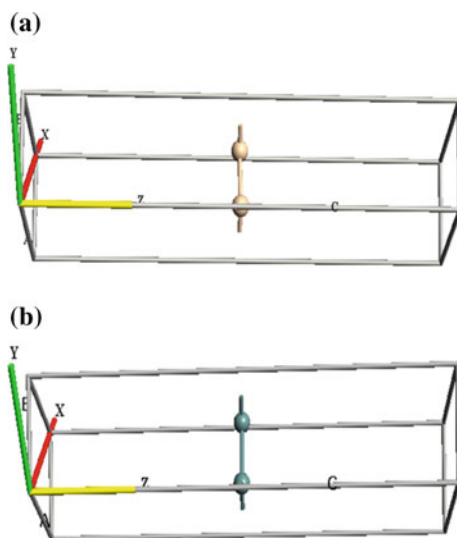
Initially, graphene molecule is taken and two atoms are replaced with silicon and germanium. Graphene has different lattice constants compared to silicene and germanene. The lattice constant of silicene is 3.8 \AA and for germanene it is 4.03 \AA . Lattice constant of silicene is kept 20 and 40 \AA for germanene. Rattle tool is used for adding small perturbation to co-ordinates. For good accuracy $21 \times 21 \times 1 \text{ k}$ points and double zeta polarized basic set are used along with GGA-PBE in exchange correlation. While optimizing, the low force tolerance is set to 0.01 eV/\AA and lower stress tolerance is set to 0.0005 eV/\AA . Unconstrained X and Y directions and structure are optimized. After optimization, we find buckling of 0.5 \AA for silicene and 0.7 \AA for germanene. After optimization, silicene and germanene molecules are formed. Optimized structure is seen in Fig. 1.

3 Computational Procedure

In this work, we have calculated charging energies of graphene, silicene and germanene for plotting charge stability diagram both in isolated and SET environments. We analysed total energy as function of gate voltage. For different charge states, total energy and charging energies are calculated using ATK 13.8.0 Tool Kit [22]. It is based on DFT and NGEF.

In computation, we used graphene, silicene and germanene molecule as island in SET. The island is placed between the source and drain electrode above gate dielectric. We have taken gate dielectric to be 3.8 \AA thick and the dielectric is kept

Fig. 1 **a** Unit cell of silicene molecule and **b** unit cell of germanene molecule



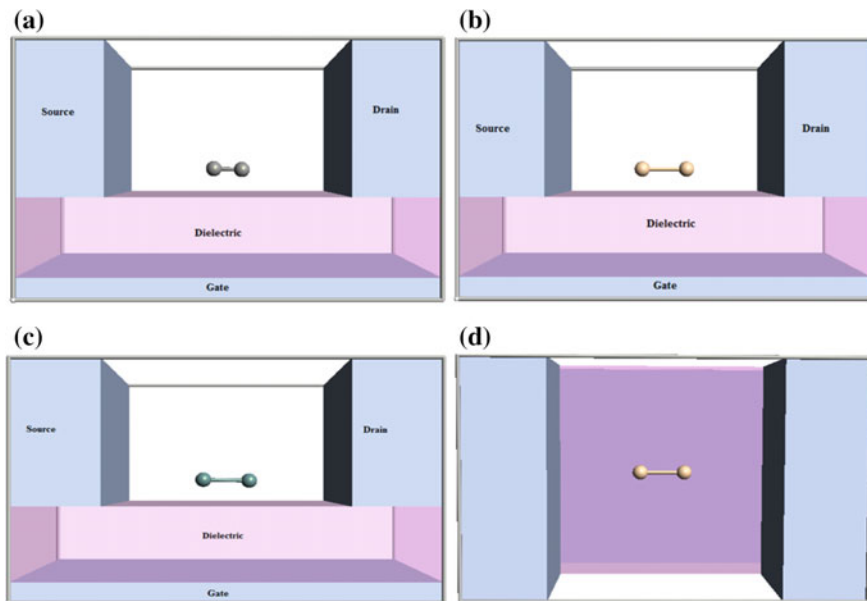


Fig. 2 **a** Graphene molecule in SET environment, **b** silicene molecule in SET environment, **c** germanene molecule in SET environment, **d** top view of silicene molecule in SET

on metallic back gate with dielectric constant of $10 \epsilon_0$. These molecules are placed above the dielectric. Figure 2 shows the graphene, silicene and germanene molecules in SET environments. We have taken gold electrode and its work function is 5.28 eV [23, 24]. Figure 2d shows the top view of SET environment of silicene molecular single-electron transistor. For calculating total energies of the molecule in isolated condition, we used DFT and LDA as exchange correlation function. We have taken Neumann boundary condition to keep the perpendicular component of electric field zero at boundaries.

4 Results and Discussion

Initially, for graphene, silicene and germanene molecules, total energies and charging energies are calculated in isolated condition (gas phase). Tables 1, 2, and 3 define the charging energies of the graphene, silicene and germanene for different charge states of the molecule ($-2, -1, 0, 1, 2$) in both SET and isolated (gas phase) environments.

The Total Energy of graphene in isolated condition with zero charge state is -304.11 eV.

Table 1 Charging energies of graphene in isolated and SET environment

Environment	Energy	State	Charging energy (eV)
Energies of graphene molecule in isolated condition	E^{I+1}	2	-34.17
	E^I	1	-11.89
	E^A	0	+2.21
	E^{A-1}	-1	-4.76
Energies of graphene molecule in SET environment	E^{I+1}	2	-26.96
	E^I	1	-9.60
	E^A	0	+4.52
	E^{A-1}	-1	+4.04

Table 2 Charging Energies of silicene in Isolated and SET environment

Environment	Energy	State	Charging energy (eV)
Energies of silicene molecule in isolated condition	E^{I+1}	2	-22.23
	E^I	1	-7.88
	E^A	0	+1.24
	E^{A-1}	-1	-3.64
Energies of silicene molecule in SET environment	E^{I+1}	2	-13.57
	E^I	1	-5.46
	E^A	0	+3.37
	E^{A-1}	-1	+4.52

Table 3 Charging Energies of germanene in Isolated and SET Environment

Environment	Energy	State	Charging energy (eV)
Energies of germanene molecule in isolated condition	E^{I+1}	2	-22.01
	E^I	1	-7.62
	E^A	0	+1.23
	E^{A-1}	-1	-3.61
Energies of germanene molecule in SET environment	E^{I+1}	2	-13.29
	E^I	1	-5.42
	E^A	0	+4.50
	E^{A-1}	-1	+3.35

The Total Energy of silicene in isolated condition with zero charge state is -349.79 eV.

The Total Energy of germanene in isolated condition with zero charge state is -554 eV.

The total energy of these molecules is very low so the coupling of molecule with the source and drain electrodes will be low, so when the electron hops from source to island the localization time increases. On increasing the localization time, the

electron loses the previous information. For electron to hop from source to island, the energy level of island must be low compared to source. When energy level of drain is low compared to island, the electron hops from island to drain.

Highest Occupied Molecular Orbital (HOMO) of graphene, $E_{\text{HOMO}} = -6.91$ eV, Lowest Unoccupied Molecular Orbital (LUMO) of graphene, $E_{\text{LUMO}} = -0.14$ eV. Highest Occupied Molecular Orbital (HOMO) of silicene, $E_{\text{HOMO}} = -4.43$ eV. Lowest Unoccupied Molecular Orbital (LUMO) of silicene, $E_{\text{LUMO}} = -0.19$ eV. Highest Occupied Molecular Orbital (HOMO) of germanene, $E_{\text{HOMO}} = -4.38$ eV. Lowest Unoccupied Molecular Orbital (LUMO) of germanene, $E_{\text{LUMO}} = -0.18$ eV.

Normally Fermi-level lies between occupied and unoccupied states and it can be anywhere in the HOMO-LUMO gap. From Tables 1, 2, and 3, we can see the charging energies of graphene, silicene and germanene for different charge states. From simulations, we got lower charging energy of -34.17 eV for graphene at charge state 2 in isolated condition and in SET environment at same charge state graphene has -26.96 eV. Here positive charges are stabilized in SET environment. At zero

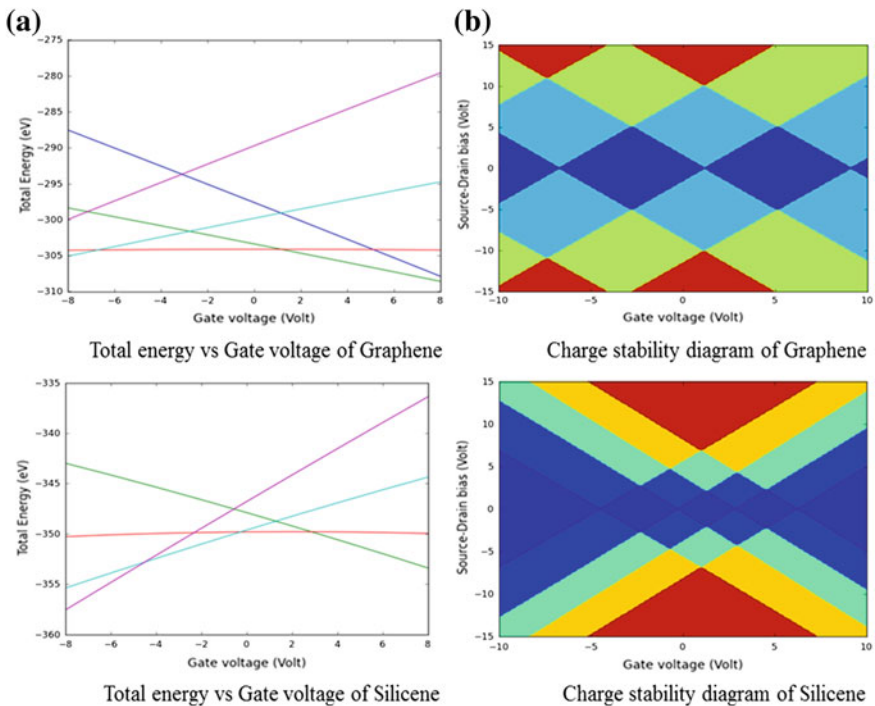


Fig. 3 **a** Total energy as the function of gate voltage for graphene and silicene molecules. Curve represents the charge state: violet [2], turquoise [1], red [0], green [-1] and blue [-2]. **b** Charge stability diagram of graphene and silicene. The colours represent: (0) blue, (1) light blue, (2) green, (3) orange and (4) red

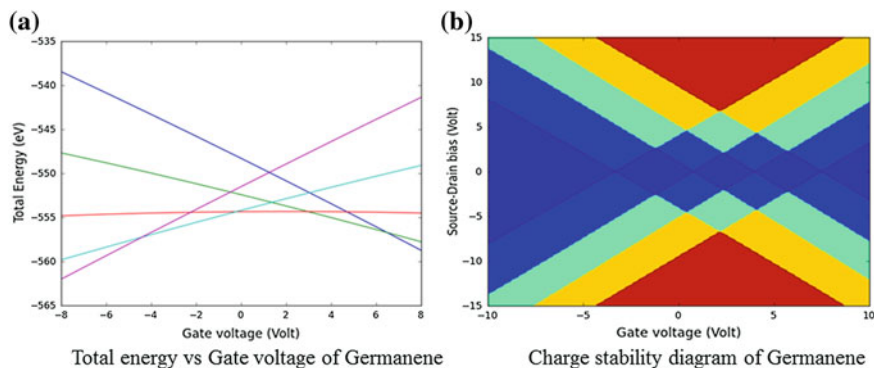


Fig. 4 **a** Total energy as the function of gate voltage for germanene molecule. *Curve* represents the charge state: *violet* [2], *turquoise* [1], *red* [0], *green* [-1] and *blue* [-2]. **b** Charge stability diagram of germanene. The colours represent: (0) *blue*, (1) *light blue*, (2) *green*, (3) *orange* and (4) *red*

state, graphene has high charging energy in both isolated and SET environments. From Table 2, the charging energy of silicene is high at state 1 in SET environment compared to state zero, and at state 2 it has lowest charging energy -22.23 eV. As in graphene, germanene has highest charging energy at state zero in SET environment as shown in Table 3. In SET environment, the positive charges are stabilized.

We calculated total energy of the different charge states of gas phase (isolated) and in SET environment as function of the gate potential. From Figs. 3 and 4, we can see the plots of the Total energy versus Gate voltages of graphene, silicene and germanene. Energy levels in the bias window are directly related with conductance. By using gold electrode in SET environment, the charge stability diagrams of graphene, silicene and germanene are plotted as shown in Figs. 3 and 4.

5 Conclusions

We have exploited the use of graphene, silicene and germanene molecule for the single-electron transistor. We used Ab initio framework for calculating the charging energies of graphene, silicene and germanene in metallic environment. We had calculated the charging energies as function of gate potential. The present study demonstrates the use of the first-principles to gain new properties of single-electron transistors operating in Coulomb blockade regime. We calculated the Total Energy, Charging Energy, HOMO and LUMO. Performance analysis of these molecules was done and we got highest negative total energies by using these molecules. Hence it causes weak coupling of molecule with source and drain. Therefore, the localization time increases and the electron loses the previous information and operates in the incoherent transport regime.

Total energies of silicene and germanene are nearly equal to graphene, so we can replace graphene with silicene and germanene. Silicene and germanene share important electronic properties and are easily incorporated in present technology with huge advantage. Hence, silicene and germanene can replace graphene and can be used for broad range of applications.

Acknowledgments The authors thank the department of science and technology of the government of India for partially funding this work.

References

1. 1965—"Moore's Law" predicts the future of integrated circuits. In: Computer History Museum. 2007. Retrieved 2009-03-19
2. Mollick, E.: Establishing Moore's law. *IEEE Ann. Hist. Comput.* **28**(3), 62–75 (2006)
3. Brown, E.R., Söderström, J.R., Parker, C.D., Mahoney, L.J., Molvar, K.M., McGill, T.C.: *Appl. Phys. Lett.* **58**, 2291 (1991)
4. Broekaert, T.P.E., Brar, B., van der Wagt, J.P.A., Seabaugh, A.C., Morris, F.J., Moise, T.S., Beam III, E.A., Frazier, G.A.: *IEEE J. Solid State Circuits* **22**, 1342 (1998)
5. Mathews, R.H., Sage, J.P., Sollner, T.C.L.G., Calawa, S.D., Chen, C.L., Mahoney, L.J., Maki, P.A., Molvar, K.M.: *Proc. IEEE* **87**, 596 (1999)
6. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Zhang, Y., Dubonos, S.V., Grigorieva, I.V., Firsov, A.A.: *Science* **306**, 666 (2004)
7. Binnig, G., Rohrer, H., Gerber, Ch., Weibel, E.: *Phys. Rev. Lett.* **49**, 57 (1982)
8. Binnig, G., Quate, C.F., Gerber, Ch.: *Phys. Rev. Lett.* **56**, 930 (1986)
9. Novoselov, K.S., Geim, A.K., Morozov, S.V., Jiang, D., Katsnelson, M.I., Grigorieva, I.V., Dubonos, S.V., Firsov, A.A.: *Nature* **438**, 197 (2005)
10. Fock, V.: *Z. Phys.* **61**, 126 (1930)
11. Katsnelson, M.I., Novoselov, K.S., Geim, A.K.: *Nat. Phys.* **2**, 620 (2006)
12. Zhang, Y., Tan, Y.-W., Stormer, H.L., Kim, P.: *Nature* **438**, 201 (2005)
13. Berger, C., Song, Z., Li, X., Wu, X., Brown, N., Naud, C., Mayou, D., Li, T., Hass, J., Marchenkov, A.N., Conrad, E.H., First, P.N., de Heer, W.A.: *Science* **312**, 1191 (2006)
14. Geim, A.K., Novoselov, K.S.: *Nat. Mater.* **6**, 183 (2007)
15. Abhinav, E.M., Chary, D.V.: Strain-induced on germanene monolayer 6 nm short channel FET from first-principle study. In: *Circuits, Communication, Control and Computing (I4C), 2014 International Conference on*, pp. 1, 4, 21–22 Nov 2014. doi:[10.1109/CIMCA.2014.7057743](https://doi.org/10.1109/CIMCA.2014.7057743)
16. Wang, X., Ouyang, Y., Li, X., Wang, H., Guo, J., Dai, H.: *Phys. Rev. Lett.* **100**, 206803 (2008)
17. Son, Y.-W., Cohen, M.L., Louie, S.G.: *Nature* **444**, 347 (2006)
18. Son, Y.-W., Cohen, M.L., Louie, S.G.: *Phys. Rev. Lett.* **97**, 216803 (2006)
19. Takeda, K., Shiraishi, K.: *Phys. Rev. B* **50**, 14916 (1994)
20. Zhang, M., Kan, Y.H., Zang, Q.J., Su, Z.M., Wang, R.S.: *Chem. Phys. Lett.* **379**, 81 (2003)
21. Durgun, E., Tongay, S., Ciraci, S.: *Phys. Rev. B* **72**, 075420 (2005)
22. Atomistix Toolkit Version 13.8.0 Quantum wise A/S: <http://quantumwise.com>
23. Riviere, J.C.: The work function of gold. *Appl Phys Lett* **8**, 172 (1966)
24. Seldenthuis, J.S., van der Zant, H.S.J., Ratner, M.A., Thijssen, J.M.: *Phys. Rev. B* **81**, 205430 (2010)

Identification of the Plants Based on Leaf Shape Descriptors

Pradip Salve, Milind Sardesai, Ramesh Manza and Pravin Yannawar

Abstract Plants are living organisms belonging to the vegetal kingdom that can live on land and in water. Plants form the critical base of food chains in nearly all ecosystems. Plants are vitally important for environmental protection and contribute to maintain biodiversity. Plant taxonomy has attracted many researchers to study the bio-diversities based on plants. Automated identification of plant species using leaf shape descriptor addresses the automatic classification of plants and simplifies taxonomic classification process. In this research work, we used Zernike moments (ZM) and Histogram of Oriented Gradient (HOG) method as a shape descriptor resulting 84.66 and 92.67 % accuracy for ZM and HOG, respectively, on ‘VISLeaf’ database.

Keywords Plant recognition · Zernike moments · Histogram of oriented gradients · Leaf shape

P. Salve (✉) · P. Yannawar
Vision & Intelligence Lab, Department of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India
e-mail: pradipslv@gmail.com

P. Yannawar
e-mail: pravinyannawar@gmail.com

M. Sardesai
Floristic Research Lab, Department of Botany,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India
e-mail: sardesaiimm@gmail.com

R. Manza
Biomedical Image Processing Lab, Department of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, (MS), India
e-mail: manzaramesh@gmail.com

1 Introduction

Most of the plants carry significant information and are considered as an essential resource for the well-being of humans. Scientifically, the discrimination among entities are based on their unique characteristics. In plants, leaves retain such unique and discriminative characteristics by means of venation architecture, geometric representation (size and shape). This complexity has generated curiosity in the minds of researchers from *Plant Sciences*, *Computer Sciences*, *Medicine*, *Pharmaceutical sciences*, *Mathematician* etc. Leaf patterns of different species exhibit a large variety of structures. Leaf shape has been investigated for its possible use in the systematic determination of species, by extracting its shape and size features, more precisely morphological and geometric features. This approach helped in characterizing the shape of leaf using machines.

Several attempts were made in order to classify these patterns; it was initiated by Von Ettinghausen [1]; his classification was refined and completed by Hickey [2]. However, the only color and margin of the leaf have not been considered as the safe criteria for the systematic classification and identification of plants because of its observed variability [3]. Plant identification can be performed using many organs namely *flowers*, *seeds*, *fruits*, *leaves*, and *woody* parts. Among these, leaves are the most appropriate for our experimental work targeted under this research work. Unlike other organs, leaves are easily available and they are generally observed throughout the year. Moreover, they contain a lot of information that are generally used for plant metadata formation and it becomes first possible method for plant identity and description.

Leaf images can be recognized either by color, texture, shape, or by an appropriate combination of these characteristics. Particularly for the plant recognition task, shape descriptors are mostly preserved. There are many methods reported by researchers in the literature for shape representation like *Chain code*, *Fourier descriptors*, *moments*, and *curvature scale space* which are just few of them. Two-dimensional shapes can be described either by encapsulating the information provided or by object's boundary, its features by description of the region occupied by the object on the image plane. An appropriate shape descriptor should be invariant to several geometrical transformations such as, *rotation*, *reflection*, *scaling*, and *translation*. Shape descriptor is a highly informative characteristic, since it is utilized in this research work for automatic recognition of plants based on leaf descriptor. This paper is organized in five section: Sect. 1 deals introduction; Sect. 2 deals with related work; the methodology adapted in this research work was discussed in Sect. 3; the obtained results were presented in Sect. 4 and conclusion in Sect. 5.

2 Related Work

Many researchers have contributed in this area; some of most promising works were discussed here; Singh et al. [4] proposed three techniques of plants classification based on leaf shape; they are as Probabilistic Neural Network with Principal Component Analysis (PCA), Support Vector Machine (SVM) utilizing Binary Decision Tree and Fourier Moment. These methods were helped in solving multiclass classification problems. The SVM-based Binary Decision Tree architecture has improved in both, that is the stable and efficient decision tree architecture resulting high classification accuracy of Support Vector Machine. Krishna Singh et al. observed that SVM-BDT is efficient than Fourier Moment and PNN techniques [4]. Bama et al. [5] used shape, color, and texture features. These features were used in HSV color space to extract the features. Texture feature extraction was carried out using Log-Gabor wavelet on the input image. All feature points were extracted using the Scale Invariant Feature Transform (SIFT). B. Sathya Bama et al. extended their work by incorporating HU's moments for shape feature extraction, and these computed features were matched using Euclidean distance classifier [5]. Jyotismita et al. [6] proposed Moments-Invariant (M-I) and the Centroid-Radii (C-R) modeling techniques. Jyotismita et al utilized M-I model and normalized central moments, and its combinations were considered for generation of optimal result. The C-R model was used as an edge detector for identification of leaf boundary generating its shape. The feature vector for all samples with criteria of 36 radii at 10° angular separation for marking leaf boundary was applied and passed to Neural Networks for recognition purpose [6]. Kadir et al. [7] incorporated *shape*, *vein*, *color*, and *texture features* to classify a leaf using Probabilistic Neural Network (PNN). Fourier Descriptors, Slimness Ratio, Roundness Ratio, and dispersion were also used for representing shape features [7]. Wang et al. [8] proposed the method to improve leaf image classification by using global features and local features of the leaves. Shape context was used as global feature, and SIFT (Scale Invariant Feature Transform) descriptors were also utilized for local features. Weighted K-NN algorithm was utilized for classification purpose [8]. Wijesingha et al. [9] used leaf length, width, area and perimeter and other morphological features to achieve automatic leaves image identification. Extracted features were fed to Probabilistic Neural Network (PNN) for classification [9]. Kadir et al. [10] used the combination of three geometric features that are Zernike Moments (ZM), Color Moments (CM), and Gray-Level Co-Occurrence Matrix (GLCM) for preparation of feature vector. For implementation purpose, two approaches have been investigated. In first approach, distance measures were used and in second Probabilistic Neural Networks (PNN) was implemented. The results show that the Zernike Moments have more clear features supporting for leaf identification [10].

Aptoula et al. [11] investigated performance of descriptors based on mathematical morphology. The first descriptor consists of the computation of morphological covariance on the leaf contour profile and the second descriptor was Circular Covariance Histogram for capturing leaf venation characteristics. These descriptors

were compared with *Contour Morphological Covariance* (CC) and *Extended Circular Covariance Histogram* (ECCH) against standard *Morphological Covariance* (MCOV), *Angle Code Histogram* (ACH), *Contour Point Distribution Histogram* (CPDH), *Rotation Invariant LBP*, and the standard *Circular Covariance Histogram* (CCH) descriptors. Classification was done with the help of nearest neighbor classifier [11]. Bhardwaj et al. [12] used the Nearest Neighborhood Classifier in their study. Work presented the use of Hu's moment, that gives seven features descriptor of different plant leaves and different morphological feature extraction and parametric calculations such as smooth factor, aspect ratio, leaf area, rectangularity, circularity, eccentricity, etc., and area convexity was computed for recognition purpose [12]. Bong et al. [13] suggested the Centroid Contour Gradient (CCG) feature extraction method calculates the gradient between pairs of boundary points corresponding to interval angle. It was observed that the CCG had better efficiency compared to Centroid Contours Distance (CCD), because it captures the curvature of the tip and base of leaf. Mei Fern Bong et al. used the Feed-forward Back-Propagation Neural Network as a classifier [13]. Satti et al. [14] adopted a combination of color, shape, tooth features as well as morphological features of the leaves. The classification was performed using Neural Networks and Euclidean classifier [14]. Mouin et al. [15] classify the plants based on the visual information provided by the plant leaves. They considered two sources of information which are leaf margin and the leaf salient points. Sofiene Mouin et al. have introduced two shape context-based descriptors: first one represents the leaf boundary while the second represents the spatial correlation between salient points of the leaf and its margin. Sofiene Mouin et al. also studied the performance of the fusion of descriptors [15].

Kulkarni et al. [16] used the shape, vein, color, and texture features with the combination of pseudo Zernike movements. Classification was done by Radial Basis Probabilistic Neural Network (RBFNN) [16]. Larese et al. [17] used leaf shape, size, texture, and color features for plant recognition. The segmentations were completed by the means of the unconstrained hit-or-miss transform and adaptive thresholding. Classification was done by the Support Vector Machines (SVM), Penalized Discriminant Analysis (PDA), and Random Forests Methods (RF) [17]. Mythili et al. [18] classified the medical plant leaves using the geometric features such as diameter, leaf length, leaf width, and tooth features. Leaves were segmented using Effective Robust Kernelized Fuzzy C-Means (ERKFCM). Classification has been performed using Support Vector Machine Classifier (SVM), and finally Artificial Neural Network (ANN) was used to recognize leaf [18]. Pradhan et al. [19] proposed physiological features for recognition of plant leaves. The features were classified using Probabilistic Neural Network (PNN) [19]. Amlekar et al. [20] extracted leaf venation morphological feature to classify them. Multi-layer perceptron based Artificial Neural Network method [20]. These trends in the design of automatic plant recognition that based on leaf shape have been tested on various databases as well as variety of feature extraction schemes.

3 Methodology

This piece of work primarily focused on Zernike Moments (ZM) and HOG Descriptor applied over Centered binary leaf images and cropped grayscale leaf images from ‘VISLeaf’ dataset, respectively. The dataset was constituted with 180 plant species collected from the Botanical garden of Dr. Babasaheb Ambedkar Marathwada University campus, Aurangabad (MS) India. Total volume of ‘VISLeaf’ contains 1800(180*10) leaf samples. In this research work we used subset of ‘VISLeaf’ containing 50 kinds of plant species and each species includes 10 sample images. Hence there are totally 500(50*10) images in the proposed dataset (Fig. 1).

Automatic plant identification process begins with the input leaf image. At the time of data collection, fresh leaf samples were plucked from Botanical Garden and were scanned with the resolution of 300 dpi. Metadata related to the sample was also recorded with taxonomical classification. These scanned samples were enhanced and passed for feature extraction. All extracted features of training set were saved on the disk and later used for testing and classification purpose.

3.1 Feature Extraction

3.1.1 Zernike Moments (ZM)

We use Zernike Moments (ZM) to extract features using the shape of leaf. We compute the Zernike moments from an input leaf image sample following these three steps: computation of radial polynomials, computation of Zernike basis functions, and computation of Zernike moments by casting the image on to the Zernike basis functions. The procedure for obtaining Zernike moments from an input image begins with the computation of Zernike radial polynomials [21].

$$Z_{mn} = \frac{m + 1}{\pi} \int_x \int_y I(x, y) [V_{mn}(x, y)] dx dy, \tag{1}$$

where ‘*m*’ defines the order of Zernike polynomial of degree ‘*m*’, ‘*n*’ defines the angular dependency, and *I(x, y)* be the gray level of a pixel of image on which the moment is calculated.

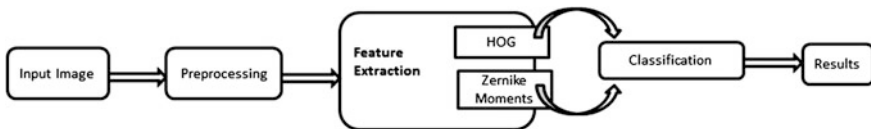


Fig. 1 Overall process for plant recognition

The Zernike polynomials $V_{mn}(x, y)$ are expressed in polar coordinates using radial polynomial (R_{mn}) as per Eqs. (2) and (3).

$$V_{mn}(r, \theta) = R_{mn}(r)e^{-jn\theta} \quad (2)$$

$$R_{mn}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s \frac{(m-s)!}{s! \left[\frac{m+|n|}{2} - s \right]! \left[\frac{m-|n|}{2} - s \right]!} r^{m-2s}. \quad (3)$$

The resultant Zernike moments Z_{mn} are invariant under rotation, scale and translational changes. Zernike Moments are the pure statistical measure of pixel distribution around center of gravity of shape and allows capturing information just at single boundary point. They can capture some of the global properties missing from the pure boundary-based representations. Zernike moments have mathematical properties that make them the ideal image features to be used as shape descriptors in shape classification problems. Michael V. et al. [22] presented the technique which explored the optimal use of Zernike Moments as a descriptor and some experiments have been done by Borde et al. [23] on ‘vVISWa’ dataset. We have calculated Zernike Moments up to order $n = 12$ and formulated Zernike feature vector containing 49 features for each sample. These features were calculated for all the samples of ‘Training set’ and ‘Test set.’ The ‘Zernike Feature Matrix’ for the samples of ‘Training set’ is shown in Table 1 and ‘Test set’ is shown in Table 2.

3.1.2 Histogram of Oriented Gradients (HOG)

The HOG descriptors are local statistic of the orientations of the image gradients around key points [24, 25]. A HOG descriptor is extracted from image local region, and we considered each leaf sample as a one single local region in this research work. The image sample was divided into $M*M$ cells after preprocessing; each cell has a pixel size of $N*N$. The gradient magnitude g and the gradient orientation θ were computed for all the pixels in the block using Eqs. (4) and (5); the derivatives (g_x) and (g_y) of the image I were computed with pixel differences using Eqs. (6) and (7).

$$g(\theta, \omega) = \sqrt{g_x(\theta, \omega)^2 + g_y(\theta, \omega)^2} \quad (4)$$

$$\theta(\theta, \omega) = \arctan \frac{g_y(\theta, \omega)}{g_x(\theta, \omega)} \quad (5)$$

$$g_x(\theta, \omega) = I(\theta + 1, \omega) - I(\theta - 1, \omega), \quad (6)$$

$$g_y(\theta, \omega) = I(\theta, \omega + 1) - I(\theta, \omega - 1). \quad (7)$$

Table 1 Zernike moment features for training set

Known leaf samples	Moment 1	Moment 2	Moment 3	Moment 4	Moment 5	...	Moment 47	Moment 48	Moment 49
1	77.75856	0.421456	-181.26	-6.85063	-0.89602	...	4.339012	0.668848	0.029231
2	110.5445	-3.38274	-234.741	-12.704	8.903311	...	3.743267	0.685658	0.039514
3	64.70785	-0.61887	-143.405	-6.02157	0.27391	...	-1.71328	0.001795	0.010238
4	78.66801	-6.33958	-184.329	-4.99839	17.85077	...	4.484767	0.656743	0.03365
5	97.99397	-4.12805	-216.46	-14.9965	11.56273	...	5.5169	0.738296	0.037488
6	56.84105	-2.98124	-117.703	-8.62756	5.070737	...	-0.29347	-0.47975	-0.01262
7	126.2326	-10.0928	-234.874	-23.1098	28.07237	...	-3.0651	1.289855	0.174406
8	109.3622	-3.45815	-237.331	-10.92	8.982125	...	3.105148	0.472418	0.026597
9	91.76419	1.091348	-193.86	-18.8074	-5.70007	...	-0.77591	-1.02049	-0.15835
:	:	:	:	:	:	...	:	:	:
:	:	:	:	:	:	...	:	:	:

Table 2 Zernike movement features for test set

Known leaf samples	Moment 1	Moment 2	Moment 3	Moment 4	Moment 5	...	Moment 47	Moment 48	Moment 49
1	65.57184	-0.86953	-157.864	-7.19407	3.224065	...	5.893298	0.956701	0.052449
2	74.80282	8.151839	-176.22	-6.86686	-24.9254	...	5.294342	0.517398	0.000907
3	97.72114	-1.25771	-215.353	-7.0441	3.617263	...	3.566638	0.900063	0.061227
4	83.7155	-0.10869	-194.423	-7.56672	2.266452	...	-1.30561	-0.57115	-0.04454
5	89.44508	-2.28251	-203.655	-4.01741	6.650382	...	5.368352	0.782443	0.008044
6	64.93522	3.260735	-156.295	-9.23458	-9.95418	...	6.996193	1.464478	0.103416
7	68.11832	-3.94394	-165.013	-6.03066	12.84926	...	4.559873	0.632003	0.023117
8	114.9099	-0.76084	-237.057	-15.6788	1.92062	...	3.01589	1.157132	0.078043
9	65.57184	-0.86953	-157.864	-7.19407	3.224065	...	4.335175	1.217952	0.096438
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	...	:	:	:	:

After gradient computation, each pixel within a cell casts a weighted vote for an orientation-based histogram based on the gradient magnitude and orientation. This histogram divides the gradient angle range into K bins. The cells were normalized by histogram so that the effect of noise and overlapping cells was reduced. We consider the histogram in the same cell as a feature. Let F be the feature, and for getting feature vector we used the function as follows:

$$F = \frac{F}{\sqrt{\|F\|_2^2 + \varepsilon}}. \quad (8)$$

The histograms of all the blocks combined into a whole HOG descriptor were processed. For example, in this work of feature extraction using HOG, the block size of 3×3 cells was considered and the number of bins i.e. K is set to 9. As a result, algorithm calculates $81(3 * 3 * 9)$ blocks, so the dimension of overall HOG feature is 81 for each image. This was used further for classification purpose. Dataset was divided into 70 % known samples and 30 % unknown samples. The System was trained over 70 % known set and 30 % unknown samples were tested on the known set and success recognition of plant species was evaluated. Table 3 shows the training set and Table 4 shows the testing set.

4 Experimental Results

The recognition of plant species was carried out by Zernike moment and HOG-based features. These features were calculated for all samples of training set and stored for recognition purpose as a feature vector. The recognition of leaf samples was divided into two parts: first recognition by Zernike features and second recognition by HOG features.

4.1 Zernike Features-Based Recognition

For plant species, recognition using Zernike features was computed for all leaf samples and considered it as training and test set. The calculated training and test set were loaded into the machine for further process of plant species recognition. These feature vectors of training samples and test samples were tested for getting the minimum distance index using *minimum* distance classifier. We considered the 'index' provided by classifier as the correctly classified plant species class from the information obtained between each pair (one vector from test set and other vector from training set) of observations. The resultant confusion matrix of training samples and test samples using Zernike moment is shown in Table 5.

Table 3 HOG-based features for train set

Known leaf samples	F1	F2	F3	F4	F5	...	F79	F80	F81
1	0.208353	0.234566	0.442596	0.26377	0.507841	...	0.282264	0.260643	0.217594
2	0.213664	0.233403	0.415034	0.269162	0.521817	...	0.300757	0.311725	0.275429
3	0.264304	0.279324	0.416075	0.225411	0.493739	...	0.291849	0.245247	0.254807
4	0.233169	0.277899	0.442072	0.265329	0.435073	...	0.357607	0.311253	0.270283
5	0.271641	0.228432	0.369083	0.251361	0.522816	...	0.264287	0.241343	0.293376
6	0.24964	0.28438	0.439906	0.250599	0.436839	...	0.441629	0.259656	0.223369
7	0.287118	0.169654	0.577579	0.194941	0.526851	...	0.367056	0.315031	0.253209
8	0.226611	0.153502	0.592977	0.193349	0.535891	...	0.332522	0.326393	0.285111
9	0.268931	0.198983	0.505586	0.212044	0.491421	...	0.316813	0.28886	0.287233
:	:	:	:	:	:	...	:	:	:
:	:	:	:	:	:	...	:	:	:

Table 4 HOG-based features for test set

Known leaf samples	F1	F2	F3	F4	F5	...	F79	F80	F81
1	0.212627	0.23506	0.441837	0.217808	0.556511	...	0.280534	0.259609	0.260309
2	0.21471	0.23796	0.427174	0.26698	0.507448	...	0.41566	0.238652	0.204317
3	0.236453	0.235787	0.397133	0.265942	0.523465	...	0.385834	0.307147	0.251926
4	0.252054	0.256935	0.434326	0.215459	0.498048	...	0.290687	0.252896	0.245929
5	0.227718	0.251636	0.424993	0.256594	0.499526	...	0.297719	0.291904	0.266785
6	0.220281	0.22336	0.460269	0.236036	0.522398	...	0.427789	0.257082	0.237993
7	0.220226	0.249701	0.43528	0.219691	0.537259	...	0.405236	0.244232	0.257477
8	0.232914	0.279924	0.440937	0.240242	0.457245	...	0.372766	0.285954	0.237928
9	0.223486	0.273273	0.445324	0.265882	0.441633	...	0.345817	0.302861	0.241916
:	:	:	:	:	:	:	:	:	:
:	:	:	:	:	:	...	:	:	:

Table 5 Confusion matrix produced on 50 plant datasets using ZM method

Class no.	Class name—species	No. of test samples/class	Recognized	Missed	Accuracy (%)
1	<i>Alstonia scholaris</i>	3	3	0	100
2	<i>Crescentia alata</i>	3	2	1	66.66
3	<i>Syzygium cumini</i>	3	3	0	100
4	<i>Tridax procumbens</i>	3	3	0	100
5	<i>Eriogloss amedule</i>	3	2	1	66.665
6	<i>Lantana camara</i>	3	3	0	100
7	<i>Holoptelia integrifolia</i>	3	2	1	66.66
8	<i>Milletia puguensis</i>	3	3	0	100
9	<i>Annona reticulata</i>	3	3	0	100
10	<i>Ficus benghalensis</i>	3	3	0	100
11	<i>Citrus limon</i>	3	2	1	66.66
12	<i>Catharanthus roseus</i>	3	1	2	33.33
13	<i>Dregea volubilis</i>	3	3	0	100
14	<i>Santalum album</i>	3	2	1	66.665
15	<i>Cocculus hirsutus</i>	3	3	0	100
16	<i>Acalypha indica</i>	3	2	1	66.66
17	<i>Barleria prionitis</i>	3	3	0	100
18	<i>Rauvolfia serpentina</i>	3	3	0	100
19	<i>Carissa carandas</i>	3	3	0	100
20	<i>Oxalis corniculata</i>	3	3	0	100
21	<i>Ficus racemosa</i>	3	2	1	66.66
22	<i>Sterculiaceae abroma</i>	3	0	3	0
23	<i>Pisum sativum</i>	3	3	0	100
24	<i>Senna occidentalis</i>	3	3	0	100
25	<i>Hibiscus syriacus</i>	3	3	0	100
26	<i>Jatropha curcas</i>	3	2	1	66.66
27	<i>Adansonia digitata</i>	3	3	0	100
28	<i>Pithecellobium dulce</i>	3	2	1	66.66
29	<i>Sphagneticola trilobata</i>	3	3	0	100
30	<i>Euphorbia grantii</i>	3	3	0	100
31	<i>Murraya paniculata</i>	3	3	0	100
32	<i>Bryophyllum pinnatum</i>	3	2	1	66.66
33	<i>Passiflora edulis</i>	3	3	0	100
34	<i>Combretum indicum</i>	3	3	0	100
35	<i>Turnera ulmifolia</i>	3	2	1	66.66

(continued)

Table 5 (continued)

Class no.	Class name—species	No. of test samples/class	Recognized	Missed	Accuracy (%)
36	<i>Hamelia patens</i>	3	3	0	100
37	<i>Jasminum multiflorum</i>	3	3	0	100
38	<i>Helianthus cucumerifolius</i>	3	2	1	66.66
39	<i>Nerium oleander</i>	3	3	0	100
40	<i>Tabernaemontana divaricata</i>	3	3	0	100
41	<i>Ipomoea purpurea</i>	3	2	1	66.66
42	<i>Ipomoea cairica</i>	3	3	0	100
43	<i>Cestrum nocturnum</i>	3	2	1	66.66
44	<i>Tecoma stans</i>	3	3	0	100
45	<i>Acalypha wilkesiana</i>	3	2	1	66.66
46	<i>Euphorbia heterophylla</i>	3	2	1	66.66
47	<i>Lythraceae lagerstroemia</i>	3	3	0	100
48	<i>Rhamnaceae ventilago</i>	3	2	1	66.66
49	<i>Gmelina arborea</i>	3	3	0	100
50	<i>Jasminum azoricum</i>	3	2	1	66.66
Total			127	23	84.66

4.2 HOG Features-Based Recognition

HOG feature samples from training set and test set were computed. The minimum distance classifier was used for measuring the similarity between test samples and training sample. Table 6 shows confusion matrix for HOG features-based recognition. It was seen that 139 samples were recognized correctly out of 150 samples, and over all recognition rate was measured to be 92.67 %. The resultant confusion matrix of training samples and test samples using HOG features is shown in Table 6.

Table 6 Confusion matrix produced on 50 plant datasets using HOG method

Class no.	Class name—species	No. of test samples/class	Recognized	Missed	Accuracy (%)
1	<i>Alstonia scholaris</i>	3	3	0	100
2	<i>Crescentia alata</i>	3	3	0	100
3	<i>Syzygium cumini</i>	3	3	0	100
4	<i>Tridax procumbens</i>	3	3	0	100
5	<i>Eriogloss amedule</i>	3	3	0	100
6	<i>Lantana camara</i>	3	3	0	100
7	<i>Holoptelia integrifolia</i>	3	1	2	33.33
8	<i>Milletia peguensis</i>	3	3	0	100
9	<i>Annona reticulata</i>	3	2	1	66.66
10	<i>Ficus benghalensis</i>	3	3	0	100
11	<i>Citrus limon</i>	3	3	0	100
12	<i>Catharanthus roseus</i>	3	3	0	100
13	<i>Dregea volubilis</i>	3	3	0	100
14	<i>Santalum album</i>	3	3	0	100
15	<i>Cocculus hirsutus</i>	3	3	0	100
16	<i>Acalypha indica</i>	3	3	0	100
17	<i>Barleria prionitis</i>	3	3	0	100
18	<i>Rauvolfia serpentina</i>	3	2	1	66.66
19	<i>Carissa carandas</i>	3	2	1	66.66
20	<i>Oxalis corniculata</i>	3	2	1	66.66
21	<i>Ficus racemosa</i>	3	2	1	66.66
22	<i>Sterculiaceae abroma</i>	3	3	0	100
23	<i>Pisum sativum</i>	3	3	0	100
24	<i>Senna occidentalis</i>	3	3	0	100
25	<i>Hibiscus syriacus</i>	3	3	0	100
26	<i>Jatropha curcas</i>	3	3	0	100
27	<i>Adansonia digitata</i>	3	3	0	100
28	<i>Pithecellobium dulce</i>	3	3	0	100
29	<i>Sphagneticola trilobata</i>	3	3	0	100
30	<i>Euphorbia grantii</i>	3	3	0	100
31	<i>Murraya paniculata</i>	3	3	0	100
32	<i>Bryophyllum pinnatum</i>	3	3	0	100
33	<i>Passiflora edulis</i>	3	3	0	100
34	<i>Combretum indicum</i>	3	2	1	66.66
35	<i>Turnera ulmifolia</i>	3	3	0	100

(continued)

Table 6 (continued)

Class no.	Class name—species	No. of test samples/class	Recognized	Missed	Accuracy (%)
36	<i>Hamelia patens</i>	3	3	0	100
37	<i>Jasminum multiflorum</i>	3	3	0	100
38	<i>Helianthus cucumerifolius</i>	3	3	0	100
39	<i>Nerium oleander</i>	3	3	0	100
40	<i>Tabernaemontana divaricata</i>	3	3	0	100
41	<i>Ipomoea purpurea</i>	3	3	0	100
42	<i>Ipomoea cairica</i>	3	3	0	100
43	<i>Cestrum nocturnum</i>	3	3	0	100
44	<i>Tecoma stans</i>	3	3	0	100
45	<i>Acalypha wilkesiana</i>	3	2	1	66.66
46	<i>Euphorbia heterophylla</i>	3	2	1	66.66
47	<i>Lythraceae lagerstroemia</i>	3	2	1	66.66
48	<i>Rhamnaceae ventilago</i>	3	3	0	100
49	<i>Gmelina arborea</i>	3	3	0	100
50	<i>Jasminum azoricum</i>	3	3	0	100
Total			139	11	92.67

Table 7 Comparison of recognition rate of proposed methods

Method	Recognition rate (%)
Zernike moments	84.66
HOG	92.67

Table 7 shows the overall performance of recognition system among fifty plant species. By using Zernike Moments as a features we achieved 84.66 % of recognition rate. And it indicates that ZM has low accuracy as compared with HOG based features. In recognizing 50 types of leaves using two different feature extraction techniques, i.e., Zernike Moments and Histogram of Oriented Gradients, the accuracy was 84.66 and 92.67 %, respectively, using 'Euclidean' minimum distance classifier.

5 Conclusion

In this research work, the results show that Zernike moments (ZM) and Histogram of Oriented Gradients (HOG) approach have contributed in the design of automatic plant recognition system based on the leaf shapes descriptors. The performance of HOG features was found to be excellent over Zernike moments. Increase in the HOG performance ratio was majorly due to its robustness and feature persistence capability. HOG generates more robust shape descriptor features compared to Zernike Moments. These features may be combined together with Neural network or other classifiers. And this may lead for design of more robust automatic plant identification system based on leaf shapes.

References

1. Von Ettinghausen, C.R.: Die Blatt-skelette der Dykotyledonen (Kaiserlich Ko'nigliche Hof-und Stattsdruckerei, Wien (1861)
2. Hickey, L.J.: *Am. J. Bot.* **60**, 17 (1973)
3. Klucking, E.P.: *Leaf Venation Patterns*, Vol. 1. J. Cramer, Berlin (1986)
4. Singh, K., Gupta, I., Gupta, S.: SVM-BDT PNN and fourier moment technique for classification of leaf shape. *Int. J. Signal Proc. Image Proc. Pattern Recogn.* **3**(4), (2010)
5. Bama, B.S., Valli, S.M., Raju, S., Kumar, V.A.: Content based leaf image retrieval (CBLIR) using shape, color and texture features. *Indian J. Comput. Sci. Eng.* **2**(2), 202–211 (2011)
6. Jyotismita, C., Parekh, R.: Plant leaf recognition using shape based features and neural network classifiers. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2**(10), (2011)
7. Kadir, A., Nugroho, L.E., Susanto, A., Santosa, P.I.: Leaf classification using shape, color, and texture features. *Int. J. Comput. Trends Technol.* (2011)
8. Wang, Z., Lu, B., Chi, Z., Feng, D.: Leaf image classification with shape context and SIFT descriptors. In: *International Conference on Digital Image Computing: Techniques and Applications* (2011)
9. Wijesingha, D., Marikar, F.M.M.T.: Automatic detection system for the identification of plants using herbarium specimen images. *Trop. Agric. Res.* **23**(1), 42–50 (2011)
10. Kadir, A., Nugroho, L.E., Santosa, P.I.: Experiments of Zernike moments for leaf identification. *J. Theor. Appl. Inf. Technol.* **41**(1), (2012)
11. Aptoula, E., Yanikoglu, B.A.: Morphological features for leaf based plant recognition. In: *ICIP* (2013)
12. Bhardwaj, A., Kaur, M., Kumar, A.: Recognition of plants by leaf image using moment invariant and texture analysis. *Int. J. Innov. Appl. Stud.* **3**(1), 237–248 (2013)
13. Bong, M.F., Sulong, G.B., Mohd Rahim, M.S.: Recognition of leaf based on its tip and base using centroid contour gradient. *IJCSI Int. J. Comput. Sci. Issues* **10**(2), (2013)
14. Satti, V., Satya, A., Sharma, S.: An automatic leaf recognition system for plant identification using machine vision technology. *Int. J. Eng. Sci. Technol.* (2013)
15. Mouine, S., Yahiaoui, I., Blondet, A.V.: Plant species recognition using spatial correlation between the leaf margin and the leaf salient points. In: *ICIP 2013—IEEE International Conference on Image Processing* (2013)
16. Kulkarni, A.H., Rai, H.M., Jahagirdar, K.A., Kadko, R.J.: A leaf recognition system for classifying plants using RBPNN and pseudo Zernike moments. *Int. J. Latest Trends Eng. Technol.* **2**, 1–11 (2013)

17. Larese, M.G., Namias, R., Craviotto, R.M., Arango, M.R., Gallo, C., Granitto, P.M.: Automatic classification of legumes using leaf vein image features. *Pattern Recogn.* **47**, 158–168 (2014)
18. Mythili, C., Kavitha, V.: Recognition of plant leaf in medicine. *J. Convergence Inf. Technol. (JCIT)* **9**, (2014)
19. Pradhan, A.K., Mohanty, P., Behera, S.: A real time based physiological classifier for leaf recognition. *Int. J. Adv. Comput. Res.* **4**, (2014)
20. Amlekar, M., Manza, R.R., Yannawar, P., Gaikwad, A.T.: Leaf features based plant classification using artificial neural network. *J. Manag. Res.* **3**, (2014)
21. Tsolakidis, D.G., Kosmopoulos, D.I., Papadourakis, G.: Plant leaf recognition using Zernike moments and histogram of oriented gradients. In: *Artificial Intelligence: Methods and Applications*, pp. 406–417. Springer International Publishing (2014)
22. Michael, V.: Shape classification using Zernike moments. In: *Technical Report. iCamp-University of California Irvine* (2011)
23. Borde, P., Varpe, A., Manza, R., Yannawar, P.: Recognition of isolated words using Zernike and MFCC features for audio visual speech recognition. *Int. J. Speech Technol.* (2014)
24. Xiao, X.-Y., Hu, R., Zhang, S.-W., Wang, X.-F., et al.: “HOG-based approach for leaf classification. In: *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pp. 149–155. Springer Berlin Heidelberg (2010)
25. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection, pp. 886–893. *CVPR, San Diego* (2005)

Use of Global Positioning System to Track Movement of Mobile Node in Proxy Mobile Internet Protocol Version 6

Nitesh M. Tarbani and A.S. Alvi

Abstract In Proxy Mobile Internet Protocol version 6 (PMIPv6), while moving in networks, whenever mobile node (MN) gets connected to new mobile access gateway (MAG), every time MAG needs to verify authenticity of MN due to which, there is a significant increase in hand over delay. To reduce this, researchers have proposed that as soon as MN gets disconnected from current MAG, current MAG should send the authentication information of MN to new MAG to which MN next gets connected. To predict new MAG, movement of MN should be tracked. In this paper, a method is proposed to track the movement of MN using GPS.

Keywords Pmpiv6 · Mobile node · GPS · AAA server · MAG

1 Introduction

In the mobile world, being connected to the network always is a huge challenge. While moving, mobile node may get disconnected from one point of access and get connected to another point of access. When MN gets connected to new access point, authenticity of mobile node is verified every time. During this time of authentication, though mobile node is in the network of new access point, it cannot use the network to send or receive data. This results in increase in handover delay. To avoid this, researchers have proposed that as soon as MN gets disconnected from current MAG, current MAG should send the authentication information of MN to new MAG to which MN gets connected next. This solution requires the prediction of next MAG to which MN is going to attach and prediction of next

N.M. Tarbani (✉) · A.S. Alvi
Department of Computer Science & Engineering, Pof. Ram Meghe Institute
of Technology & Research, Badnera-Amravati, India
e-mail: ntarbani@gmail.com

A.S. Alvi
e-mail: abrar_alvi@rediffmail.com

MAG requires MN movement tracking. In this paper, a method is proposed in which MAG will track the movement of MN using GPS. As many of today’s mobile nodes have GPS tracking system installed into it. In the future, it may be the case that most mobile nodes would have installed GPS tracking system into it. Using GPS tracking, next MAG to which MN is going to attach can be predicated.

The rest of the paper is organized as follows. Section 2 describes PMIPv6. Section 3 describes the literature review related to this field. Section 4 describes the proposed work. Section 5 describes handover. Section 6 presents the conclusion.

2 Working of PMIPv6

As shown in Fig. 1, PMIPv6 has three main entities: authentication authorization and accounting (AAA) server, MAG, and local mobility anchor (LMA). LMA is responsible for maintaining the MNs reachability state and is the topological anchor point for the MNs home network prefix. The mobile access gateway is the entity that performs the mobility management on behalf of an MN and it resides on the access link where the MN is anchored. When MAG senses MN in its area, it confirms MN’s authenticity using policy profile. This policy profile is stored in the central server called the authentication, authorization, and accounting (AAA) server.

Every MN in PMIPv6 has to be first registered to LMA. When MN wants to use the network, it needs to send router solicitation message to MAG. This router

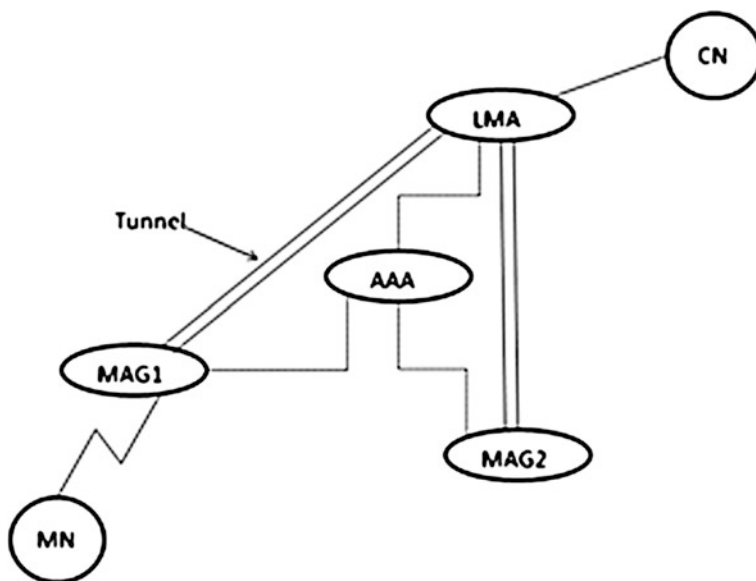


Fig. 1 Working of PMIPv6

solicitation message includes MN-ID. Upon receiving router solicitation message, MAG sends query message containing MN-ID to AAA server to know LMA to which this MN is registered.

On AAA server, the policy profile of each MN is stored. On receiving the query message from MAG, AAA server finds the policy profile of that MN using its MN-ID. If policy profile of that MN is not found, then MN is not registered. If policy profile of that MN is found, then LMA address is sent to MAG using reply message.

On receiving a reply message from AAA server, MAG retrieves LMA address message from the reply message. MAG creates proxy binding updates message (PBU) and sends this PBU to LMA.

On receiving PBU, LMA creates bidirectional tunnel between LMA and MAG and creates a message called proxy binding acknowledgement (PBA) which includes the address prefix(s) for the MN.

Upon receiving PBA from LMA, MAG sends router advertisement message to MN. After receiving router advertisement message finally, MN is able to send or receive the message from the network [1]. PMIPv6 is implemented in NS-2 but without AAA server [2]. AAA server in NS-2 is implemented in another research work [3].

3 Handover

Handover refers to the process of getting disconnected from one point of attachment to another point of attachment as shown in Fig. 2. Handover delay is the difference between the time when MN gets disconnected from one point of attachment and the time at which MN is again able to send or receive data in the network. In Fig. 2, MN gets disconnected from MAG1 at time t_1 . MN comes in the network of MAG2, MN sends router solicitation message to MAG2. MN is able to send or receive data in the network at time t_3 as MAG2 receives PBA at time t_3 , so, for the above figure (Fig. 2), handover delay is $t_3 - t_1$.

4 Literature Review

Worldwide research work is being carried out by academia and industries to reduce handover delay in both the host-based mobility model such as MIPv6 and HMIPv6, and network-based mobility model.

Ming-Chin Chuang and Jeng-Farn Lee proposed FH-PMIPv6 that uses predictive handoff mode to inform the target MAG about the handoff information (i.e., MNID and authentication information) in advance. Then, in FHPMIP the authentication and registration phases are performed simultaneously to decrease the handoff latency [4].

Ahmad Rasem et al. proposed a new protocol called Optimized Proxy Mobile IPv6 which combines the handover with route optimization (O-PMIPv6).

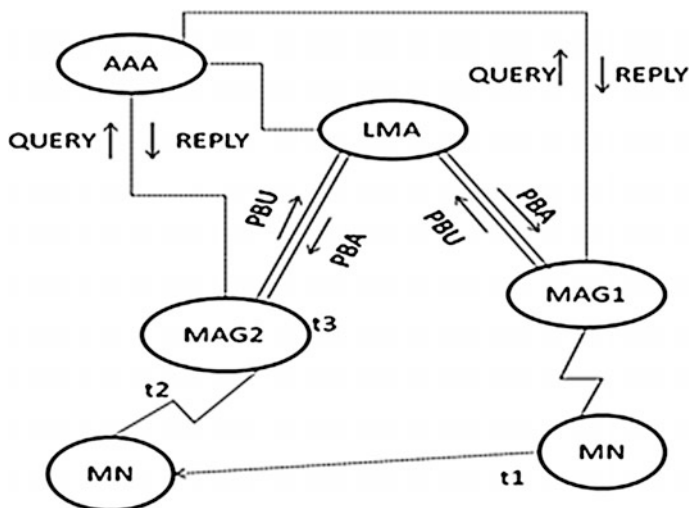


Fig. 2 Handover

Their proposed protocol enhances the performance of PMIPv6 and F-PMIPv6 by optimizing handover delay, signaling cost, and network utilization [5].

Zhiwei Yan et al. proposed FC-PMIPv6, which establishes pointer forwarding chain between MAGs to cut down location update traffic to an LMA during handovers of an MN [6]. In FC-PMIPv6, each MAG maintains a pointer table (PT). Each entry in the PT consists of three fields: HNP, previous, and next fields. The HNP field holds the MN's HNP. The previous field holds the MN's care-of address (CoA) used in the previous MAG area, which is the address of MN's previous MAG. Then, the next field supplies the MN's CoA for the next MAG area, which is the address of target MAG. If the next field is null, the MN resides in that MAG area.

Geumsan Jo et al. proposed a scheme in which the local mobility anchor (LMA) maintains the movement history of an MN, and anticipates the nMAG by comparing the present location of the MN and the previous location [7].

5 Proposed Work

Many of the above researchers have proposed various ways to reduce the handover delay of PMIPv6 in which current MAG sends authentication information of MN to new MAG, but there is need for a mechanism that can anticipate new MAGs. In real circumstances, each MAG would be surrounded by many other MAGs. So before sending authentication information of MN to a new MAG, the current MAG should anticipate MAG to which MN will be attached next. This paper proposes a method

used by the current MAG to predict new MAGs to which MN will attach next using GPS. In this proposed method, it is assumed that GPS tracking unit will be installed in each and every mobile node (Even today many cell phones have installed GPS tracking unit). Following are the steps in prediction of new MAG.

- For prediction of next MAG, it is necessary for current MAG to detect the movement of MN.
- For movement detection, MAG should have the locations of MN from time to time. To achieve this, MAG should be enabled to get MN locations from GPS on a timely basis.
- So, a new pack is introduced called LOCQ (Location Query) which is sent by MAG to GPS and GPS will reply to MAG by LOCR (Location Reply) to MAG which include the current location of MN.
- MAG will receive LOCR packets time to time. These packets will be stored in buffer called LOC buffer.
- As MAG will be receiving many LOCR packets, buffer may get full in a very short period of time.
- Solution for this problem is to delete the older data of LOC buffer as they show stale location of MN.
- When current MAG discovers that link between MN is going down, it will try to find direction of MN using locations stored in LOC buffer of that particular MN.

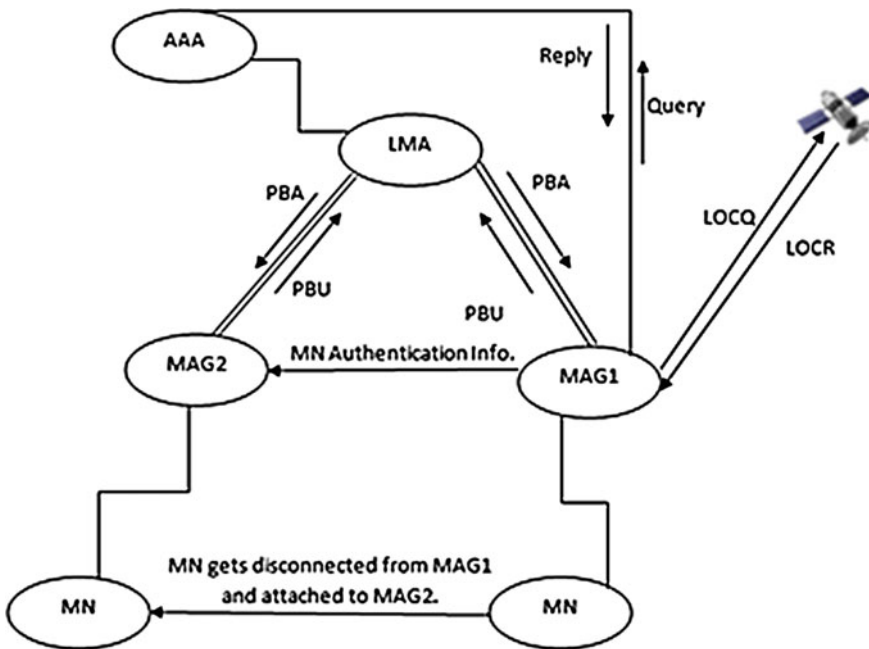


Fig. 3 Proposed methodology

- Using this direction, MAG can predict the next MAG to which MN may attach.
- As current MAG has predicted new MAG, current MAG can send the authentication information of MN to predict the next MAG.
- After handover, when MN attaches to the next MAG, MAG does not need to send query and reply messages to AAA for getting authentication information of MN (which is working of PMIPv6) as it is already available in MAG.
- After getting attached to new MAG, MN does not need to wait for authentication procedure as this new MAG already has authentication information of MN which leads to reducing the handover delay.
- As soon as this MN gets attached to new MAG, MN will be able to send/receive packet to/from corresponding node (Fig. 3).

6 Conclusion

In PMIPv6, minimizing handover delay is a big challenge. To reduce handover delay there is mainly one way, that is, to send authentication information of MN from current MAG to new MAG to which MN will get attached next. To predict the next MAG, tracking MN's movement is essential. In this paper, a method is proposed in which current MAG will track movement of MN using GPS. Using this movement next MAG is predicted and authentication information will be received by new MAG before MN is reached and this will lead to handover delay reduction.

References

1. Devarapalli, V., Wichorus, K.C.: Proxy mobile IPv6. In: RFC 5213, August 2008
2. Liza, F.F., Yao, W.: Implementation architecture of proxy mobile IPv6 protocol for NS2 simulator software. In: 09th, International Conference on Communication Software and Networks IEEE, July 2009
3. Tarbani, N., Chandavarkar, B.R.: Implementation of AAA sever for PMIPv6 in NS-2. In: Springer: International Conference on Parallel, Distributed Computing Technologies and Applications, September 2011
4. Chuang, M.-C., Lee, J.-F. FH-PMIPv6: a fast handoff scheme in proxy mobile IPv6 networks. In: International Conference on Consumer Electronics, Communications and Networks IEEE, November 2011
5. Rasemy, A., Makaya, C., St-Hilairey, M.: O-PMIPv6: efficient handover with route optimization in proxy mobile IPv6 domain. In: International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) IEEE, December 2012
6. Yan, Z., Lee, J.-H., Senior Member: State-aware pointer forwarding scheme with fast handover support in a PMIPv6 domain. *IEEE Syst. J.* 7(1), (2013)
7. Jo, G., Choe, H.J., Choo, H.: Predictive handover scheme using mobility history in PMIPv6. In: *Research in Adaptive and Convergent Systems*, vol. 13, ACM, October 2013

Is Mobile Cloud Computing Efficient for E-Learning?

Aftab Alam, Abdul Mateen Ansari, Mohammad Mujahid Barga and Ahmad Albhaishi

Abstract In the modern era, cloud computing is an emergent research area, spanning over multiple disciplines. It is a new service-centric technology that reduces the cost of resources by leveraging the efficient pooling of on-demand, self-management computer-generated infrastructure, consumed as a service through the Internet. Mobile network integrated with cloud computing to create new technology that is mobile cloud computing (MCC) to bring benefits for mobile users. Cloud computing popularity is rising by leaps and bounds in all domains like IT, business, electronic media, communication, academia, scientific research and education, etc. In present era; cloud computing has changed the traditional education systems to modern education systems like e-learning. E-Learning is the virtualized distance learning system; which is an electronic communication mechanism over the internet. In this paper, we will explore “is mobile cloud computing useful/efficient for E-Learning,” the most recent advancement of mobile cloud computing. We join the several new technologies to reach this goal. We are going to summarize present features and characteristics of E-learning and also look at the idea of E-learning through mobile cloud computing (MCC). We will also discuss the mobile cloud computing architecture (MCCA) and explain the design of mobile cloud computing stage by merging the feature of E-Learning.

A. Alam (✉) · M.M. Barga · A. Albhaishi
Department of Computer Science, King Khalid University, Abha, Kingdom
of Saudi Arabia
e-mail: aftabjh@gmail.com

M.M. Barga
e-mail: connecttomujju@gmail.com

A. Albhaishi
e-mail: albhaishi@gmail.com

A.M. Ansari
Department of Information Systems, Bisha University, Bisha, Kingdom
of Saudi Arabia
e-mail: mateenjmi@gmail.com

Keywords Cloud computing · Mobile cloud computing · E-learning · Mobile cloud computing architecture (MCCA)

1 Introduction

Earlier food, clothing, and shelter were the basic needs for a human being but at present time mobile phone has been included in that list. Our attachment and dependency toward mobile phones increases day-by-day and it also affects our daily routine activities. People use it for both personal and professional purposes. According to IDC (International Data Corporation), the number of PCs will fall from 28.7 % of the device market in 2013 to 13 % in 2017. Tablets will increase from 11.8 % in 2013 to 16.5 % by 2017, and smart phones will increase from 59.5 to 70.5 %. The internet has changed today's traditional education system forever. Mobile computing devices like smart phones, tablets and laptops are playing significant roles (Fig. 1).

E-learning plays significant roles in this new learning process. E-learning is the term used to describe the learning format, where training is accessible quickly online. E-learning eliminates the time bound barriers and brings concept of 24/7 learning. It has changed the style of learning and acquiring the skills. Many students and working employees are using e-learning to get their degrees. Today's people use some different kind of term to describe e-learning like web-based learning, online learning, distributed learning, and virtual learning. The growth of e-learning is depending on the widespread accessibility to ICT (information and communications technology) as well as lowering its cost's. E-learning has registered exceptional growth in the last



Fig. 1 Mobile e-learning cloud

decade and there is no sign of decline its growth in future also. In 2011, e-learning market was \$35.6 billion and the prediction is that it will be \$51.5 billion in 2016. E-learning is not just for education businesses also gradually start relying on e-learning. Across the globe, e-learning is growing. Cloud computing is acting like a game changer. The cloud is transforming the way employees, partners, and organizations interact and cooperate with each other. Cloud technology provides more efficient collaboration and improves organization’s efficiency and effectiveness. Cloud computing offers many solutions, SaaS (Software-as-a-Service) is playing key role. Gartner has predicted a rapid growth of cloud computing in 2015.

2 Cloud Computing

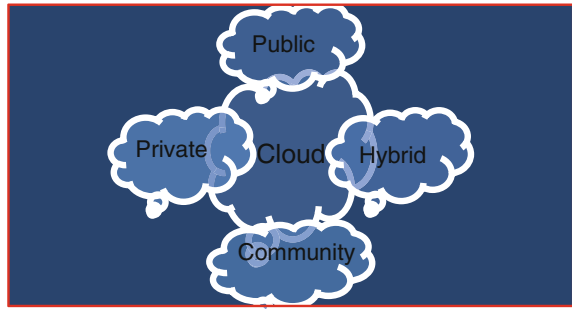
The fame of cloud computing is growing by the day, it is not just a promise or rather only discussed at school, college, university, seminars, and conferences anymore. It is the scalability and reality to fast, on-demand, automated, and low cost management of IT resources that have joint to make information technology a gripping paradigm for larger productivity and improved efficiency. There is an increasing apparent idea that cloud computing will make computing the sixth service after mobile, telephone, gas, water, and electricity that will offer the elementary level of computing services considered important for fulfilling daily life routines (Fig. 2).

Cloud computing uses the Internet as its backbone to offer elastic on-demand and energetically scalable computing infrastructure for various applications using any of its four deployment models: public, private, community, or hybrid cloud.



Fig. 2 Illustration of cloud computing

Fig. 3 Cloud computing deployment models



2.1 Cloud Computing Deployment Models

Public: It is standard cloud computing model; all the resources are available in this model such as application software and storage, available for universal public over the internet. Public cloud services accessible to user on pay-per-usage or may be offered free.

Private: The cloud infrastructure is provisioned for special use by a single organization embracing several clients (e.g., commercial units). It may be owned, managed, and functioned by the organization, a third party, or some combination of them, and it may exist on or off premises. Upgrade of the private cloud model is possible with the request and that is given using third party.

Community: A community cloud is a multitenant infrastructure that is shared among multi organizations from a specific community with common computing concerns (jurisdiction, security, compliance, etc.), hosted internally or externally and whether operated by a third party or internally. The expenses divided over rarer users than a public cloud (but more than a private cloud), so only some of the expenses reserves potential of cloud computing are recognized.

Hybrid: A hybrid cloud is a cloud computing environment in which configuration of two or more clouds (Private, Community, or Public) that becomes remain single entities but are bound together, present the benefits of various deployment models. In this atmosphere, organization provides and manages some resources internally and others provided externally (Fig. 3).

2.2 Cloud Computing Service Delivery Models

There are three-service delivery models: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) (Fig. 4).

Software as a Service (SaaS): In the commercial model using software as a service (SaaS), users are provided access to application software and databases. Infrastructure and Platforms are managed by cloud service providers which are

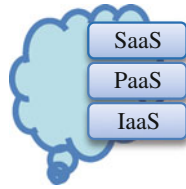


Fig. 4 Cloud computing service delivery models

hosted applications. It is frequently offered to as “on-demand software” and generally its cost is on a pay-per-use basis. Its providers usually cost software application using a subscription fee. Cloud providers offer a facility to install all software in the cloud and cloud machinists to authorized extract of permissible services and the required facilities from cloud consumers like software application. Cloud machinists do not prosper the hardware and operating system where the software is going on. In this situation, cloud provider removes all the major work of operators such as run and install of all software on the own personal electronic machine, which is easy to maintain, upgrading of software, and hardware support. Cloud applications are better and efficient from other comparative applications in their scalability such as Google, Twitter, Facebook, and Flickr are all examples of SaaS.

Platform as a Service (PaaS): Platform as a Service is a type of cloud services and plays a vital role in cloud computing. It is a new heterogeneous type of cloud computing that provides an environment and platform to permit developers to figure applications and services over the internet. PaaS services are hosted in the cloud and accessed by users simply via their web browser. PaaS services are normally paid for on a payment basis with clients finally paying just for what they use.

Infrastructure as a Service (IaaS): Infrastructure as a Service (IaaS) is one of the core service of cloud computing. It offer’s best cloud computing services to access the computing infrastructure (All hardware resources) in a virtualized healthy atmosphere “the Cloud” across a public connection, usually the internet. It offers network connections, bandwidth, IP address, virtual server space, and load balancers. IaaS can be operated by enterprise clients to make cost effective and easy scalable IT solutions where the complexities and costs of managing the underlying hardware are outsourced to the cloud service provider (Fig. 5).

3 Mobile Cloud Computing

Mobile cloud computing is the combination of mobile devices and cloud computing to create a new infrastructure by which cloud execute the heavy lifting of computing tasks and storing huge amounts data. In MCC data is executed and stored outside the mobile devices (Fig. 6).

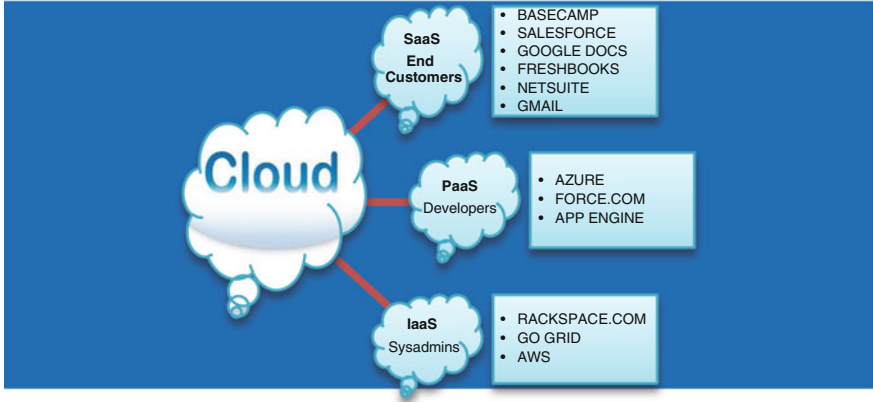


Fig. 5 Diagram of the model for the services oriented architecture

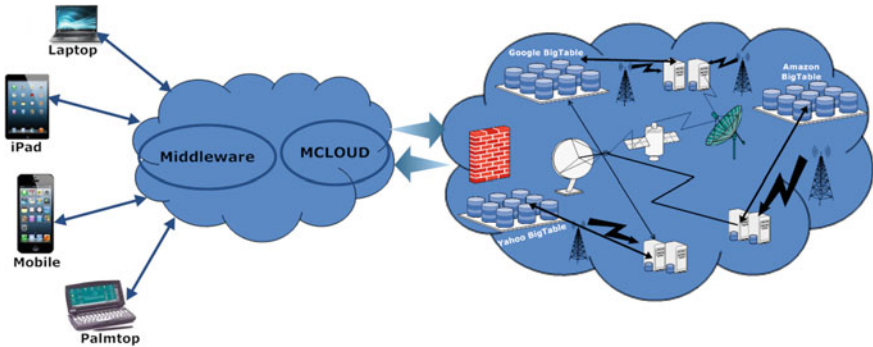


Fig. 6 Mobile cloud computing

Mobile cloud computing is the new paradigm of cloud computing. Its popularity increases day-by-day. “According to the modern research from Juniper Research, the number of mobile cloud computing subscribers is expected to grow rapidly in the next 5 years. Cloud-based mobile market will generate annual revenue of \$9.5 billion in 2014 from \$400 million in 2009, at an average annual increase of 88 %” [1].

3.1 Key Factors or Reasons Behind the Acceptance of Mobile Cloud Computing

Enabling technologies: “HTML5, CSS3, hypervisor for mobile devices, cloudlets and Web 4.0 will drive adoption of mobile cloud computing” [2].

Improved and increased broadband coverage: 3G and 4G along with Wi-Fi offering better connectivity for smart phones.

Trends and demands: Customers want to access the companies' web URL or applications through the internet anywhere and anytime. Smart phones can deliver such kind of facilities. The business person wants to connect business applications and services to increase their production from anywhere and anytime even they are on travel.

3.2 Major Trends in Mobile Cloud Computing

- Acceleration in the “Consumerization” of IT through Mobile Computing.
- Risk Challenges are Evolving.
- Mobile Computing will be Revolutionize How Work is done.
- Mobile Computing will become the “Internet of Things”.
- Mobile Computing is Here to Stay Whether People or IT are Prepared or Not.

3.3 Advantages of Mobile Cloud Computing

- Extending battery lifetime.
- Improving data storage capacity and processing power.
- Improving reliability and scalability.
- Offer facility to store all the data in one place and access it from anywhere.
- Multi-tenancy.
- Dynamic provisioning.
- Scalability.

3.4 Applications of Mobile Cloud Computing

- Mobile commerce
- Mobile learning
- Mobile healthcare
- Mobile gaming

3.5 Challenges of Mobile Cloud Computing

- Signal disturbance.
- Limitation of bandwidth and communication latency.
- Limited power, memory screen.
- Low computing ability-processing power.
- Quality of service (affected by weather and buildings).

4 Mobile Cloud Computing (MCC) Architecture

This is the basic architecture of mobile cloud computing (MCC). Figure 7 is the basic architecture having two mobile networks A and B. Mobile network A and B may have devices which are attached via a wireless access point, satellite or BTS. Different network services like database, home agent (HA), authentication, authorization and accounting (AAA) is running on the servers available in the mobile network [3]. The mobile user's request is executed by the central processor; who are directly associated with servers. Afterward, on the basis of HA and the user's information stored in the databases, the mobile network operator provides AAA services to the users [3]. After this process the user credentials quit mobile network and linked to the cloud (possess by data center owners or cloud service provider) via internet. When the user enters the cloud, the cloud controller connects the user's credential with the significant cloud depending upon the service credentials requested by that user. The user can ask for different kind of services like computing resources, virtualization, applications, databases, and storage services in the cloud.

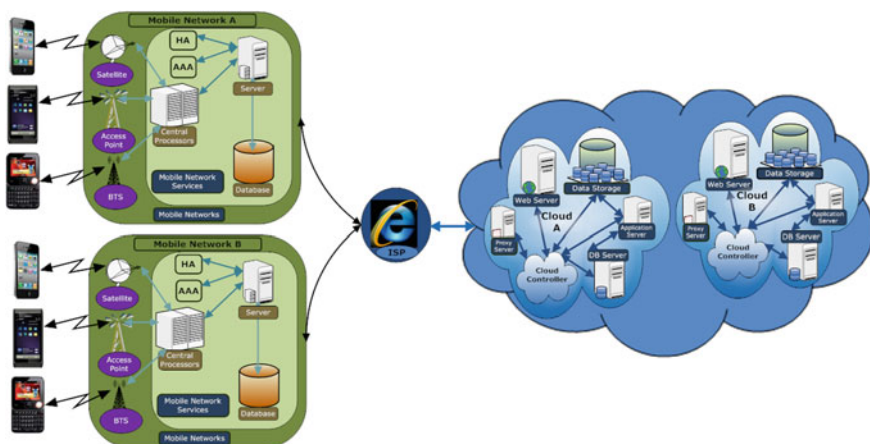


Fig. 7 Mobile cloud computing architecture

5 E-Learning

E-Learning includes all the pattern of electronic learning in which the internet and other learning techniques are used for the delivery of the teaching and training process, and for organizing and managing the interaction between teacher, instructor, trainer, and student, among students themselves, and between students and sources of learning. It is considered to be both face-to-face and electronic learning (Fig. 8).

E-learning applications and processes include web-based learning, computer-based learning, virtual education opportunities, and digital collaboration [4]. It can be self-paced or instructor-led and includes media in the form of text, image, animation, streaming video, and audio [4]. In the present time, it is broadly used on various educational organizations in different levels: distance education, supportive learning, corporate training, university organization courses, etc. There are several e-learning solutions from free source to trade. Students and instructors are most famous and powerful entities associated with an e-learning system. E-learning provides no restriction of period, location, and communication. Heterogeneous environment, for running top-level e-learning, needs huge resources and infrastructure of any educational organization or company in low cost. Cloud computing have ability to provide all resource and infrastructure economically

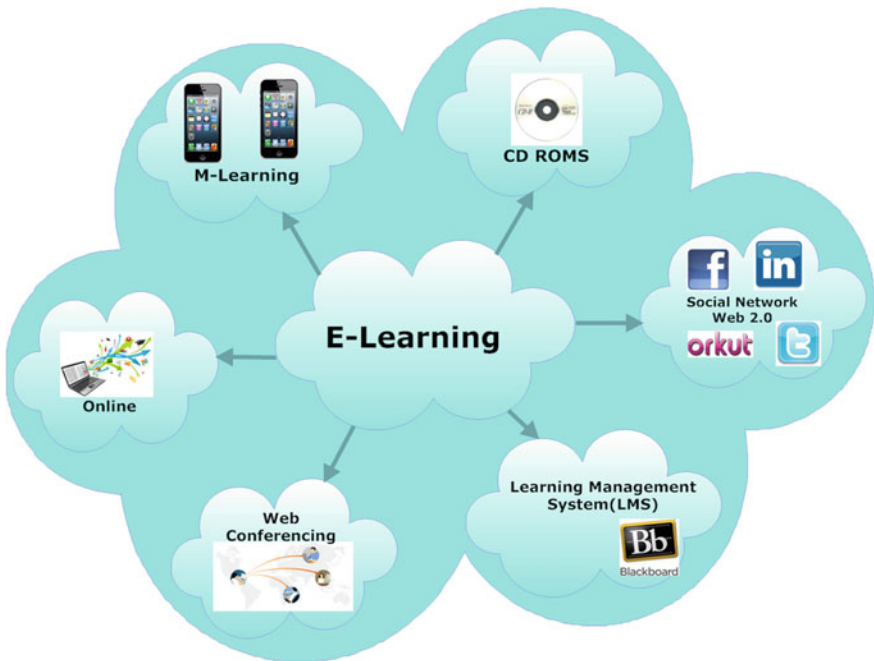
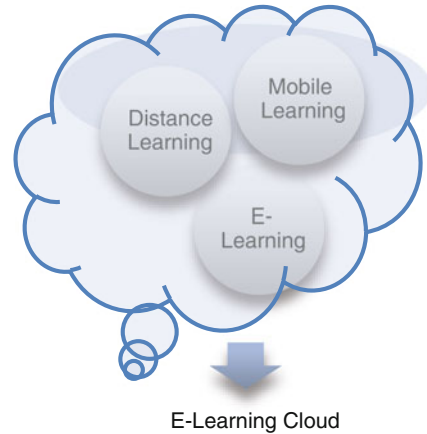


Fig. 8 E-learning

Fig. 9 E-learning cloud

on-demand and pay-per-use at any time anywhere. In e-learning atmosphere, there are two biggest players such as Black Board and Moodle. Now it cloud have oriented version of base application. Modern and developing period is became high-tech savvy, all the users are using e-learning through the smart electronic devices (like ipad, ipod, laptop, tab, and mobile phone) in heterogeneous and healthy environment (Fig. 9).

6 Usage of Mobile Cloud Computing for E-Learning

At present, e-learning is used various academia, education, and learning stage through smart electronic devices in heterogeneous mobile cloud computing environment. For example, software training for software engineering in corporate sector, hardware training for commercial network engineering, academic courses, all application trainings for all users in different sectors like health, banking, posting, army, etc. There are several e-learning modules available in the market from free source to corporate. “There are at least two entities involved in an e-learning system: the students and the trainers” [5]. E-learning systems can use advantage from mobile cloud computing using:

The students:

- Take online course
- Take online quiz
- Take online attendance
- Take online Exam
- Take online discussion
- Send feedback
- Send homework, projects

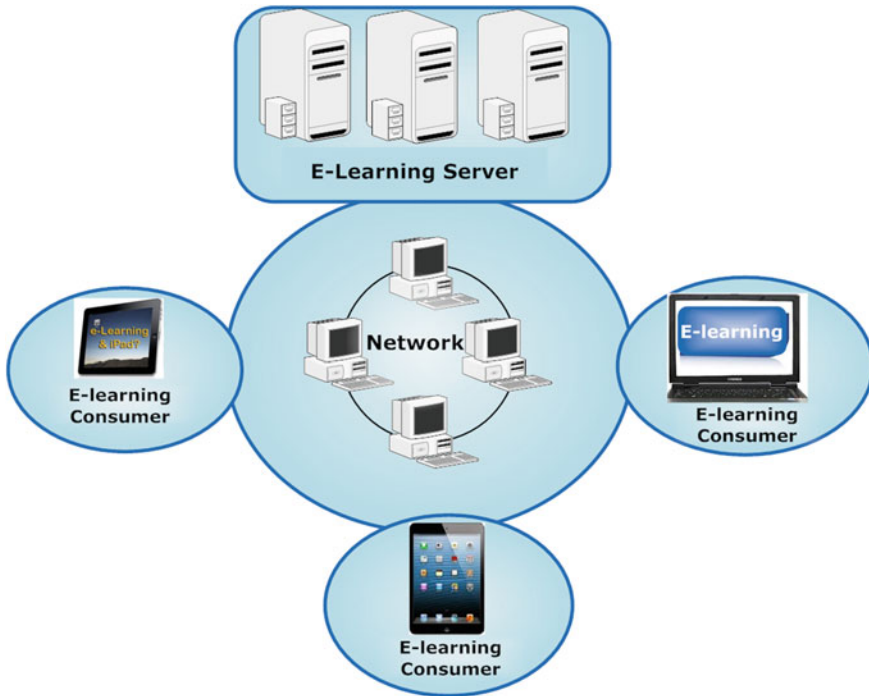


Fig. 10 Working model of e-learning

The trainers:

- Prepare quiz, discussion, assignment, question paper
- Match able content of the course
- Assessment tests, projects taken by students
- Send opinion
- Edugram (SMS-Chatting Application)
- Connect with learner (forums)

This is a mega system. This system is run smooth due to developed and de-centralized applications, but this is not compulsory. The design of a de-centralized e-learning system includes applications devices, such as a database server, client application, an application server (see Fig. 10), and the essential hardware devices (client healthy electronic devices, communication setup, and e-learning servers).

The consumer infrastructure could be electronic smart device, that device uses a humble web browser or a committed application as client software. Even with the present software and hardware restrictions, electronic smart devices are associated with multimedia-based applications. Currently, electronic smart devices applications, particularly multimedia-based applications, have solicitous restrictions due to

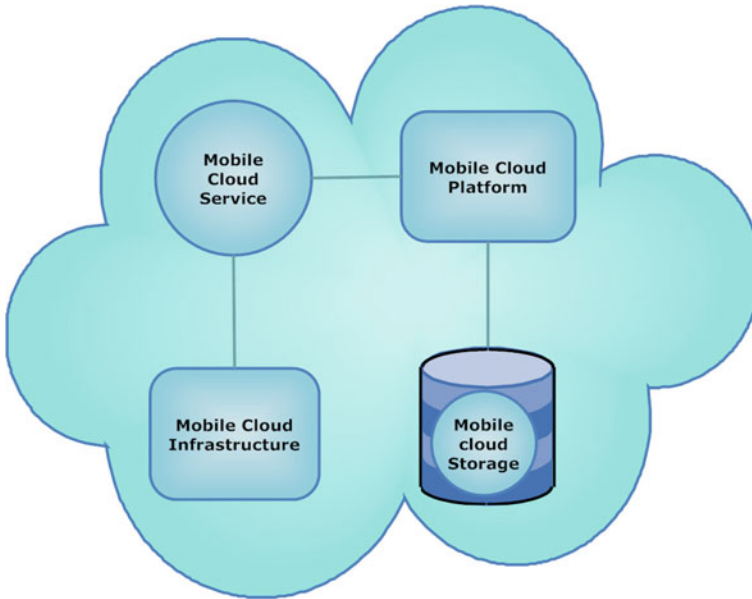


Fig. 11 Services provider

the memory limitations and executing power. Therefore, data are executing on the user opposite side and the habit of electronic smart machine for learning is rapid. The e-learning server will use mobile cloud computing, so that completely the essential resources will be adjusted as desired. Online learning system structures can take advantage from mobile cloud computing using:

Services: Implement online learning solution through internet given by the mobile cloud owner.

Operating System: Implement and improve an online learning solution based on the mobile cloud owner's progress edge.

Infrastructure: Take an online learning solution through internet on the mobile cloud owner's infrastructure (Fig. 11).

In e-learning, mega anxiety is associated to the information and data safety because both the software and data are placed on isolated servers that can crash or vanish without any additional warnings.

Even if it appears not very serviceable, the mobile cloud computing offers some main and huge securities, profits for commercial organization, and specifications that are using/emerging e-learning elucidations, as follows:

Central data storage: Losing a mobile cloud consumer is no longer time, a main occurrence while the core part of the solicitations and information is stored into the mobile cloud so a new consumer can be associated very fast. Envision what is happening nowadays if an electronic smart device that keeps the test paper is taken; observing the information use to develops simply in observation of a point

that single location should be managed, not more than hundreds of electronic smart devices using in an organization, institution, and academia for illustration. Also, the safety changes can be effortlessly verified and executed since the mobile cloud signifies a distinctive access point for entirely the consumers. Another significant benefit is related to budgets. If the e-learning facilities are used for a relative short time (several hours, a day, a week, a quarter, a semester an academic year), the savings are very significant.

Enhanced improbability: There are no chance for any concerned person to determine where is placed the device that stores specific wanted information (thesis, research results, project, research paper, etc.) or to determine which is the hardware element he wants to steal in demand to get an electronic property.

Virtualization: Creates probably the quick substitution of a compromised mobile cloud situated server without main prices and reimbursements. This is very simple to construct a group of simulated systems so the mobile cloud interruption is probable to be decreasing significantly.

7 Analysis of E-Learning System Through Electronic Smart Devices

We have gone through the survey in respective Computer Science Department in various universities of the KSA such as King Abdul Aziz University (KAUU), King Khalid University (KKU), and King Saud University (KSU). We are analyzing almost all faculty members and students who are using e-learning for all subjects in three famous universities in KSA, e-learning modules like black board for quiz, assignment, discussion, virtual class, attendance of student, exam, etc.

Black board was used through electronic smart devices by faculties as well as students. According to our analysis in computer science department subject name “Computer Skills”, subject code “101 HAL”, we have founded this comparative figure between mobile devices and non-mobile devices. The percentage of mobile consumer in this diagram (Fig. 12) shows that mobile devices are sufficient for e-learning.

We have survey of different organizations, institutions, and academics; and conclude that almost all institutions are using e-learning system at present through electronic smart devices. E-learning system is used through electronic smart devices, mobile cloud computing, and the growing rate increases day-by-day due to following facts.

- a. Minor rates
- b. Reduced software costs
- c. Improved performance
- d. Improved document format compatibility

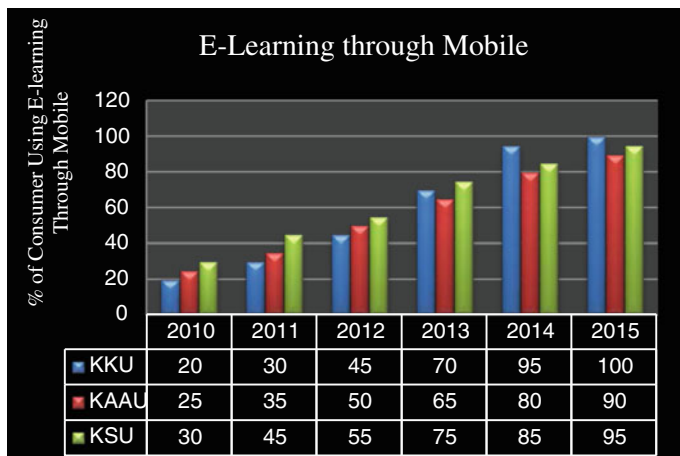


Fig. 12 Mobile versus non-mobile devices

- e. Increased data reliability
- f. Instant software updates
- g. Device independence
- h. Universal document access.

8 Conclusions

This paper discussed an architecture of mobile cloud computing (MCC) for e-learning, “is mobile cloud computing suitable for e-Learning” and most recent development of mobile cloud computing. Learning through tablet, iPad, and mobile are gaining popularity each and every day. It has changed the whole traditional education system and transforming it into a new paradigm shift in the field of education. The population of smartphone users is increasing day-by-day. Connectivity, multifunctionality, and portability are the major reasons for their popularity. Students, employees, and businesses are taking advantages of these features to get their degree or training hassle free. Cloud-based m-learning will help out students, teacher, working professionals, trainers, institutions, and mainly rural students will get an opportunity to acquire knowledge shared by the trainer across the globe. We believe that all over the world, government institutions will show their enthusiasm to adopt cloud-based mobile e- learning in the area of research, university, and school.

References

1. <http://cloudtimes.org/mobile-cloud>
2. <http://thoughtsoncloud.com/2013/06/mobile-cloud-computing>
3. Prasad, M.R., Gyani, J., Murti, P.R.K.: Mobile cloud computing: implications and challenges. *J. Inf. Eng. Appl.* **2**(7), (2012)
4. Bora, U.J., Ahmed, M.: E-learning using cloud computing. *Int. J. Sci. Mod. Eng. (IJISME)* **1**(2), (2013). ISSN:2319-6386
5. Pocatilu, P., Boja, C.: Quality characteristics and metrics related to M-learning process. *Amfiteatru Econ.* **Year XI**(26), (2009)

A Comparative Analysis of Different Social Network Parameters Derived from Facebook Profiles

Paramita Dey and Sarbani Roy

Abstract In social network analysis (SNA), using online social media, it is possible to collect large open source information and to analyze those data for knowing the characteristics of these networks. The main objective of this work is to study online social network parameters commonly used to explain social structures. In this paper, we have extracted data from the three real-time facebook accounts using Netvizz application. Gephi, a open source free software, is used for analysis and evaluation of these network parameters. This analysis shows some well-known network parameters like calculating clustering coefficient (CC) of clusters, group formation, finding node degree distribution (NDD), identifying influential node etc., which can be used for further feature extraction.

Keywords Social network analysis · Facebook profiles · Social network parameters · Netvizz · Gephi

1 Introduction

Recently, with the advent of online social networks (OSNs), there is a boom in social information [1]. In December 2014, the monthly active users of facebook, a popular social network service reach the value over 1.39 billion (source: www.facebook.com). But the main drawback of using these data is the huge size of the network. These data are mainly stored in the form of graph, analysis of which required large time and computation overhead. The main objective of this paper is to study the dynamics of different parameters involved in social network analytics. We have considered three real-time Facebook profiles. Some selected network

P. Dey

Department of Information Technology, GCECT, Kolkata, India

S. Roy (✉)

Department of Computer Science & Engineering, Jadavpur University, Kolkata, India
e-mail: sarbani.roy@cse.jdvu.ac.in

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_13

125

parameters are derived from these profiles for analysis. A comparative analysis is also presented here. Filtering approaches are used to reduce the data set but without effecting basic properties like node degree distribution (NDD) and clustering coefficient (CC).

Different social network parameters relevant to this paper are briefly discussed in Sect. 2. Results derived from real-time Facebook accounts are presented in Sect. 3 and a comparative analysis is done in perspective of social networking. In Sect. 4, the analysis is concluded and future scope of this work is reviewed.

2 Social Network Parameters

Social network analysis (SNA) is the qualitative and quantitative measuring technique to find the relationships among different social community like people, groups, organizations, etc. Some useful network parameters are discussed in the Table 1.

3 Analysis and Comparative Study

For the analysis purposes, three Facebook profiles were used to extract the data using Netvizz [2]. These three profiles are three disjoint data sets as they are not in the friend list of each other. From the perspective of social network, it is informative to study their interactions within the network and find the implication of these results [3]. A partial view of data table of profile 1 is shown in Fig. 1. After importing the data tables from Netvizz (this application is inbuilt within facebook profile), an initial hairball like network can be formed using Gephi [4], where each node represents an individual user and each edge represents the communication between them. Though from these networks, no direct information can be derived but an initial idea about the profile can be made. From the following profiles, it can be seen that the profile 1 consists of small number of edges which signifies the users within these group shares less communication within them. On the other hand, profile 2 and profile 3 have almost similar initial network due to the size of network in terms of nodes (profile 2: 215 #nodes; profile 3: 234 #nodes) and edges (profile 2: 3114 #edges; profile 3: 3310 #edges). A comparative view of these three initial networks is given in Fig. 2. Different network layouts are available in Gephi software. Among them, for our simulation, we have used Force Atlas 2 model [5] and Fruchterman-Reingold model [6]. Force Atlas 2 use different techniques such as degree-dependent repulsive force, Barnes Hut simulation, and adaptive temperatures for their simulation purposes. The main idea of simulation is that the nodes repulse and the edges attract. It is a continuous force directed layout. Network layouts using Force Atlas 2 with dissuade hubs mode of three profiles is shown in Fig. 3a. Dissuade hubs prefer authorities (nodes with high in-degree) in

Table 1 Network parameters

Social network parameters	Definition
Radius	Minimum path between any two nodes of the network, represented as $rad(G)$ [7]
Diameter	Maximum path between any two nodes of the network, represented as $diam(G)$ [7]
Shortest path	The minimum distance between any two nodes is the shortest path between two nodes [7]
Average path length	Arithmetic mean distance among all possible shortest path between any two nodes of the network signifies the rank of the network [7]
Node degree distribution	For a directed graph, <i>in degree</i> denotes the number of edges ending at that node and <i>out-degree</i> denotes the number of edges beginning at that node [8]
Rank	Rank counts the the number and quality of a links connected to that node [9]
Node degree centrality	Node degree centrality depends on the node degree distribution (i.e., in degree and out degree) of individual node. The node with maximum node degree represents the maximum centrality [10]
Betweenness Centrality	It denotes the number of shortest path passes through a node. The nodes with high betweenness centrality implies maximum connectedness in the network and vulnerability of the network is dependent on that nodes [10]
Closeness centrality	It denotes the average shortest path of a node with other nodes [10]
Eigen vector centrality	This centrality measures is based on Eigen vector matrices [10]
Community detection	It is the parameters of a network to classify the nodes into separate groups according to some properties. In social network analysis, <i>mutuality</i> , <i>reachability</i> , <i>vertex degree</i> and internal versus external <i>cohesion</i> are the four properties, which are used for Community detection [11, 12]
Clustering coefficient	The measurements of average distance of connected clustering nodes in a graph is called the clustering coefficient (CC) [10]

the center than hubs (nodes with higher out-degree). This system pushes the hubs in the periphery of the network. Force Atlas 2 network layout of these profiles with LinLog mode is shown in Fig. 3b. In this mode, LinLog energy model is used to make the cluster more dense but the convergence time with this model is much higher. According to Fruchterman-Reingold model [6], continuous network modeling was done depending on even distribution of the vertices in the frame, making edge lengths uniform and reflects inherent symmetry. Network layout of these profiles using Fruchterman-Reingold model are shown in Fig. 3c.

In profile 1, numbers of clusters are more but with less density which can be seen in two modes (dissuade hubs and LinLog) of Force Atlas 2. In profile 2, clusters are evenly distributed in nature. It is quite likely that there are better communications

1	Id	Label	sex	locale	agerank
2	10131316	Anthony Fernandes	male	en_US	69
3	5.4E+08	Atasi Deb Ray	female	en_US	68
4	5.6E+08	Raktim Ghosal	male	en_US	67
5	5.66E+08	Swaty Mitra	female	en_US	66
6	5.71E+08	Sohini Dasgupta	female	en_US	65
7	6.65E+08	Prabal Bagchi	male	en_US	64
8	6.71E+08	Urmi Choudhury	female	en_US	63
9	6.72E+08	Ram Rup Sarkar	male	en_US	62
10	6.96E+08	Madhumita Barua	female	en_US	61
11	7.01E+08	Alok Sharma	male	en_US	60
12	7.7E+08	Ranjan Som	male	en_US	59
13	8.22E+08	Sanjit Kumar Das		en_US	58
14	8.24E+08	Dipak Chatterjee	male	en_US	57
15	1.02E+09	Tapas RayMahapatra	male	en_US	56
16	1.06E+09	Gourab Ghosh	male	en_US	55
17	1.08E+09	Som Subhra Chakraborty	male	en_US	54
18	1.15E+09	Poulomi Chakraborty	female	en_US	53
19	1.18E+09	Samrat Laskar	male	en_US	52
20	1.27E+09	Sudip Chakraborty	male	en_US	51

Fig. 1 Partial data table derived from facebook using Netvizz

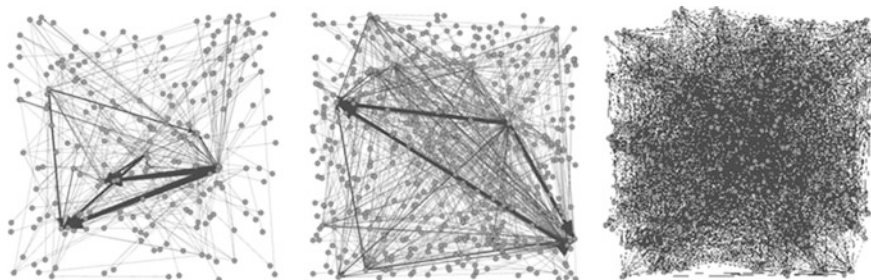


Fig. 2 Initial network formation from three facebook profiles

within the members of these profiles and the network is very stable. On the other hand, in profile 3, there are only two distinguishable clusters apart from very small clusters or isolated nodes. These clusters are very closely connected (as reflected from LinLog mode), which imply there is a regular communication between these dense clusters whereas it is less on the other nodes or clusters. In all three profiles, some of the links are very dense than the other which imply these nodes that share these links have a large betweenness, that means these nodes are very influential in that network. Node degree distribution (NDD) is an important networking property. NDD can be improved by reducing less important nodes and edges. This is the

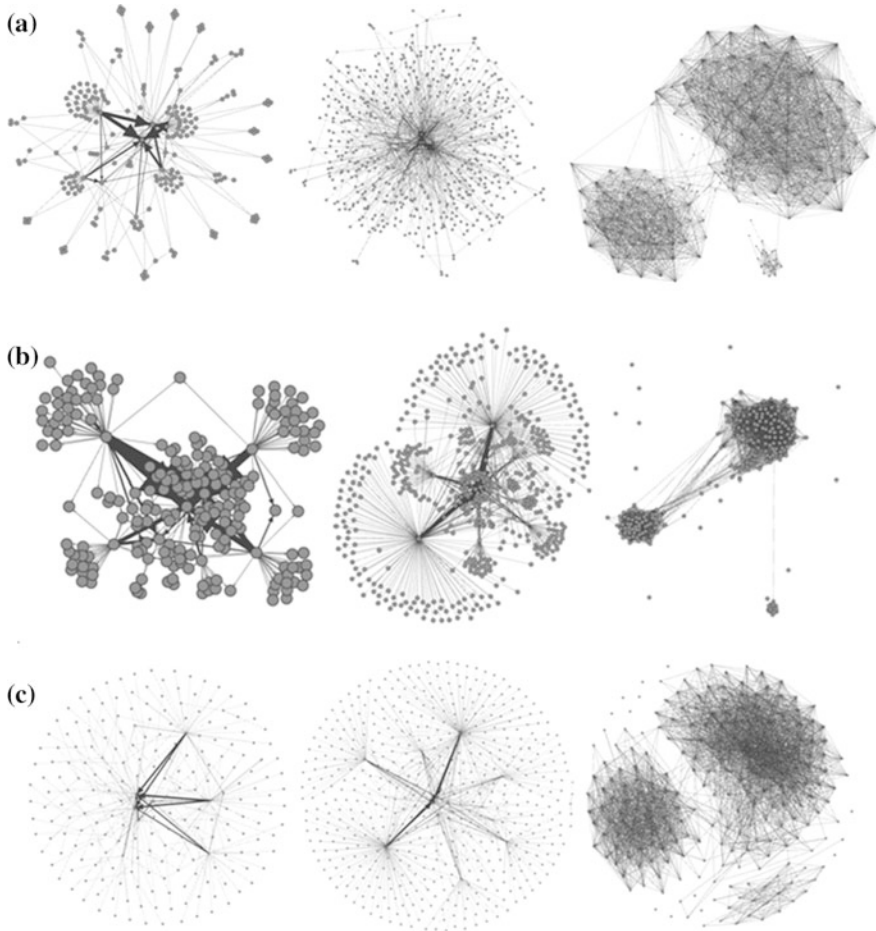


Fig. 3 Network formation of three profiles using different techniques. **a** Dissuade hubs mode of force Atlas 2. **b** LinLog mode of force Atlas 2. **c** Fruchterman-Reingold mode

basic process of filtering. Rank denotes the number of connections of each node. The nodes with maximum ranks can form the most significant network. In Fruchterman-Reingold distribution, the nodes with maximum nodes are connected with darker edges. Maximum eccentricity is the diameter and minimum eccentricity is the radius of a network. From the simulation result of eccentricity distribution and centrality distribution of the facebook profiles different parameters like radius, diameter, average path length, number of shortest paths can be calculated.

In social networking terms, centrality defines how fast the information can be spread. The distribution for closeness centrality distribution is shown in Fig. 4.

Clustering coefficient (CC) defines the centers of different communities. Community distribution based on the sex of these three profiles is shown in Fig. 5.

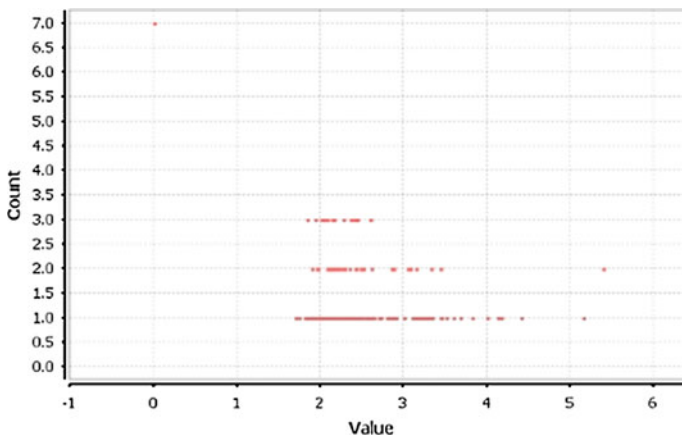


Fig. 4 Closeness centrality distribution of profile 2

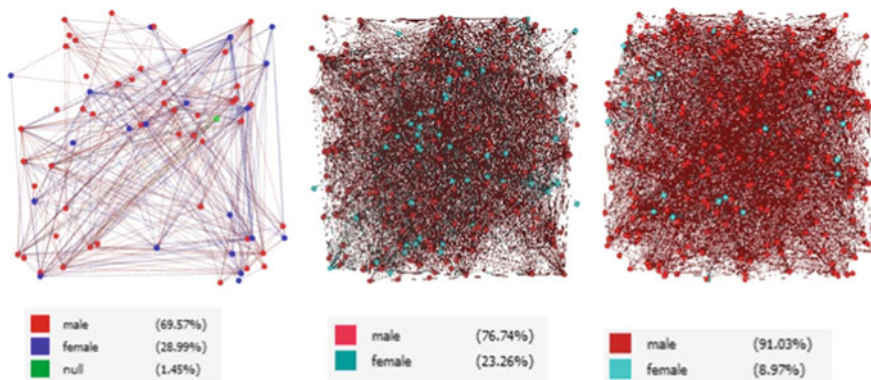


Fig. 5 Male–female percentage distribution of three profiles

It can also be done for other types of community distribution. Summarize all the properties derived from above distributions, it is shown in the Table 2. Though a number of nodes in all three profiles are almost same but in the first profile number of edges are significantly less. It signifies the cohesion between all the nodes are relatively less. It also reflects in density parameters which is only 1 % compare to 13 and 12 % of other two profiles. Diameter of profile 3 is maximum, refers that any two nodes of that profile attached through a distant communication. So, profile 3 plays a powerful role in the network. Average clustering coefficient in profile 2 and profile 3 is around 0.5 whereas for profile 1, it is less than 0.01, which signifies clusters of this profile are less connected compare to other two profiles. In this work, we have considered three profiles of male users. Here one observation is that female community in these profiles in respect to male community is very less.

Table 2 Comparative study of three facebook profiles

Matrices	Profile 1	Profile 2	Profile 3
Number of nodes	259	215	234
Number of edges	357	3114	3310
Diameter	6	7	9
Radius	1	0	0
Number of communities	21	17	13
Density	0.012	0.135	0.121
Average weighted degree	3.202	28.967	28.291
Average clustering coefficient	0	0.561	0.592
Average path length (after clustering)	3.492	2.538	2.794
Number of shortest path (after clustering)	47,326	43,056	49,062

4 Conclusion and Future Scope

The main motivation of this work is to extract the well-known parameters of the social network, make an analysis and make a comparative study between all the profiles. As we are using open-source tools like GEPHI software, which can use external modules like JAVA net-beans as add-on, these data can be classified further for more advanced feature extraction. Extracted information can be used for designing graph sampling algorithms and game theory-based social network designing. In future, we will try to design an interactive model for finding a relation between an individual with population to measure the influence of that person in the society.

Acknowledgments We would like to thank Subrata Dey, Aniket Sinha and Dr. Krishnendu Dutta for sharing their Facebook profiles with us for data collection purposes.

References

1. Benevenuto, F., Rodrigues, F., Cha, M., Almeida.: Characterizing user behavior in online social networks. In: Proceedings of ACM IMC (2009)
2. Rieder, B.: Studying facebook via data extraction: the Netvizz application. In: WebSci, Paris, France, ACM 978-1-4503-1889-1, May 24 2013
3. Wilson, R.E., Gosling, S.D., Graham, L.T.: A review of facebook research in the social sciences perspectives. Psychol. Sci. **7**(3), 203–220 (2012)
4. Gephi Official Website: <https://gephi.github.io>
5. Jacomy, M., Heymann, S., Venturini, T., Bastian, M.: ForceAtlas2, a continuous graph layout algorithm for handy network visualization. PloS One **9**(6), (2010)
6. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. Softw: Pract. Expert **21**(11), 1129–1164 (1991)
7. Wayne, G., Oellermann, R.O.: Distance in Graphs, Structural Analysis of Complex Networks, pp. 49–72. Springer (2011)

8. Wang, T., Chen, Y., Zhang, Xu, T., Jin, L., Hui, P., Deng, B., Li, X.: Understanding graph sampling algorithms for social network analysis. In: Simplex, IEEE ICDCS, Minneapolis, USA, pp. 123–128 June 20–24 2011
9. Kleinberg, J.: Authoritative sources in a hyperlinked environment. In: Proceedings of 9th ACM-SIAM Symposium on Discrete Algorithms (1998)
10. Freeman, L.C.: Centrality in social networks: conceptual clarification. *Soc. Netw.* **1**, 215–239 (1978/1979)
11. Ronald, S.B.: Social contagion and innovation: Cohesion versus structural equivalence. *Am. J. Sociol.* **92**(6), 1287–1335 (1987)
12. Fortunato, S., Castellano, C.: Community structure in graphs. *Phys. Rep.* (2007)

A Fast and Hardware-Efficient Visual Cryptography Scheme for Images

Dipesh Vaya and Sarika Khandelwal

Abstract Shamir's encryption method is to share a secret image by n number of shadow images and then r shadow images can be used to obtain the original secret image. In proposed method, the size of secret image is greater than the size of the shadow image used for the encryption. Such shadow image is beneficial to process in image hiding, transmission, or storage. For this purpose Shamir's encryption technique is used in this proposed work. In Shamir's encryption technique, author used the equation to encrypt the data into multiple parts. In the equation, time required for the multiplication and division of the input components of the data is more than addition and subtraction. Hence, the encryption time is little bit high. To reduce this encryption and decryption time, Shamir's equation is modified by converting all the multiplication part into addition and division parts into subtraction in this proposed work.

Keywords Secret sharing · Lossless reveal · Shadow images · Shamir's encryption technique

1 Introduction

Secret images and texts often exist in the military or commercial application. About the storage of the secret images or data, the security is one of the most important concerns. In last few years, different methods were proposed by different researchers to improve the security of data or images; examples include image hiding and watermarking. But one of the major drawback of these methods are, the

D. Vaya (✉) · S. Khandelwal
Department of Computer Science & Engineering, Geetanjali Institute
of Technical Studies, Udaipur, India
e-mail: dipesh.vaya88@gmail.com

S. Khandelwal
e-mail: sarikakhandelwal@gmail.com

secret data is transfer through single information carrier and if this carrier is lost or crippled during transmission, the secret data cannot be revealed completely. In order to overcome this problem, multiple duplicates of the information carrier can be used, but because of this the security danger of exposure will also increase. Secret sharing method can be used to solve this problem.

Blakley [1] and Shamir [2] proposed the concept of secret sharing independently. This proposed method was known as the $(k; n)$ threshold scheme. In most of the studies, main focus is on the security of the secret key of the encryption data [3–5]. Number of bytes used in the digital image are generally very large and the gray value of the digital image is bounded (0–255). Hence when the image is used as the secret data in the threshold scheme directly, this will waste memory space and more time required for encryption of image and obtaining original image from the shadow image. Therefore, need of the specific method for secret image sharing is obtained. A new secret image sharing method derived using the $(k; n)$ threshold scheme proposed by Shamir is proposed in this paper. The time required to convert the shadow images is much less than that of the secret image in our method. We will require that

- The n shadow images are generated using secret image.
- To reconstruct the image any k or more than k shadow images can be used.
- Any $k - 1$ or less shadow images cannot get sufficient information to reveal the secret image, i.e., secret image cannot be obtain from any $k - 1$ or less shadow images.

2 Shamir's (k, n) Threshold Scheme

This paper roughly introduces his scheme first, because proposed method is based on the (k, n) threshold scheme proposed by Shamir. If the secret data D is divided into n shadows, i.e., (D_1, D_2, \dots, D_n) , and we want original secret data cannot be retrieve from the encrypted data without k or more shadows. Without loss of generality, let us take any number as the D . Pick randomly a prime number p and a $k - 1$ degree polynomial to split D into n shadows [1, 6].

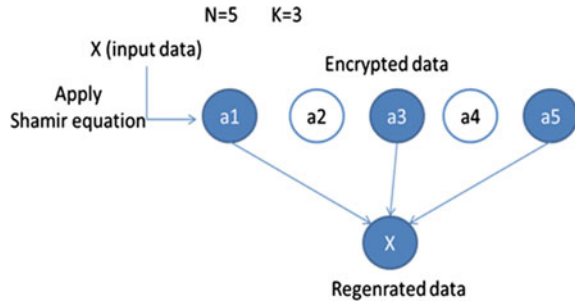
$$q(x) = (a_0 + a_1x + \dots + a_{k-1}x_{k-1}) \bmod p, \quad (1)$$

In which $a_0 = D$ and then evaluate

$$D_1 = q(1), \dots, D_i = q(i), \dots, D_n = q(n) \quad (2)$$

Note that each D_i is a shadow. The coefficients a_1, \dots, a_{k-1} in $q(x)$ are randomly chosen from the integers in $0 - (p - 1)$. Given any k pairs of these n pairs $\{i, D_i\}_i^n = 1$ we can find the coefficients $a_0 - a_{k-1}$ of $q(x)$ by the Lagrange's

Fig. 1 Scenario of Shamir threshold equation



interpolation, and hence the secret data $D = a_0$ is also revealed. Figure 1 shows the overall functionality of the Shamir threshold equation where n is number of encryption parts and k is number of reconstruction parts.

In image sharing, to use Shamir’s (k, n) threshold scheme directly, the gray value of the first pixel is taken as a_0 , then we obtain the corresponding output $q(1) - q(n)$; Once the output is obtained then the value of a_0 is replaced by second pixel’s gray value, and this process is repeated until all the pixels of secret image are processed.

If we want to divide D secret image into n number of shadow images, i.e., (D_1, \dots, D_n) , and r or more shadow images are used to reveal the secret image from the information carrier. We generate the $r - 1$ degree polynomial, by letting the r coefficients, will be the gray values of r pixels in our proposed method. We use no random coefficient in proposed method where in Shamir’s method he uses random coefficient, this is the main difference between our and Shamir’s method. The gray value of the image pixel is between the 0 and 255, because of this Shamir let the prime number p be 251 which is the greatest prime number not larger than 255. We must truncate all the gray value to 250 from 251 to 255 of the secret image so that all the gray values are in the range of 0–250.

First the image is divided into several sections for the encryption process. Every section of the secret image has k number of pixels, and every pixel of the image belongs to one and only one section of the secret image. For every section of j define the following $k - 1$ degree polynomial

$$q(j) = (a_0 + a_1x + \dots + a_{k-1}x_{k-1}) \bmod 251, \tag{3}$$

where a_0, \dots, a_{k-1} are the k pixels of the section, and then evaluate

$$q_j(1), q_j(2), \dots, q_j(n). \tag{4}$$

Output obtained from the equations is the n output pixels $q_j(1) - q_j(n)$ of the j section are assigned sequentially to the n shadow images. Every shadow image receives one generated pixel of the secret image for each section (of k pixel). Secret image sharing method proposed here follows the following steps:

- Reduce all the gray values of the secret image larger than 250 to in the range of 0–250.
- To generate a permutation sequence to permute the pixels of the secret image use a secret key.
- To form a section take k not-shared-yet pixel of the permuted image sequentially.
- To generate n pixels for the n shadow images, use the section formed in Step 3 and Eqs. (3) and (4).
- Until all pixels of the permuted Images are processed repeat Steps 3 and 4 constantly.

The reveal phase using any k shadow images are as follows:

- From every k shadow images take the first non-used pixel.
- Improvise every k pixels and apply the Lagrange's interpolation to solve for the coefficients $a_0 - a_{k-1}$ in Eq. (3). The coefficients $a_0 - a_{k-1}$ are then the corresponding k pixel values of the permuted image.
- Until all pixels of the k shadow images are processed repeat Steps 1 and 2.
- To get the secret image, apply inverse permutation operation to the permuted image.

Figure 2 shows the experimental results of the Shamir's secret sharing encryption technique taken from paper presented by Thien and Line [7].

The value of $n = 4$ used in the encryption technique and value of $k = 2$. So after encryption of the original image using Shamir's secret sharing encryption technique four shadow images are generated and to obtain the original image from the encrypted shadow image, two shadow images are combined because value of $k = 2$ is used in the Shamir's encryption equation.

3 Related Work

El-Tigani et al. [8] introduced a new (x, n) threshold secret sharing method for grayscale images. Here quadratic residues method is incorporated to encrypt the image and after that Shamir's methodology is used to create the shadow images. The major concern of author is to allow any number of participants for sharing secret image in any communication channel. As compared to Chen-Chang's technique [9], this work provides more flexibility and security for image sharing.

Patil et al [10] proposed a method of encryption to construct colored EVC scheme to improve visual image quality with error diffusion and VIP synchronization. For retaining the original values of pixel, the pixels positions are synchronized that contain visual information of original images over the color channels.

Kalai Selvi et al [11] proposed new algorithm to generate a random polynomial of the secret image for the sharing purpose. This proposed process offers enhanced technique of encryption. In this algorithm, the image is encrypted after several

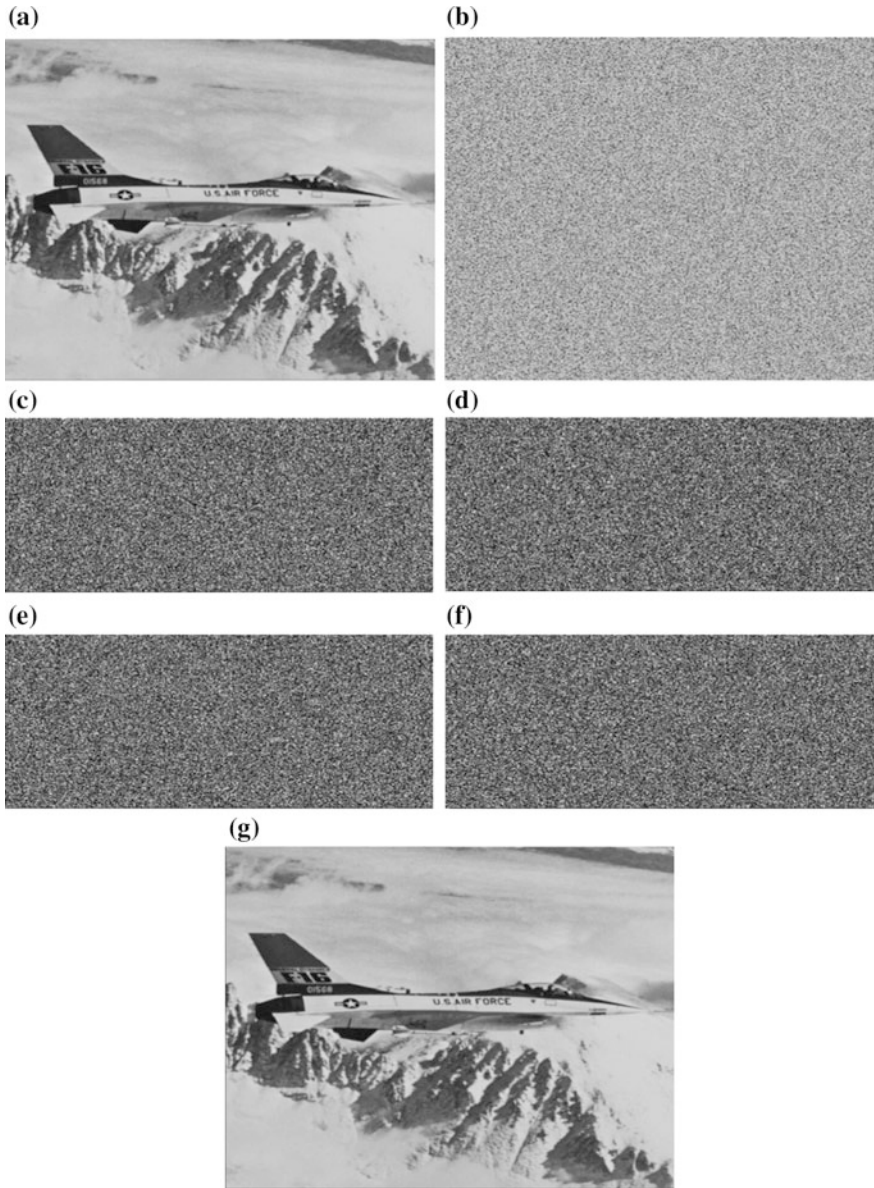


Fig. 2 a A 512×512 secret image; b the permuted image of (a); c–f the four shadow images; g the revealed image

numbers of rounds, which makes the computation more complex. As Compared to the encryption schemes based on the secret sharing, image size is far bigger than the secret share's size.

4 Proposed Methodology

The proposed work aims to reduce the time required for encryption and regeneration of image so that the process of sharing and regeneration is fast. For image sharing, Shamir's secret sharing encryption technique is applied with enhanced features.

This paper proposes a method for enhancement of Shamir's secret sharing encryption algorithm where we convert all higher level mathematical operation of Shamir's encryption to lower level operations i.e., this work enhances Shamir's work by implementing all multiplication operations into addition and all division operations by subtraction.

In Shamir's encryption equation, array multiplier is used for encrypting the image. A carry save array multiplier is also known as a simple parallel array multiplier. But it can be used to perform only over signed bits due to restrictions. To remove the logic registers from the array multiplier, array of AND gates are used in the structure and adders in iteratively arranged manner [12]. In original Shamir encryption algorithm, image pixels are stored inside array and then multiplication is applied on the different arrays to give the output array.

Data 1 = array of pixels

Data 2 = array of pixels

Data 1 \times Data 2 = Output Array

More processes are required to implement the multiplication array, because more adders are required to execute the multiplication and it requires more time and energy to process the encryption. For any array multiplier area, computation speed and dissipation of power are the important criteria. To improve these parameters, array adders are used in the proposed method. Parallel prefix adder can perform operations in higher speed. It is most flexible and commonly used adder for binary addition. For optimizing the fan out, area, inter connection count, and logic complexity, different researchers have been proposed number of architectures for designing parallel prefix adder. Lots of works so far have been done in past and considerably various architectures have been proposed for designing an adder. When operations required to be occur on high speed, parallel prefix adders are required to be in tree like structure [13]. To implement array adders in the Shamir's encryption equation, higher order operations are converted to the lower order operations. The original Shamir's encryption equation is

$$x^{(1:k-1)*r} + s \quad (5)$$

By converting the higher operation in the equation by lower operation, we obtained the equation as

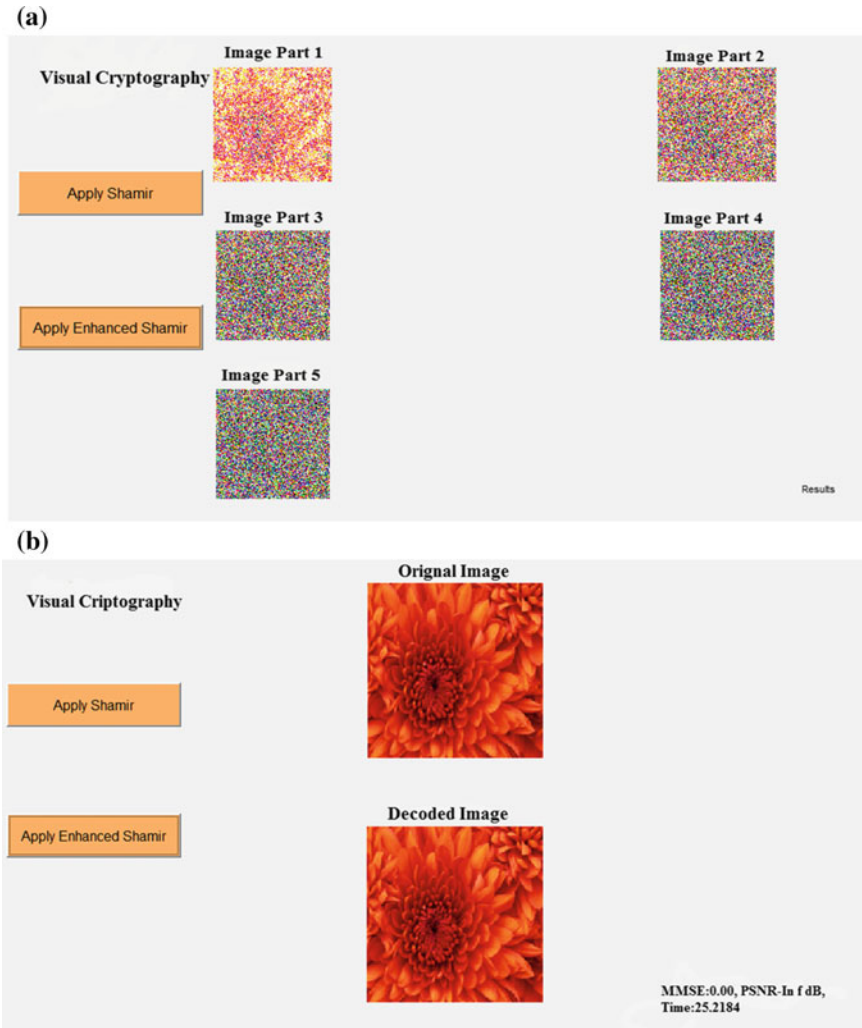


Fig. 3 **a** Output after applying proposed encryption technique; **b** output after applying proposed decryption technique on an image

$$x \sum_{i=1}^r (1:k-1) + s \tag{6}$$

5 Experimental Result

MATLAB software is used to implement the proposed encryption technique on images. The experimental results are shown below:

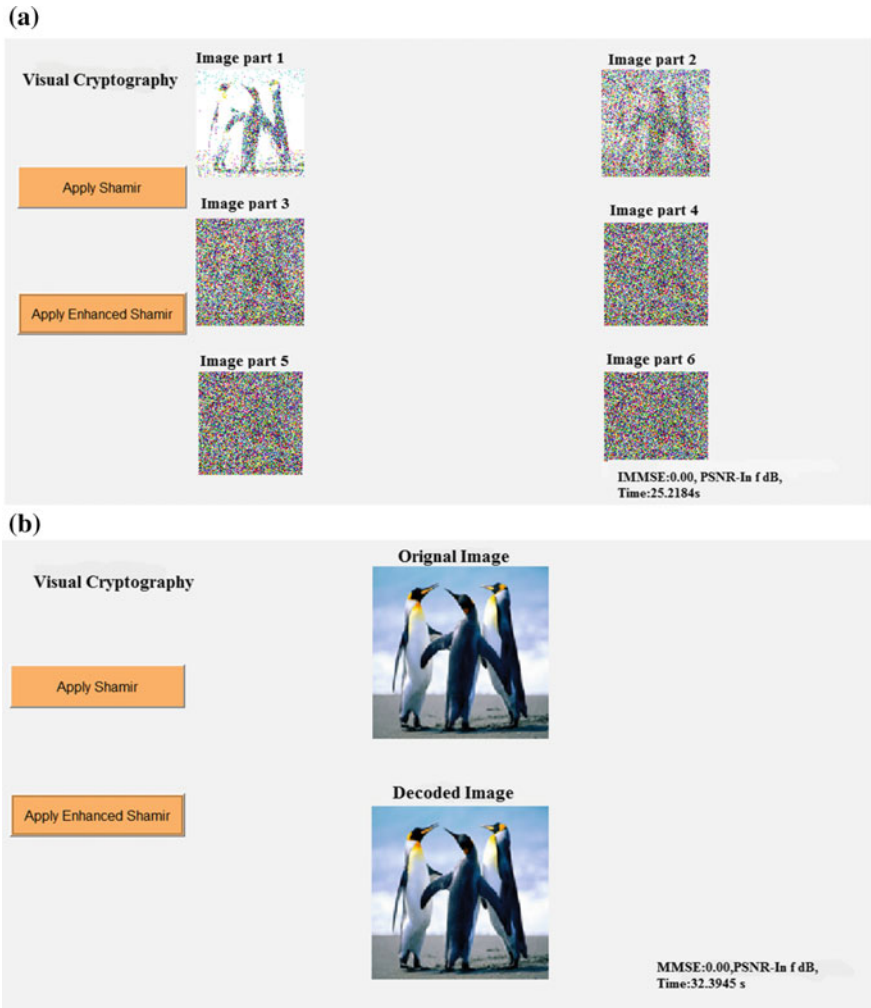


Fig. 4 **a** Output after applying proposed encryption technique; **b** output after applying proposed decryption technique on an another image

Figures 3a, b and 4a, b show the experimental results of proposed encryption and decryption techniques output on images of flower and penguins respectively.

Table 1 shows the comparison of time required to encrypt and decrypt the image using proposed technique and Shamir’s technique and Fig. 5 shows the graphical representation of the comparison.

Table 1 Result of proposed and simple Shamir’s techniques

Image size	Time (s) (Shamir)	Time (s) (Enhanced Shamir)
128 × 128	47.25	25.3
256 × 256	75.36	38.45
512 × 512	102.97	47.44
1024 × 1024	165.89	74.56

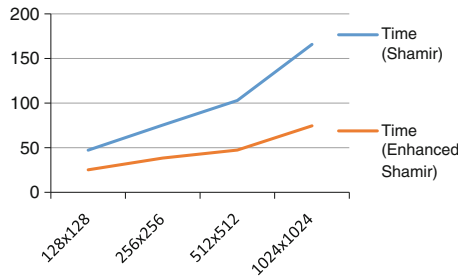


Fig. 5 Graphical representation of the comparison between Shamir and Enhanced Shamir algorithms

6 Conclusion

The proposed method works effectively and efficiently over time optimization for the image transmission. In this paper, we improved the Shamir’s encryption and decryption equation for the better utilization of the time. As shown in the Table 1, we are able to reduce the time of encryption and decryption of the image by 45–60 %. For reducing the time, we have converted all the multiplication into addition and all the division into subtraction in the original Shamir’s equation to obtain enhanced Shamir’s equation. In this paper, Shamir’s encryption technique is used for the encryption; in future any other advance encryption technique can be used to improve the efficiency of the proposed system.

References

1. Blakley, G.R.: Safeguarding cryptographic keys. In: Proceedings AFIPS 1979 National Computer Conference, vol. 48, pp. 313–7. New York, USA, 4–7 June 1979
2. Shamir, A.: How to share a secret. *Commun. ACM* **22**(11), 612–613 (1979)
3. Ahlswede, R., Csiszár, I.: Common randomness in information theory and cryptography-part I: secret sharing. *IEEE Trans. Inf. Theory* **39**(4), 1121–32 (1993)
4. Beimel, A., Chor, B.: Secret sharing with public reconstruction. *IEEE Trans. Inf. Theory* **44** (5), 1887–1896 (1998)

5. Beimel, A., Chor, B.: Universally ideal secret-sharing schemes. *IEEE Trans. Inf. Theory* **40**(3), 786–794 (1994)
6. Stinson, D.R.: Decomposition constructions for secret-sharing schemes. *IEEE Trans. Inf. Theory* **40**(1), 118–124 (1994)
7. Thien, C.-C., Jin, J.-C.: Secret image sharing. *Comput. Graph.* **26**, 765–770 (2002)
8. Abdelsatir, E.-T.B., Salahaldeen, S., Omar, H., Hashim, A.: A novel (K, N) secret sharing scheme from quadratic residues for grayscale images” *Int. J. Netw. Secur. Appl. (IJNSA)* **6**(4), 65–72 (2014)
9. Chen, C.-C., Chang, C.-C.: Secret image sharing using quadratic residues. *Intell. Inf. Hiding Multimedia Sig. Process. (IIHMSP)* **1**, 515–518 (2007)
10. Patil, S., Rao, J.: Extended visual cryptography for color shares using random number generators. *Int. J. Adv. Res. Comput. Commun. Eng.* **1**(6) (2012)
11. Kalai Selvi, A., Mohamed Sathik, M.: Secret sharing scheme for image encryption based on primitive root theorem. *Int. J. Control Autom* **5**(3), 37 (2012)
12. Thakur, M., Ashraf, J.: Design of braun multiplier with Kogge Stone Adder & it’s implementation on FPGA. *Int. J. Sci. Eng. Res.* **3**(10) (2012)
13. Seng, Y.K. Roy, K.: Low Voltage, Low Power VLSI Subsystems. TMC (2009)

Energy-Efficient Modified Bellman Ford Algorithm for Grid and Random Network Topologies

Rama Devi Boddu, K. Kishan Rao and M. Asha Rani

Abstract Energy-efficient routing techniques are required for mobile ad hoc Networks (MANETs) to improve the lifetime of the network. The lifetime of the network depends on the battery capacity of nodes. The link failure due to the battery discharge of node can be avoided by considering the nodes having good residual energy (RE) with less change in their battery capacity. In this paper, the Bellman–Ford algorithm (BFA) is considered to find the shortest path for routing. Bellman–Ford algorithm is modified and the nodes whose change in battery capacity is less than a predefined threshold value are considered for routing to avoid the link failures and to enhance the lifetime of the network. In the proposed modified Bellman–Ford algorithm (MBFA), residual energy (RE) is considered as a metric to find the shortest path. IEEE 802.11 a/g standards using orthogonal frequency division multiplexing (OFDM) are considered for simulation. Energy consumed by the radio transceiver, processor, losses in the battery, and DC–DC converter are taken into consideration for energy calculation. The performance of BFA and MBFA for the grid and random network topologies is simulated by considering the network with multiple sources and destinations are compared with and without mobility by assuming various densities, i.e., 15, 30, 45, and 60. The mobility of the node increases the loss of orthogonality among the OFDM subcarriers and results inter carrier interference (ICI). The effect of mobility and network size on

R.D. Boddu (✉)

Department of E.C.E, Kakatiya Institute of Technology and Science,
Warangal, Telangana, India
e-mail: ramadevikitsw@gmail.com

K. Kishan Rao

Department of E.C.E, Vaagdevi College of Engineering, Warangal,
Telangana, India
e-mail: prof_kkr@rediffmail.com

M. Asha Rani

Department of E.C.E, Jawaharlal Nehru Technological University,
Hyderabad, Telangana, India
e-mail: ashajntu1@yahoo.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_15

throughput, delay, jitters for the grid, and random network topologies using BFA and MBFA are compared. Simulation results show that the performance of proposed MBFA is better compared to BFA.

Keywords Bellman ford algorithm · Inter carrier interference · Modified bellman ford algorithm · OFDM · Residual energy

1 Introduction

The increase in demand and usage of mobile ad hoc networks in the present generation has to concentrate on the battery power usage of nodes in the ad hoc networks. For the efficient usage of battery power, we require an energy-efficient routing algorithm to improve the lifetime of the network. Minimum total power routing (MTPR) use distributed Bellman–Ford algorithm (BFA) and the path with a minimum total power cost from the source node to the current node is selected. In the minimum battery cost routing (MBCR) [1] route with maximum remaining battery capacity is selected to prevent overuse of nodes. Route in min-max battery cost routing (MMBCR) selects the route using MBCR by considering the battery capacity above the threshold. Toh et al. [2] proposed power aware conditional max-min battery capacity routing for a wireless ad hoc networks. A modified version of BFA was proposed to find out the shortest path using maximum residual energy (RE) [3] to improve the network lifetime. Neyre Tekbiyik et al. [4] presented a review on energy-efficient routing, link cost based, and an energy-efficient shortest path routing algorithms. Mostafa Dehghan et al. [5] developed an opportunistic routing algorithm and computed the optimum path using the stochastic version of BFA. Efficient Power Routing [6] considered energy consumption metric and battery discharging loss.

Node placement [7] affects the node density, network topology, routing, network delay, node energy and lifetime, and transmission range. Node placement can be classified into: (i) Deterministic (ii) Semi-deterministic (iii) Non-deterministic types [8]. Energy considerations based on node placement and various routing protocols for the grid topology with varying network size were investigated [9].

In this paper, the effect of node deployment with mobility and varying network size is considered (i) to propose energy-efficient routing algorithm and to investigate the effect of node deployment on its performance (ii) to analyze the performance of the grid and random networks with and without mobility with varying network size.

This paper is organized as follows: Sect. 2 describes energy consumption of OFDM transceiver; Sect. 3 describes proposed modified Bellman–Ford algorithm and energy calculation. Numerical results are analyzed in Sect. 4. Finally, conclusions are drawn in Sect. 5.

2 Energy Consumption in IEEE 802.11a/g OFDM Transceiver

The optimum energy consumed by the node can be calculated by considering all the blocks of transceiver circuit as shown in Fig. 1. Let us assume P_{PA} , P_{ct} , P_{ADC} , P_{mix} , P_{LNA} , P_{IFA} , P_{DAC} , P_{syn} , P_{filt} and P_{filr} represents power consumed by power amplifier (PA), circuit, analog-to-digital converter (ADC), mixer, low-noise amplifier (LNA), intermediate frequency amplifier (IFA), digital-to-analog converter (DAC), frequency synthesizer, active filter at transmitter, and receiver, respectively.

Power consumed by the node in active mode can be written as

$$P_{on} = P_{PA} + P_{ct} \quad (1)$$

Power consumed by PA is a function of transmission power P_{out} and is given by

$$P_{PA} = (1 + \alpha')P_{out} = \frac{\xi}{\eta}P_{out}, \quad (2)$$

where $\alpha' = \frac{\xi}{\eta} - 1$. $\eta = \frac{\eta_{max}}{\xi}$ is drain efficiency of the power amplifier, η_{max} = maximum drain efficiency, ξ = peak to average ratio (PAR).

Transmission power is a function of distance and can be written as [10]

$$P_{out} = \bar{E}_b R_b \left[\frac{(4\pi)^2 d^\alpha}{G_r G_t \lambda^2} \right] M_l N_f, \quad (3)$$

where G_t , G_r represents gain of transmitting and receiving antennas, respectively; d is the distance between transmitter and receiver; R_b , λ , M_l , α , N_f , \bar{E}_b represents bit rate, carrier wavelength, link margin, path loss coefficient, noise figure, and energy required per bit per hop for a given symbol error rate, respectively.

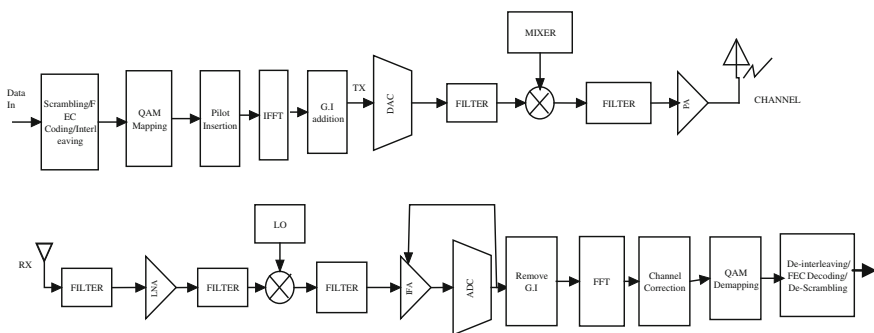


Fig. 1 OFDM transmitter and receiver with transceiver circuit

The total power consumed by the circuit (P_{ct}) is the sum of circuit power consumed at transmitting node (P_{ctx}) and receiving node (P_{ctr}).

$$P_{ctx} = P_{DAC} + P_{mix} + P_{filt} + P_{syn} \quad (4)$$

$$P_{ctr} = P_{LNA} + P_{mix} + P_{IFA} + P_{ADC} + P_{syn} + P_{filr} \quad (5)$$

3 Proposed Modified Bellman Ford Algorithm

The wireless ad hoc network consists of n nodes represented by a node set $N = (n_1, n_2, n_3, \dots, n_n)$. The network was represented by a graph $G = (N, L)$, L represents the link set, C_{ij} represents the cost of direct edge $e_{i,j}$ between (n_i, n_j) and D_{n_i} represents the minimum cost from n_i to sink. BFA computes the shortest path using a distance vector routing protocol. Route update messages will be sent periodically to update the routing table. Shortest path using the BFA was evaluated [11] using the following iterative steps:

- Initialize with $D_1^h = 0$ for all h , $D_i^0 = \infty$ for all $i = 2, 3, \dots, n$
- Calculate $D_i^{h+1} = \min_{j \in N} [C_{ij} + D_j^h]$ for all $i \neq 1$

Let $h = h + 1$

- Bellman–Ford Algorithm terminates when $D_i^h = D_i^{h-1}$.

Modified Bellman–Ford Algorithm (MFBA): In MFBA, residual energy (RE) is considered as a metric in the routing table of BFA in addition to the destination sequence which was the standard metric. Initially, all nodes configured with BFA and broadcast the route advertisement packet to find the shortest distance to reach other node. Meanwhile, all nodes update their energies in the routing table as a route metric by communicating with the physical layer. When the new route advertisement packet arrives, the energy in the incoming route advertisement packet is compared with the energy of the node in the route table for which the packet is being advertised. The route table is updated when its energy is changed by the certain offset. To eliminate the nodes with the leakage battery capacity, always it checks the change in the battery capacity of a node is greater than the threshold or not. When a source node wants to transmit the packet, it finds the optimal path toward the destination by consulting the routing table. During the phase of packet forwarding, we can calculate the total energy consumed by the route has a sum of all hops, individual node energies present in the route, efficiency losses in the battery, energy consumed by the processor, and DC–DC converter are also to be considered. All the nodes in the route of MBFA are aware of the total energy utilized by that route. During the simulation, the nodes gradually get discharged and update their energies periodically with the help of route update packets and include in the routing table. Finally, all the nodes are aware of the energies of all other nodes.

Energy Calculation for MBFA: Let $E_{\text{tras_amp}}d_{k,k+1}^z$ represents transmission energy required to transmit a distance of $d_{k,k+1}$ between the nodes n_k to n_{k+1} ; $E_{\text{trans_Tx}}, E_{\text{trans_Rx}}, E_{\text{Idle}}$ represents energy consumed by the transceiver circuit during transmission, reception, Idle state, respectively. Node initial energy is represented by E_k^i . The energy consumed by the processor, the efficiency losses in the battery and DC–DC converter are represented by $E_{\text{CPU}}, E_{\text{Bat}}, E_{\text{DC}}$, respectively.

The energy consumed to transmit a data packet of L bits from node n_k to n_{k+1} can be given by

$$\begin{aligned} E_k^{Tx} &= E_{\text{trans_Tx}} + E_{\text{tras_amp}}d_{k,k+1}^z \\ &= P_{\text{ctx}}T_{\text{on}} + \frac{\xi}{\eta}\bar{E}_bR_b \left[\frac{(4\pi)^2}{G_rG_t\lambda^2} \right] M_tN_f T_{\text{on}}d_{k,k+1}^z \end{aligned} \quad (6)$$

The energy consumed by the node n_k to receive a data packet of L bits is given by

$$E_k^{Rx} = E_{\text{trans_Rx}} = P_{\text{ctx}}T_{\text{on}} \quad (7)$$

After transmitting or receiving 1-bit data, the RE of n_k can be given by

$$E_k^R = \begin{cases} E_k^i - E_k^{Tx} & \text{for } L\text{-bit transmission} \\ E_k^i - E_k^{Rx} & \text{for receiving } L\text{-bit} \end{cases} \quad (8)$$

In general, the energy consumed by the system per cycle is given by [12] the energy consumed by the radio transceivers, protocol processor, the efficiency losses in the battery, and DC–DC converter which regulates voltage for different components.

MBFA calculates the total energy consumed by the route during packet forwarding. For MBFA and total energy consumed for Z hops can be written as

$$E_T = \sum_{k=0}^{Z-1} [E_k^{Tx} + E_{k+1}^{Rx}] + \sum_{k=0}^Z [E_{\text{CPU}} + E_{\text{Bat}} + E_{\text{DC}} + E_{\text{Idle}}] \quad (9)$$

4 Numerical Results

Two network topologies of Grid and random for Varying network Traffic with different node densities 15, 30, 45, 60 over an area 1000 m^2 are configured on the QualNet Simulator using BFA and MBFA routing algorithms. For simulation, IEEE 802.11a/g standards are considered as shown in the Table 1. A Bit Error Rate (BER) reception model and 64-QAM modulation scheme with simulation time of 3000 s and Group Mobility, Random Mobility models are considered for mobility.

Table 1 OFDM simulation parameters for IEEE 802.11a/g

Parameter	Value
BW : channel bandwidth	20 MHz
N_{FFT} : Size of FFT	64
N_{SD} : Number of data subcarriers	48
N_{SP} : Number of pilot subcarriers	4
$N_{ST}(= N_{SD} + N_{SP})$: Total number of subcarriers	52
$\Delta_F = \frac{F_s}{N_{FFT}}$: Subcarrier frequency spacing	0.3125 MHz (=20 MHz/64)
$T_b = \frac{1}{\Delta_F}$: Useful symbol time	3.2 μ s
T_g : Cyclic prefix (CP) time	0.8 μ s
$\beta = \frac{T_g}{T_b}$: Cyclic prefix ration	$\frac{1}{4}$
$T_s = T_b + T_g$: Symbol interval	4 μ s
$T_{Preamble}$: PLCP preamble duration	16 μ s ($T_{Short} + T_{Long}$)

Table 2 Simulation parameters generic radio energy model

Parameter name	Symbol	Value (mW)
Transmit power	P_{Tx}	100
Receive power	P_{Rx}	130
Mixer power consumption	P_{mix}	30.3
Filter power consumption	$P_{filt} = P_{flr}$	2.5
Synthesizer power consumption	P_{syn}	50
LNA power consumption	P_{LNA}	20
DAC power consumption	P_{DAC}	15.4
ADC power consumption	P_{ADC}	14
IFA power consumption	P_{IFA}	3

The generic radio energy model [12] in QualNet Simulator is considered for energy calculation as shown in the Table 2. The various parameters used for simulation are 120 mAh battery initial capacity, 3.37 V battery voltage, 0.000002 change in node capacity as a threshold, 217–402 m radio range, 2.3 path loss coefficient (2-ray model), a data packet of 512 bytes, and 2.4 GHz radio frequency.

The residual battery capacity of BFA and MBFA using grid and random topologies with and without mobility is shown in the Table 3. Multiple sources and multiple destinations are considered with different network sizes with node densities of 5, 30, 45, and 60. The network with varying network traffic considered for the simulation. The total packets received verses number of nodes is shown in Fig. 2. The simulation results show that the number of packets received in MBFA is better compared to BFA. As the mobility of the node increases, the delay spread increases which results in the loss of orthogonality among the OFDM subcarriers and cause intercarrier interference (ICI) due to which the total number packets

Table 3 Residual battery capacity (mAh)

No. of nodes	Without mobility				With mobility			
	Grid topology		Random topology		Grid topology		Random topology	
	BFA	MBFA	BFA	MBFA	BFA	MBFA	BFA	MBFA
15	1567.78	1775.8	1567.59	1775.6	1567.25	1776.1	1568.07	1776.1
30	3150.15	3551.2	3135.7	3549.6	3150.43	3553.2	3135.48	3550.3
45	4388.1	4968.9	4702.56	5320.9	4388.31	4973.0	4701.87	5324.5
60	6685.06	7571.3	6268.88	7096.0	6684.7	7574.7	6268.41	7098.3

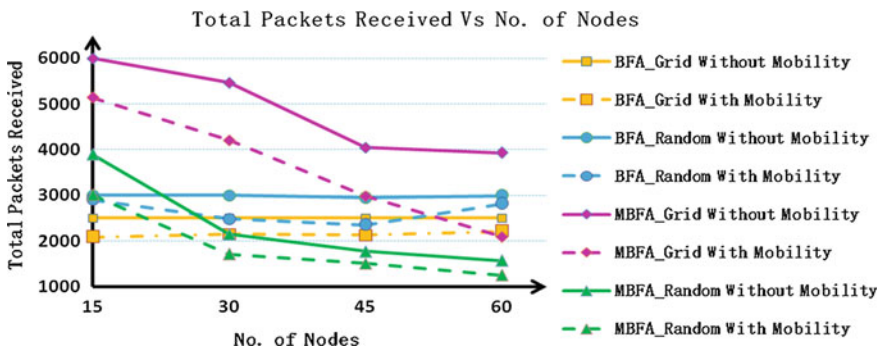


Fig. 2 Total packet received versus number of nodes with varying network traffic

received will be reduced. The performance of Grid topology is good when compared to random topology. In grid topology, the nodes are placed on exact, pre-defined points on a grid, and in the random deployment technique the randomness of the deployment is not in our control.

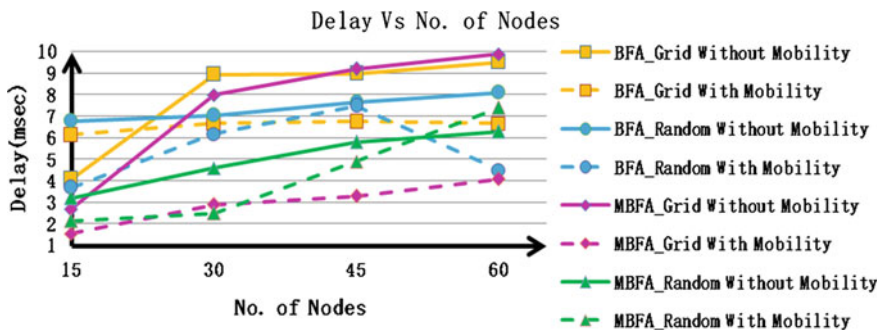


Fig. 3 Delay versus number of nodes with varying network traffic

The end-to-end delay verses number of nodes is shown in Fig. 3. As the size of the network increases, the traffic in the network increases which increases the average end-to-end delay due to multiple sources and destinations.

5 Conclusions

Mobile ad hoc networks require energy-efficient routing techniques to save battery capacity which improve the lifetime of the network. In this paper, energy-efficient modified Bellman–Ford algorithm is proposed. In modified Bellman–Ford algorithm residual energy is considered as a metric in addition to distance and hop count. All the nodes whose change in battery capacity is less than the threshold battery capacity are considered for routing. IEEE 802.11a/g OFDM standards using 64-QAM with generic radio energy model in QualNet Simulator is considered for energy calculation. In addition to the energy consumed by the radio transceiver circuit, the energy consumed by the processor, losses in the battery, and DC–DC converter are taken into consideration for energy calculation. Different scenarios of grid and random topologies using different network sizes 15, 30, 45, and 60 with multiple sources and destinations are considered for simulation. The impact of node deployment, mobility, and network size are considered for Bellman–Ford algorithm and modified Bellman–Ford algorithms. Simulation results shows that the performance of modified Bellman–Ford algorithms is better compared to Bellman–Ford algorithm and performance of grid topology is better when compared to random topology due to randomness of the deployment. Mobility of the nodes results increase in delay spread and the loss of orthogonality among the OFDM subcarriers which cause inter carrier interference hence number of packets received will be reduced. Hence, the performance of static network is better when compared to a mobile network. As the size of the network increases, the traffic in the network increase due to more number of multiple sources and multiple destinations. As the network size increases the average delay and jitter increases.

For future work, the performance of modified Bellman–Ford algorithm is improved by cooperative relays using a power saving mode will be studied.

References

1. Singh, S., Woo, M., Raghavendra, C.S.: Power-aware routing in mobile ad hoc networks. In: Proceedings of MobiCom '98, Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Dallas, pp. 181–190 (1998)
2. Toh, C.K.: Maximum battery life routing to support ubiquitous mobile computing in wireless ad-hoc networks. *IEEE Commun. Mag.* **39**(6), 138–147 (2001). doi:[10.1109/35.925682](https://doi.org/10.1109/35.925682)
3. Chang, J-H., Tassiulas L.: Maximum lifetime routing in wireless sensor networks. *IEEE/ACM Trans. Netw.* **1**(4), 609–619 (2004)

4. Tekbiyik, N., Uysal-Biyikoglu, E.: Energy efficient wireless unicast routing alternatives for machine-to-machine networks. *J. Netw. Comput. Appl.* **34**(5), 1587–1614 (2011). doi:[10.1016/j.jnca.2011.02.005](https://doi.org/10.1016/j.jnca.2011.02.005)
5. Dehghan, M., Ghaderi, M., Goeckel, D.: Minimum-energy cooperative routing in wireless networks with channel variations. *IEEE Trans. Wireless Commun.* **10**(11), 3813–3823 (2011)
6. Shivashankar, Suresh, H.N.: Design of novel protocol to increase network lifetime in MANETs. *Int. J. Adv. Res. Comput. Commun. Eng.* **2**(10), 3974–3978 (2013)
7. Sergiou, C., Vassiliou, V.: Energy utilization of HTAP under specific node placements in wireless sensor networks. In: *IEEE Wireless Conference (EW)*, pp. 482–487 (2010)
8. Rama Devi, B., Asha Rani, M., Kishan Rao, K.: Performance of static networks for power saving mode. *Int. J. Adv. Eng. Global Technol.* **2**(6), 783–789 (2014)
9. Shakywar, H., Sharma, S., Sahu, S.: Performance analysis of DYMO, LANMAR, STAR routing protocols for grid placement model with varying network size. *IJCTA* **2**(6) 1755–1760 (2011)
10. Rama Devi, B., Asha Rani, M., Kishan Rao, K.: Energy efficient cooperative node selection for OFDM systems based on SNR estimation. *Int. J. Adv. Comput. Electr. Electron. Eng.* **3**(1), 32–36 (2014)
11. Bechkit, W., Challal, Y., Bouabdallah, A., Koudil, M., Souici, B., Benatchba, K.: A new weighted shortest path tree for converge cast traffic routing in WSN. In: *IEEE Symposium on Computers and Communications (ISCC)*, pp. 187–192 (2012)
12. QualNet 6.1 Wireless Model Library.: Scalable Network Technologies, September 2012

Text-Dependent Versus Text-Independent Speech Emotion Recognition

Biswajit Nayak and Manoj Kumar Pradhan

Abstract The communication between individual and equipment is through speech emotion recognition which plays a vital role and is very exigent to handle. Today, this field has become an important area of research. It has wide range of applications. This paper analyzes the performance of emotion recognition for eight speakers. Indian Institute of Technology Kharagpur Simulated Hindi Emotional Speech Corpus (IITKGP-SEHSC) emotional speech corpora used for emotions recognition. The sentiments under surveillance for this study are anger, fear, happy, neutral, sarcastic, and surprise. The categorization is prepared using Gaussian mixture model (GMM). Mel-frequency cepstral coefficients (MFCCs) attributes have been used for defining the emotions. We have extracted the percentage of accuracy of emotion for both text-dependent data and text-independent data. We also observed that emotion recognition performance depends on text and speaker. We found that the percentage of accuracy of text-dependent data is more than the text-independent data.

Keywords Speech emotion recognition • Gaussian mixture model • Male-scale frequency cepstral coefficient • IITKGP-SEHSC

B. Nayak (✉)

Department of Computer Science & Engineering, Bhubaneswar Engineering College (BEC),
Bhubaneswar, Odisha, India

e-mail: b_nayak2007@yahoo.co.in; biswajit.nayak.mail@gmail.com

M.K. Pradhan

Department of Electronic & Communication Engineering, Gandhi Institute for Education
and Technology, Bhubaneswar, Odisha, India

e-mail: manojkumarpradhan007@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_16

1 Introduction

An individual articulates its thought through emotion and the means of conveying can be through look, motion, and verbal communication. The way of transmitting vital information is basically through emotions. Existence of feeling creates communication to be more accepted. Emotion is used comprehensively by an individual to express its intent through speech. It is pragmatic that similar message can be transmitted in various way by means of suitable emotion. The information contained in any speech signal includes proposed message, identity of speaker, and emotion status of speaker. The best ever natural method of communication between individuals is speech signal. Therefore, the fastest and competent way of interaction between an individual and machine is speech. On the other hand, it is vital that equipment should have adequate astuteness to distinguish human tone of voice. Till date, it is difficult to accomplish a natural interface linking an individual and machine since the machine is not capable to manipulate the feeling hidden in the speech of a speaker. Consequently, the recognition of sentiments in the communication is required to realize an individual's emotional state and analyze the message. Therefore, it is essential to build up a speech system which can identify the emotion in a speech efficiently.

In order to build a speech emotion recognition system, it is vital to decide the set of imperative emotions to be categorized by an automatic emotion recognizer. Classifying various types of emotion is a very difficult task. Therefore, researchers agree with the theory of "palette" which illustrates that every emotion is composed of primary emotions alike to the manner that every color is an amalgamation of a few primary colors. Key emotions are anger, fear, happy, neutral, sarcastic, and surprise. The implementation of speech emotion recognition system is extremely exigent due to the following facts. Foremost, it is not an easy task to clarify which features of a speech is more influential in differentiating among emotions. The auditory inconsistency brings in due to subsistence of dissimilar sentences, spokesperson, dialog technique, and speaking rates put in an additional barrier. Since these properties directly affect most of the common extracted speech features such as pitch, and energy form. Furthermore, there may be an additional professed emotion present in the matching sound; every emotion matches up to a different section of the vocal expression [1–3].

The formation of this article is as follows. The subsequent part, Sect. 2, illustrates the implementation of emotional speech corpus. Section 3 defines the categorization form for speech emotion recognition. Section 4 illustrates about attribute extraction. Section 5 illustrates the structural design of Emotion recognition method. Section 6 presents the experiment study, outcome of testing, and interpretation of outcome. And conclusion in Sect. 7 and the ending with reference.

2 Emotional Speech Corpus

In order to characterize the emotions synthesizing or by recognizing, we need to have an appropriate database of emotional speech. The planning and assortment of emotional speech corpora primarily depends on the objective of study. For an instance, synthesis of emotional speech can be carried out through a single speaker emotional speech corpus, but to recognize emotions we require database containing different style of emotion from multiple speakers.

The assessment obtainable in this part vitally evaluates the database of emotional speech databases on the basis of verbal communication, number of emotions, and the technique of compiling. The database considered for execution of research work is basically designed and build up at the Indian Institute of Technology, Kharagpur, in order to study speech emotion recognition. The speech database used for experiment is the first one to be built up by utilizing the frequent emotions in our everyday dialog. This corpus considered is adequately outsized to evaluate the speaker's emotion. Indian Institute of Technology Kharagpur Simulated Hindi Emotional Speech Corpus (IITKGP-SEHSC) is a Hindi speech database recorded using 10 (5 males and 5 females) proficient actors from All India Radio (AIR) Varanasi, India. In this recording, they have taken into account eight different emotions for creating a database which are anger, disgust, fear, happy, neutral, sadness, sarcastic, and surprise. Every actor is asked to articulate 15 sentences in 8 set sentiment in single gathering. The total number of sitting for creating the database is 10. The sum of all expressions in the database is 12,000 (15 transcript punctuals \times 8 sentiments \times 10 spokes persons \times 10 sitting). Every feeling has 1500 expression. The time period of database is 7 h.

In this analysis, six emotions have been considered out of eight from the database. The emotions considered for analyzing are anger, fear, happy, neutral, sarcastic and surprise. Two separate groups have been considered for data sample. One group is utilized for instructing the model for recognition and other group is used for testing the existing trained system. The different emotions extracted from these testing data samples are being classified using GMM classifier. The already trained data sample is compared with data sample considered for experiment purpose. The outcome of comparison is detection of emotion from these samples of data [4–6].

3 Classification Model

GMM model is a probabilistic model for density clustering and estimation. GMMs are very competent in representing multimodal allocations, their preparation and testing requisites are to a great extent is minimal. GMMs cannot sculpt sequential formation of the training data in view of the fact that all the training and testing equations are based on the postulation that all vectors are self-regulating. It is vital to determine the optimal figures of Gaussian components, which is a difficult task.

The categorization tools applied for developing an emotion recognition model is GMMs. They model the probability density function of observed variables using a multivariate Gaussian mixture density. Given a series of inputs, GMM refines the weights of each distribution through expectation–maximization algorithm [7–9].

4 Speech Emotion Recognition Feature Extraction

Mel-frequency cepstral coefficients (MFCCs) are based on the known variation of the human ear’s critical bandwidth with frequency filter spaced linearly at low frequencies and logarithmically at high frequencies have been used to capture the phonetically imperative distinctiveness of speech. This is expressed in the Mel-frequency scale, which is linear frequency spacing below 1000 Hz and a logarithmic spacing above the 1000 Hz. This frequency warping can allow for enhanced illustration of sound. The speech signal has to process through several steps to calculate the Mel-frequency cepstral coefficient. Initially it is preprocessed then frames are extracted, after that each individual frames are windowed upon which fast Fourier transform (FFT) algorithm is applied which is passed through the Mel-filter bank, then logarithm and discrete cosine transformation are applied to get the Mel-frequency cepstral coefficients of given emotional speech [10, 11]. In case of speech processing, the Mel-frequency cepstrum is a depiction of the short-term power spectrum of a speech frame using a linear cosine transform of the log power spectrum on a nonlinear Mel-frequency scale. Conversion from normal frequency to Mel-frequency is calculated using Eq. (1):

$$m = 2595 \log_{10} \left[\frac{f}{700} + 1 \right] \quad (1)$$

where f is the normal frequency and m is the Mel-frequency and for obtaining Mel-frequency cepstral coefficients (MFCCs) from a speech signal using certain steps given in Fig. 1.

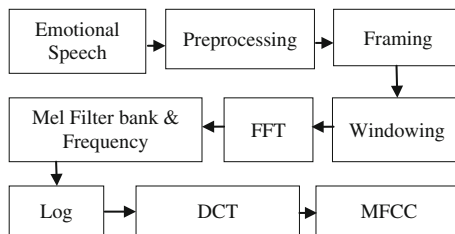


Fig. 1 Steps to calculate MFCC

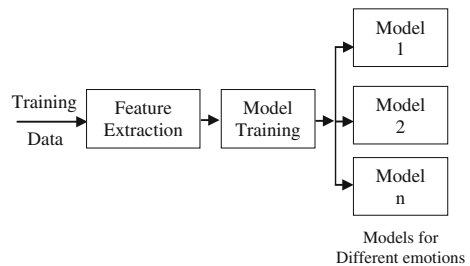
Initially, each signal is preprocessed by emphasizing on the speech signal. In the next stage, signals are divided into sequence of frame with frame size 30–40 ms. In order to minimize the signal discontinuities at the beginning and the end of each frame, windowing method is used. The magnitude spectrum for each window frame is computed by using Fourier transformation. Then Fourier transform signal passed through Mel-filter bank to obtain Mel spectrum. Then Log Mel-frequency coefficient is computed by passing Mel spectrum through Log. At the end discrete cosine transformation is applied to get the desired MFCCs [12–14].

5 Emotion Recognition System Architecture

In this work, we recognize the emotion present in speech using Gaussian mixture model (GMM). GMM on the other hand consider a signal to contain different component that are independent of each other. These components represent the broad acoustic classes. IITKGP-SEHSC database is used for the recognition. We are using Mel-frequency cepstral coefficients (MFCCs) feature of speech sample for classifying speech sample into different emotions. Our developed emotion recognition system basically has two phases, training phase and testing phase. In training phase, using training data we create the model for each emotion. Furthermore, in testing phase new speech sample is tested with all emotion models which we got in training phase and speech sample can be classified in particular emotion according to probability values of each model. In this work, we have focused on six types of emotions anger, fear, happy, neutral, sarcastic, and surprise. GMM classification model is used for recognition [8, 9]. System architecture has two phases. Training phase responsible training models for different emotion (Fig. 2).

Testing phase is responsible for finding the probability value for all models. Whichever model having highest probability value the speech utterance can be classified according to that model [7, 8] (Fig. 3).

Fig. 2 Emotion recognition architecture for training phase



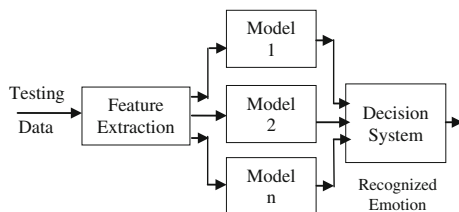


Fig. 3 Emotion recognition architecture for testing phase

6 Result and Observation

6.1 Text Dependent

For this experiment, we use three models for different emotion. For each speaker, 1–7 sessions of each emotion were used to train the model (15 sentences \times 7 sessions = 105 utterances of each emotion) and remaining 8–10 sessions (15 sentences \times 3 sessions = 45 utterances) were used for testing purpose. Therefore, in this case, training and testing data were 70 and 30 %, respectively. We have used three different components (8, 16, and 32). Table 1 shows the percentage of accuracy for speaker 7 from Hindi language database.

Tables 2 and 3 and 4 shows the text-dependent emotion recognition performance (In percentage) of speaker 7 for 8-centered GMM, 16-centered GMM, 32-centered GMM. Table 5 shows the average text-dependent emotion recognition performance (in percentage) of 8 speakers from Hindi language database.

Table 1 Accuracy of speaker 7

GMM	8 centered	16 centered	32 centered
Performance (%)	71.48	75.19	80.37

Table 2 Accuracy for 8-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	40.0	00.0	31.1	17.8	08.9	02.2
Fear	00.0	88.9	02.2	08.9	00.0	00.0
Happy	24.4	00.0	75.6	00.0	00.0	00.0
Neutral	24.4	00.0	11.1	64.4	00.0	00.0
Sarcastic	02.2	00.0	02.2	00.0	95.6	00.0
Surprise	26.7	00.0	00.0	08.9	00.0	64.4
Average accuracy	71.48 (%)					

Table 3 Accuracy for 16-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	48.9	00.0	31.1	11.1	06.7	02.2
Fear	00.0	86.7	02.2	08.9	00.0	00.0
Happy	22.2	00.0	71.1	06.7	00.0	00.0
Neutral	31.1	00.0	13.3	55.6	00.0	00.0
Sarcastic	00.0	00.0	02.2	00.0	97.8	00.0
Surprise	06.7	00.0	00.0	02.2	00.0	91.1
Average accuracy	75.19 (%)					

Table 4 Accuracy for 32-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	60.0	00.0	28.9	04.4	06.7	00.0
Fear	00.0	95.6	02.2	02.2	00.0	00.0
Happy	22.2	00.0	77.8	00.0	00.0	00.0
Neutral	24.4	00.0	08.9	66.7	00.0	00.0
Sarcastic	00.0	00.0	02.2	00.0	97.8	00.0
Surprise	11.1	00.0	02.2	02.2	00.0	84.4
Average accuracy	80.37 (%)					

Table 5 Average accuracy of the entire speakers

GMM	8 centered	16 centered	32 centered
Performance (%)	66.98	71.53	75.18

6.2 Text Independent

In text-independent case, two-third or 66 % prompts were used for training the models and one-third or 33 % prompts were used for testing purpose. In text-independent case, training data for each emotion contained 100 utterances (10 sentences \times 10 sessions) and testing data for each emotion contained 50 utterances (5 sentences \times 10 sessions). The performance of speaker 6 is mentioned in Table 6.

Tables 7, 8 and 9 show the text-independent accuracy (in percentage) of speaker 6 for three different centered GMM. Table 10 shows the average of 8 speakers.

Table 6 Accuracy of speaker 6

GMM	8 centered	16 centered	32 centered
Performance (%)	60.66	64.66	71.33

Table 7 Accuracy for 8-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	30	02	20	20	06	22
Fear	00	82	02	16	00	00
Happy	04	08	48	10	10	20
Neutral	02	04	04	64	22	04
Sarcastic	00	06	16	04	68	06
Surprise	00	02	12	12	02	72
Average accuracy	60.66 (%)					

Table 8 Accuracy for 16-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	42	02	20	18	08	10
Fear	02	90	02	06	00	00
Happy	02	04	64	08	08	14
Neutral	02	12	08	56	16	06
Sarcastic	01	04	18	08	66	02
Surprise	02	04	16	06	02	50
Average accuracy	64.66 (%)					

Table 9 Accuracy for 32-centered GMM

	Anger	Fear	Happy	Neutral	Sarcastic	Surprise
Anger	56	06	14	10	06	08
Fear	00	92	00	06	00	02
Happy	02	06	52	08	16	16
Neutral	00	00	08	82	08	02
Sarcastic	02	00	10	12	74	02
Surprise	02	06	14	04	02	72
Average accuracy	71.33 (%)					

Table 10 Average accuracy of the entire speaker

GMM	8 centered	16 centered	32 centered
Performance (%)	65.96	70.43	73.68

7 Conclusion

Even though it is complicated to obtain a precise outcome, however, we can analyze the variations that occur when emotion transforms. MFCC description of speech sample is applied for the detection. The experiment was conducted using emotional speech database of Hindi language. The execution of emotion

recognition depends on spokesman, emotion, and verbal communication implemented in the process. Gaussian mixture model has been used in order to classify six different emotions as: anger, fear, happy, neutral, sarcastic and surprise of eight different speakers from database. Three type of GMM model used namely 8-centred GMM, 16-centred GMM, and 32-centred GMM to find the performance of text-dependent data and text-independent data. It is observed that, the average recognition accuracy is observed to be 75.18, 71.53, and 66.98 %, respectively is for 32-, 16-, 8-centered GMM for text-dependent data. Whereas, 73.68, 70.43, and 65.96 %, respectively, for text-independent data. The percentage of accuracy of text-dependent data is more than the text-independent data.

References

1. Koolagudi, S., Rao, K.S.: Emotion recognition from speech using source, system, and prosodic features. *Int. J. Speech Technol.* **15**, 265–289 (2012)
2. Ververidis, D., Kotropoulos, C.: Emotional speech recognition: resources, features, and methods. *SPC* **48**, 1162–1181 (2006)
3. Rabiner, L.R., Juang, B.H.: *Fundamentals of Speech Recognition*. Prentice-Hall, Englewood Cliffs (1993)
4. Koolagudi, S.G., Maity, S., Kumar, V.A., Chakrabarti, S., Rao, K.S.: IITKGP-SESC: Speech Database for Emotion Analysis, vol. 40, pp. 485–492. Springer, Berlin (2009)
5. Koolagudi, S., Reddy, R., Yadav, J., Rao, K.S.: IITKGP-SEHSC: Hindi speech corpus for emotion analysis. In *International Conference on Devices and Communications (ICDeCom)*, pp. 1–5 (2011)
6. Moataz, M.H., Kamel, A.E., Mohamed, S., Fakhreddine, K.: Survey of speech emotion recognition: feature, classification, schemes and databases. *Elsevier* **44**(3), 572–587 (2011)
7. Cheng, X., Duan, Q.: Speech emotion recognition using Gaussian Mixture Model. In: *In the 2nd International Conference on Computer Application and System Modeling* (2012)
8. Thapliyal, N., Amoli, G.: Speech based emotion recognition with Gaussian Mixture Model. *Int. J. Adv. Res. Comput. Eng. Technol.* **1**, 65–69 (2012)
9. Reynolds, D.: Gaussian mixture models: MIT Lincoln Laboratory, 244 St Wood, emotion recognition using support vector regression. In: *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*
10. Wankhade, S.B., Tijare, P., Chavhan, Y.: Speech emotion recognition system using SVM AND LIBSVM. *Int. J. Comput. Sci. Appl.* **4**(2) (2011) ISSN: 0974-1003
11. Khanna, M.P., Kumar, S., Toscano-Medina, K., Nakano, M., Meana, H.P.: Application of vector quantization in emotion recognition from human speech. *Inf. Intell. Syst. Technol. Manage. Commun. Comput. Inf. Sci.* **141**, 118–125 (2011)
12. Panda, B., Padhi, D., Dash, K., Mohanty, S.: Use of SVM classifier & MFCC in speech emotion recognition system. *IJARCSSE* **2**(3) (2012) ISSN: 2277128X
13. Olivares-Mercado, J., Aguilar, G., Toscano-Medina, K., Nakano, M., Meana, H.P.: GMM versus SVM for face recognition and face verification. In: *Corcoran, P. (ed.) Reviews, Refinements and New Ideas in Face Recognition* (2011). ISBN: 978-953-307-368-2
14. Utane, A.S., Nalbalwar, S.L.: Emotion recognition through speech using gaussian mixture model and hidden markov model. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**(4) (2013). ISSN: 2277 128X

A Novel Approach for Speaker Recognition by Using Wavelet Analysis and Support Vector Machines

Kanaka Durga Returi, Vaka Murali Mohan and Y. Radhika

Abstract Speaker recognition approach through wavelet analysis as well as support vector machines is presented in this paper. The wavelet-based approach is used to differentiate among regular and irregular voices. The wavelet filter banks were utilized to coincide by means of support vector machine for extraction of the feature and its classification. This approach creates utilization of wavelets as well as support vector machine to separate particular speech signal through multi-dialog settings. In this approach, first we apply the wavelets to calculate audio features those have sub-band power and calculated pitch values from the given data of the speech. Multi-speaker separation of speech data is carried out by the utilization of SVM more than these audio features as well as other values of the signal. This entire database was utilized to calculate the performance of the system and it represents over 95 % accuracy.

Keywords Speaker recognition · Wavelet analysis · Support vector machines

1 Introduction

The speaker recognition through text independence has been strongly studied for so many years. The basics of the usual recognition of the speaker, extraction of the feature, and modeling for the identification of the speaker by using more advanced computational methods to discuss the strength and irregularity. Modern development

K.D. Returi (✉) · Y. Radhika
Department of Computer Science & Engineering,
GITAM University, Visakhapatnam, AP, India
e-mail: durga1210@gmail.com

V.M. Mohan
Department of Computer Science & Engineering, TRR College of Engineering,
Patancheru, Hyderabad, TS, India
e-mail: vakamuralimohan@gmail.com

of the vectors through supervectors releases the new and innovative field of investigation and technology for its representation for the recognition of speaker. Support vector machine is one of most efficient machine learning techniques to study the decision during the development of discrimination and good generalization characteristics. SVM contains and establishes to be the successful classifiers on so many pattern recognition problems. Support vector machines have been demonstrated as a powerful method for pattern classification.

SVM is utilized to insert the input features through elevated dimensional liberty to divide patterns by means of a hyperplane among improved discrimination ability. SVM with mean square error classifier gives more accurate classification. SVM is also utilized to insert inputs into elevated dimensional liberty and represented as different classes in hyperplane. Significant feature of using SVM is to design the inner invention, kernel, and stimulate by means of the elevated dimensional mapping. SVM is used to characterize a novel and innovative approach to pattern recognition by means of effectively applied series of pattern recognition approaches, as well as article recognition, speaker detection, genetic material utility prediction, etc. In these, overall performance of the SVM is equivalent or considerably improved than the established learning methods, together with artificial neural networks.

Discrete moment of the processing signals of the recorded voices can be utilized for separation of unusual acoustical individuality which differentiates the regular as well as irregular human voices. Irregular voices connected to the glottal tract were generally recognized during audio perceptual principles like breathiness, hoarseness, and harshness. The composite configuration of the genetic structure for speech synthesis, pathologies by means of insensitive distinctiveness may be confused by means of the perceptual identity as hoarse. The strength of the speech signal is measured in terms of turbulence ensuing through malfunction of the vocal folds and noise in terms of spectral mechanism. In the latest computer-based algorithms by means of laryngeal pathology identification is illustrated in literature through wavelets, neural networks, and SVM.

2 Literature Review

Several researchers did several research works related to this area. Some of them are: Truong et al. [1] presented a novel multi-speaker segmentation method that makes use of the wavelet analysis and support vector machine to separate various speech signals of the speakers through multi-dialog approach. Zhang et al. [2] reported recognition of the speech by using the function of the ORF kernel for getting accuracy with higher recognition rate. Fonseca et al. [3] described an algorithm toward the recognition of the voice by using it with discrete wavelet transform proposed by Daubechies; coefficients of the linear prediction and support vector machines by using with least squares. Sangeeth and Jothilakshmi [4] developed a novel spoken keyword detection algorithm for differentiating

keywords. Pawan and Raghunath [5] presented a text-independent speaker recognition technique with Fourier transform by means of MFCC and SVM. Huang et al. [6] presented the properties of sound features, methods for the normalization and modeling techniques for the classification of statistical state of the speaker. Wu and Lin [7] presented the learning of speaker classification for protection of systems by means of strength of the utterances spoken by speaker, signal preprocess, extraction of the features through the transform of the wavelet packet, and identification of the speaker with neural networks. Francesco et al. [8] presented an approach for the recognition with wavelet analysis and techniques for execution of SVM-based classification. Arjmandi and Pooyan [9] presented an approach for the identification of different voices with the transforms of the short-time Fourier and continuous wavelet. Campbell et al. [10] reported that as support vector machines are established as a beautiful powerful technique for the classification of the patterns, speaker, and language evaluations of the system. You and Li [11] studied the relevance factors of models in Gaussian mixture and SVM for speaker and speech recognition. Shriberg et al. [12] described a novel approach for the recognition of the speaker with different durations, pitch, and features for developed syllable in recognition of the speech. Dileep and Chandra Sekhar [13] reported the performance of the classifiers of the SVM for recognition of speech emotion and classification of speaker with kernels. Zhao et al. [14] described the system designing of the speaker verification, test normalization to further enhance performance by using SVM. Kinnunen et al. [15] presented methods for system designing of the verification speaker by using VQ and GMM. Kinnunen and Li [16] presented the recent improvement and converse of the estimation methodology of the speaker recognition systems. McLaren et al. [17] presented the rate of recognition through SVM-based identification and verification of the speaker with the model of weight-based factor analysis. Peng and Wang [18] presented the SVM model for recognition of tone in a speaker-independent task. Wang et al. [19] investigated SVM ensembles with four different collection constructing models. Hua and Sun [20] reported a new and innovative model of configuration forecast by means of binary classifiers using support vector machine. Gajšek et al. [21] presented a well-organized method for recognition of the tasks of the features by using Gaussian mixture model. Mohamed and Ramachandran [22] described the improvement in property of the normalization by using HMM and artificial neural networks. Saeedi et al. [23] proposed a method for the separation of normal and abnormal voices through wavelet analysis and support vector machines. Shih et al. [24] reported the classification of the existing speaker with SVM classifier and speaker-independent model. Haniilçi and Ertaş [25] analyzed the rate of the speaker recognition system with the duration of the training and test data. Daqrouq [26] presented the methods of wavelet transform and neural network for the development and identification of the text-independent speaker. Govindan et al. [27] reported a method of robust speaker recognition with adaptive wavelet shrinkage method for noise suppression. This identification and recognition are completed by using adaptation of Mel-frequency cepstral coefficient for mapping the voice signal segments. Rossi and Villa [28] investigated the applications of the SVMs for FDA.

3 Wavelet Analysis

Wavelet analysis is a signified technique by means of different-sized classification. This permits the utilization of the stretched instance period wherever higher frequency in sequence was achieved. ‘Wavelet’ involves a little wave of limited period and limited power and it is connected by means of signal for getting coefficients of wavelet. The orientation wavelet is recognized as the care for wavelet and it is used to evaluate the coefficients for complete variety of dilation and conversion factors. The most important improvement present in wavelets is the capability to execute analysis. The significance of the shrinkage in wavelet for noise removal, a dual threshold strategy was developed to suppress noise, preserve signal coefficients, and minimize the introduction of artifacts. The recognition is completed using modification of Mel-frequency cepstral coefficient for mapping of the overlapped voice signal segments.

It is a technique for different classifications and agrees to utilize extended instance period of particular low occurrence in sequence and shorter constituency with higher frequency. Features extraction from the data have been preprocessed and normalized to remove prospective fluctuations of baseline, interferences, noises, and WP tree decomposition. The method of wavelet packet analysis is the simplification of wavelet decomposition and it gives the richer signal analysis. The atoms of the wavelet packets are waveforms keyed by three different constraints such as position, frequency, and scale in decomposition of wavelet. The wavelet transform has been described and denoted as inner invention of a signal $x(t)$ through its original wavelet $\Psi(t)$:

$$\psi_{a,b}(t) = \psi\left(\frac{t-b}{a}\right) \quad (1)$$

where a and b are denoted as parameters of the scale and shift; original wavelet can be present and translated through modulating a and b .

The wavelet method is utilized for the investigation of the specified input speech signal. This signal refers in the direction of a specific word spoken by a specific speaker. The wavelet method consists of a detail and approximation. An approximation might be divided into different series of levels. These levels are original audio signal (shown in Fig. 1) and regenerated output waveform (shown in Fig. 2). This shows the degree of detail in this method of analysis. In this case, the number of details is taken and fixed to be five and denoted as D_1, D_2, D_3, D_4, D_5 , and A_5 . Here, in this case, for each word spoken a total of five features were extracted and presented in Table 1.

4 Support Vector Machine

SVM is found as lying on theory of statistical learning. This is utilized for learning the predicted upcoming data. SVM is trained through explaining controlled quadratic optimization. SVM executes mapping of the inputs through a high-dimensional space

Fig. 1 Original audio signal

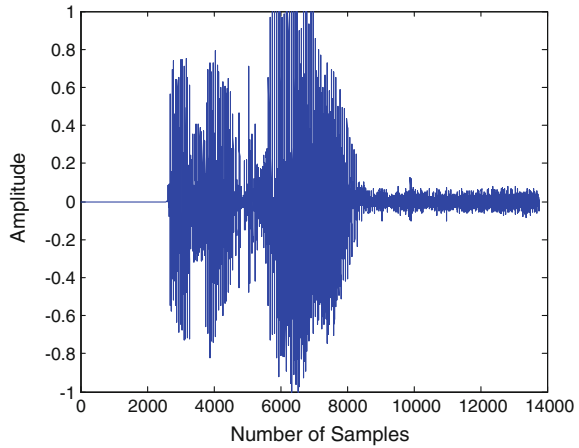
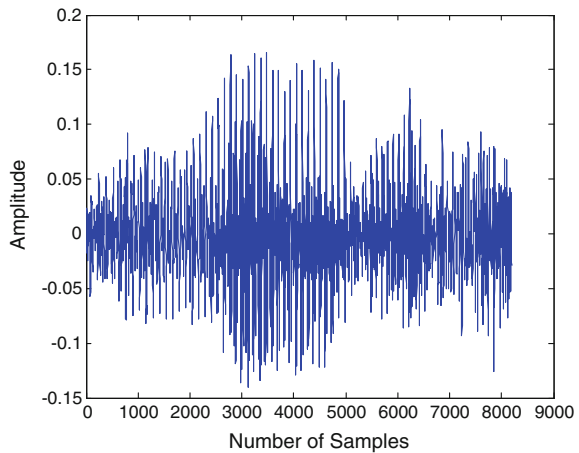


Fig. 2 Regenerated wave signal



by means of nonlinear functions. SVM is utilized to study a variety of illustrations like neural networks, polynomial estimators, splines, etc., but there is a single best possible solution designed for each one of the SVM parameters. SVM is entirely different in new learning machines like neural networks are trained through the popular backpropagation learning algorithm. SVM consists four most popular features like duality, convexity, kernels, and sparseness. Support vector machines is a technique for pattern detection and function evaluation in the construction of statistical learning theory. SVM is a commanding classifier that has been newly implemented for speaker detection. This can be useful together by means of spectral, prosodic, and high-intensity features. A SVM was distinct by a separating hyperplane and it is labeled as training data (supervised learning), the outputs of the algorithm are best possible hyperplanes which classify new examples. SVM is a valuable procedure for data classification. Although it is considered that NN are simpler to use than

Table 1 Wavelet results

Record	D_5	D_4	D_3	D_2	D_1	A_5
1	16,800	33,592	67,193	134,378	268,765	16,800
2	26,560	53,112	106,233	212,458	424,925	26,560
3	24,480	48,952	97,913	195,818	391,645	24,480
4	26,560	53,112	106,233	212,458	424,925	26,560
5	23,520	47,032	94,073	188,138	376,285	23,520
6	14,720	29,432	58,873	117,738	235,485	14,720
7	17,920	35,832	71,673	143,338	286,685	17,920
8	12,160	24,312	48,633	97,258	194,525	12,160
9	16,320	32,632	65,273	130,538	261,085	16,320
10	15,840	31,672	63,353	126,698	253,405	15,840
11	19,360	38,712	77,433	154,858	309,725	19,360
12	19,360	38,712	77,433	154,858	309,725	19,360
13	14,080	28,152	56,313	112,618	225,245	14,080
14	13,280	26,552	53,113	106,218	212,445	13,280
15	17,920	35,832	71,673	143,338	286,685	17,920
16	10,240	20,472	40,953	81,898	163,805	10,240
17	12,160	24,312	48,633	97,258	194,525	12,160
18	13,760	27,512	55,033	110,058	220,125	13,760
19	12,800	25,592	51,193	102,378	204,765	12,800
20	12,160	24,312	48,633	97,258	194,525	12,160

this, sometimes unacceptable results are attained. A categorization assignment generally engaged through training and testing the data that consists of a few data illustrations. The purpose of SVM is to create a method which forecasts objective value of data illustration in testing set. SVM classification is a supervised learning technique. Known labels specify whether the system is doing in a right way or not. This information spots to a preferred response, authenticating the correctness of the system, or be applied to help the system study to act correctly.

5 Results

The performance achieved by data is obtained by its testing and validation of the data with classifier. The number of samples shown describes the algorithm which the database corresponds to 20 % for testing and data of 80 % for training. This method which processes almost all possible grouping with the 1st, 2nd, 3rd, and 4th levels of DWT revolution was utilized to access LS-SVM by using three various kernel functions such as linear, MLP, and RBF. The data with 40 speakers (20–40 years old) which include 20 males and 20 females were taken for the present study.

They were made to say four words each ‘Bhavana,’ ‘How,’ ‘Are,’ and ‘You.’ All the speakers say again these words and were recorded. The Kit of wavelet method in MATLAB 7.7.0 particularly with toolbox was utilized in this analysis. The total number of levels was taken and fixed as five and it gives five different values from the analysis of the wavelet. In this, no straight values were taken as input and were calculated from features obtained from the method of wavelets. With this approach a large set of data were obtained for each input and is taken as the training data. The total number of data sets taken for training as well as testing was 800. These iterations were carried out by utilizing 800 data sets for total 40 individual

Fig. 3 Approximation by fixed size LS-SVM based on maximal entropy 1.851

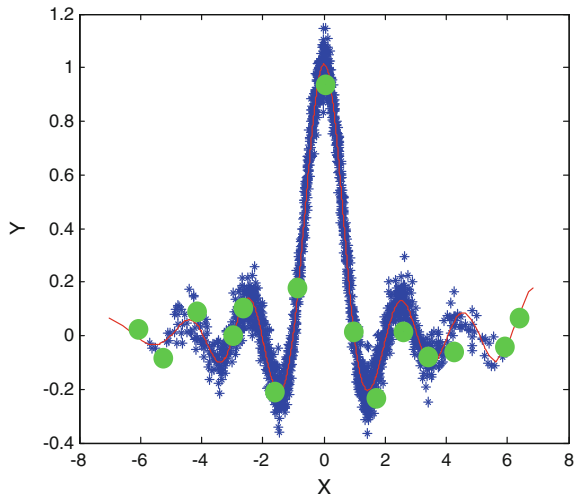
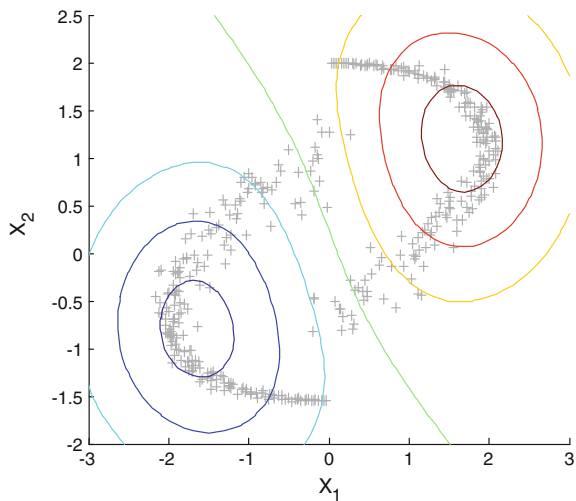


Fig. 4 Projections onto the First Kernel PC



speakers. For each speaker, 35 speech signals were obtained and were utilized for the analysis by using least square support vector machines.

The experimental result shows that approximation by fixed size LS-SVM with maximal entropy of 1.851 is shown in Fig. 3. Projections onto the First Kernel PC are shown in Fig. 4 and its denoising by computing an approximate pre-image is shown in Fig. 5.

The graphical representation of the categorization with the root mean square values for the filtered mechanism for the RBF kernels is shown in Fig. 6. The figure reveals that the classification of RBF at $\gamma = 10$, $\sigma^2 = 0.2$ in two different classes.

Fig. 5 Denoising by computing an approximate pre-image

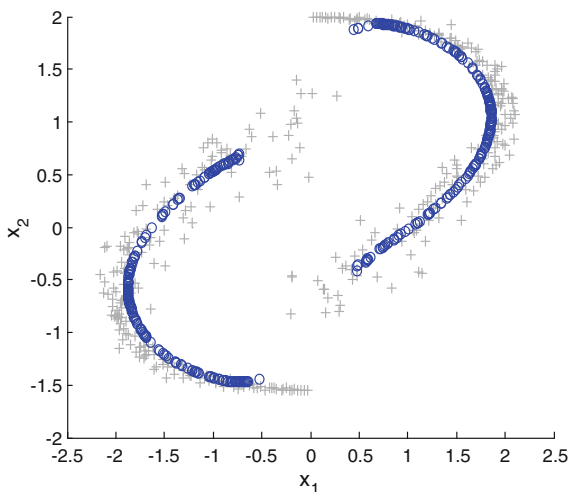


Fig. 6 LS-SVM-RBF $\gamma = 10$, $\sigma^2 = 0.2$ in two different classes

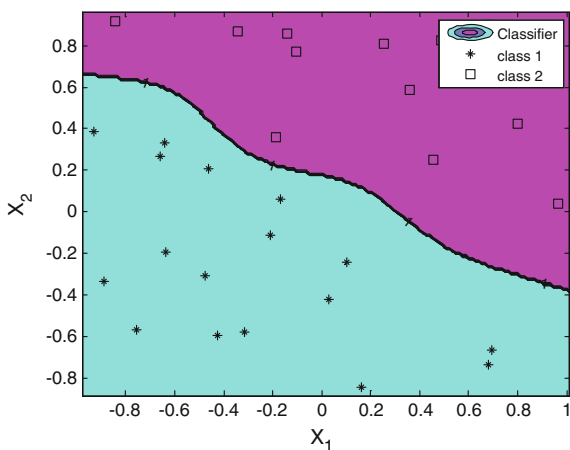


Fig. 7 Function estimation for LS-SVM-RBF $\gamma = 26.75$, $\sigma^2 = 0.99$

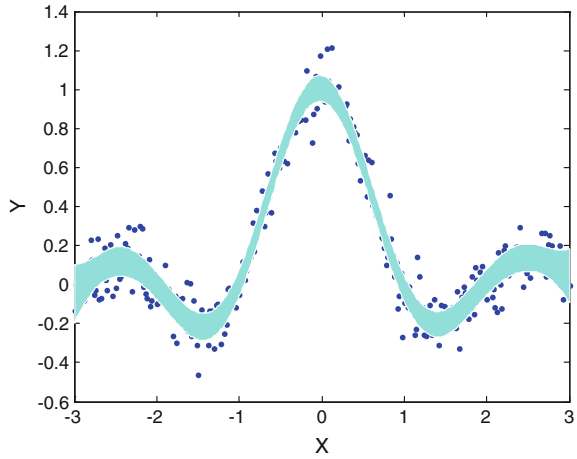
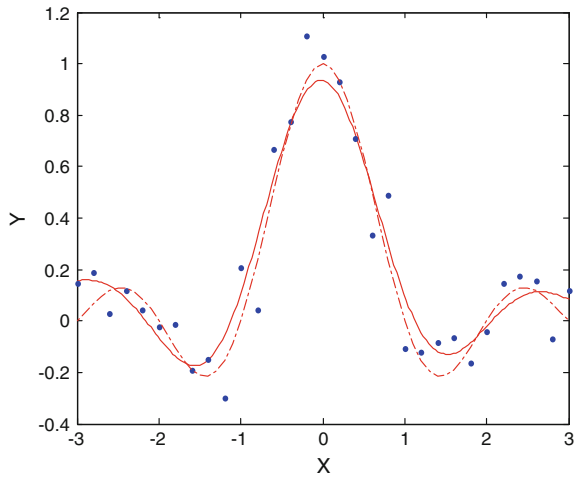
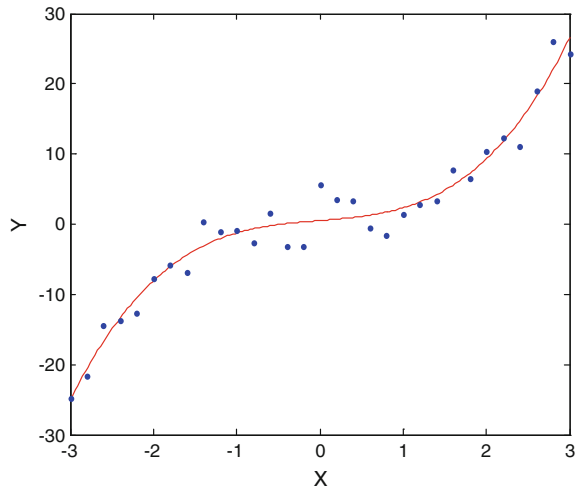


Fig. 8 Function estimation for LS-SVM-RBF $\gamma = 10$, $\sigma^2 = 0.3$



The function estimation for least square support vector machine for the RBF with $\gamma = 26.75$, $\sigma^2 = 0.99$ is shown in Fig. 7. The function estimation for least square support vector machine for the RBF with $\gamma = 10$, $\sigma^2 = 0.3$ is shown in Fig. 8. The function estimation for least square support vector machine with polynomial function with $\gamma = 3$, $t = 1$, and degree = 3 is shown in Fig. 9.

Fig. 9 Function estimation for LS-SVM-POL $\gamma = 3$, $t = 1$, degree = 3



6 Conclusions

A novel approach for speaker recognition by using wavelet analysis and support vector machine has been reported in this paper. The wavelet model is utilized to separate normal and abnormal voices. The filter banks in the wavelet were utilized in combination with feature extraction and support vector machines. This approach is utilized to separate the particular speaker signals by using multi-dialog environment with the above approaches. In this process, first wavelets are applied to speech signals to extract the features. Multi-speaker separation of the specified speech records can be expert by means of LS-SVM. The features of the parameter values were utilized to train the classifier with the approach of LS-SVM has recognized a sufficient effect of simplification to separate normal and affected voices. This database was utilized to calculate the performances of the system and it represents over 95 % of the accuracy in terms of classification and difficulty in relative to the speech signal's strength.

References

1. Truong, T.K., Chien, C.L., ShiHuang, C.: Segmentation of specific speech signals from multi dialog environment using SVM and wavelet. *Pattern Recogn. Lett.* **28**(11), 1307–1313 (2007)
2. Zhang, X., Liu, X., Wang, Z.J.: Evaluation of a set of new ORF kernel functions of SVM for speech recognition. *Eng. Appl. Artif. Intell.* **26**(10), 2574–2580 (2013)
3. Fonseca, E.S., Guido, R.C., Maciel, C.D., Pereir, J.C.: Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders. *Comput. Biol. Med.* **37**(4), 571–578 (2007)

4. Sangeeth, J., Jothilakshmi, S.: A novel spoken keyword spotting system using support vector machine. *Eng. Appl. Artif. Intell.* **36**, 287–293 (2014)
5. Pawan, K.A., Raghunath, S.H.: Fractional Fourier transform based features for speaker recognition using support vector machine. *Comput. Electr. Eng.* **39**(2), 550–557 (2013)
6. Huang, D.-Y., Zhang, Z., Ge, S.: Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines. *Comput. Speech Lang.* **28**(2), 392–419 (2014)
7. Wu, J.-D., Lin, B.-F.: Speaker identification using discrete wavelet packet transform technique with irregular decomposition. *Expert Syst. Appl.* **36**(2), 3136–3143 (2009)
8. Francesco, P., Fiore, U., Alfredo, D.S.: On the detection of card-sharing traffic through wavelet analysis and Support Vector Machines. *Appl. Soft Comput.* **13**(1), 615–627 (2013)
9. Arjmandi, M.K., Pooyan, M.: An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomed. Signal Process. Control* **7**(1), 3–19 (2013)
10. Campbell, W.M., Campbell, J.P., Reynolds, D.A., Singer, E.: Support vector machines for speaker and language recognition. *Comput. Speech Lang.* **20**(2-3), 210–229 (2006)
11. You, C.H., Li, H.: Relevance factor of maximum a posteriori adaptation for GMM–NAP–SVM in speaker and language recognition. *Comput. Speech Lang.* **30**, 116–134 (2014)
12. Shriberg, E., Ferrer, L., Kajarekar, S., Venkataraman, A.: Modeling prosodic feature sequences for speaker recognition. *Speech Commun.* **46**(3), 455–472 (2005)
13. Dileep, A.D., Chandra Sekhar, C.: Class-specific GMM based intermediate matching kernel for classification of varying length patterns of long duration speech using support vector machines. *Speech Commun.* **57**, 126–143 (2014)
14. Zhao, J., Dong, Y., Zhao, X., Yang, H., Liang, L., Wang, H.: Advances in SVM-based system using GMM super vectors for text-independent speaker verification. *Tsinghua Sci. Technol.* **13**(4), 522–527 (2008)
15. Kinnunen, T., Sidoroff, I., Marko, T., Pasi, F.: Comparison of clustering methods: A case study of text-independent speaker modeling. *Pattern Recogn. Lett.* **32**(13), 1604–1617 (2011)
16. Kinnunen, T., Li, H.: An overview of text-independent speaker recognition: from features to supervectors. *Speech Commun.* **52**(1), 12–40 (2010)
17. McLaren, M., Matrouf, D., Vogt, R., Bonastre, J.-F.: Applying SVMs and weight-based factor analysis to unsupervised adaptation for speaker verification. *Comput. Speech Lang.* **25**(2), 327–340 (2011)
18. Peng, G., Wang, W.S.-Y.: Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Commun.* **45**(1), 49–62 (2005)
19. Wang, S.-J., Mathew, V., Chen, Y., Lee, J.: Empirical analysis of support vector machine ensemble classifiers. *Expert Syst. Appl.* **36**(3), 6466–6476 (2009)
20. Hua, S., Sun, Z.: A novel method of protein secondary structure prediction with high segment overlap measure: support vector machine approach. *J. Mol. Biol.* **308**(2, 27), 397–407 (2001)
21. Gajšek, R., Mihelič, F., Dobrišek, S.: Speaker state recognition using an HMM-based feature extraction method. *Comput. Speech Lang.* **27**(1), 135–150 (2013)
22. Mohamed, A., Ramachandran Nair, K.N.: HMM/ANN hybrid model for continuous Malayalam speech recognition. *Procedia Eng.* **30**, 616–622 (2012)
23. Saeedi, N.E., Farshad, A., Farhad, T.: Support vector wavelet adaptation for pathological voice assessment. *Comput. Biol. Med.* **41**(9), 822–828 (2011)
24. Shih, P.-Y., Lin, P.-C., Wang, J.-F., Lin, Y.-N.: Robust several-speaker speech recognition with highly dependable online speaker adaptation and identification. *J. Netw. Comput. Appl.* **34**(5), 1459–1467 (2011)
25. Haniççi, C., Ertaş, F.: Investigation of the effect of data duration and speaker gender on text-independent speaker recognition. *Comput. Electr. Eng.* **39**(2), 441–452 (2013)

26. Daqrouq, K.: Wavelet entropy and neural network for text-independent speaker identification. *Eng. Appl. Artif. Intell.* **24**(5), 796–802 (2011)
27. Govindan, S.M., Duraisamy, P., Yuan, X.: Adaptive wavelet shrinkage for noise robust speaker recognition. *Digit. Signal Proc.* **33**, 180–190 (2014)
28. Rossi, F., Villa, N.: Support vector machine for functional data classification. *Neurocomputing* **69**(7-9), 730–742 (2006)

Real-Time Surveillance for Critical Activity Detection in ICUs

Dhakate Pankaj

Abstract In today's scenario, motion detection has gathered attention of researchers due to its promising applications in numerous areas. Such as video surveillance, patient monitoring, traffic management, security, video games, military armors, object classification, and sign language recognition. The review of this intelligent application demands gradation of technologies. The findings say that it is still in its early developmental stage and there is need to improve its robustness when applied to a complex and changing environment. Therefore, it is effective to improve the surveillance techniques. Upgradation can be obtained successfully using 3D camera, but it is expensive. The methodology mentioned in this paper follows the human eyes visualization concept by using pair of identical two-dimensional cameras to generate stereoscopic video. Growing number of cameras enables new signal processing applications. The amount of the data also increases which is to be processed to be supportive to design new modified algorithm to obtain accuracy. This can also track the movement using optimized Kalman filter. This real-time method is useful in monitoring and detecting every inch and second of information of interested areas. The technique mentioned in this paper is beneficial for online and offline applications. Project implemented using this paper trim down memory requirements for activity storage. It is efficient, sensitive, and absolutely useful for society welfare. The proposed method in this paper will activate a warning system, highlight the changes, and capture the live streaming video when minute movement of coma patient is detected, also it keeps track on mental stress of patients.

Keywords Kalman filter • Stereoscopy • Concatenate arrays • Stress analysis

D. Pankaj (✉)
E&TC Engineering Department, Singhad Academy of Engineering,
University of Pune, Pune, India
e-mail: pankajdhakate11@gmail.com

1 Introduction

In proposed method the concept of surveillance means to protect life. Motion detection has become more and more effective due to its promising applications in numerous areas [1]. But it is still in its early developmental stages and needs to improve its robustness when applied in a complex environment. Manual analysis of surveillance video is time intensive, prone to mistakes, and expensive. Psychological research also discovered limitation of the human mind in monitoring several signals simultaneously [2]. Several techniques for moving object detection have been developed among them; three representative approaches are temporal differencing, background subtraction, and optical flow [3]. Temporal differencing method is based on frame difference; it attempts to detect moving regions by making use of the difference of consecutive frames in a video sequence [4]. This method is highly adaptive to dynamic environments, but usually does a poor job of extracting the complete shapes of certain types of moving objects. The limitation of optical flow estimation in the multiscale variation framework is that it cannot correctly estimate the fine motion structures, especially for regions with significant and abrupt displacement variation [5]. Also these techniques are based on two-dimensional images which are sometimes incapable to detect changes hidden to the front view of cameras.

Figures 1 and 2 show that same object is captured from different positions for plane object and rounded object, respectively, using two-dimensional cameras. Left image was captured by shifting left camera to left direction by 3.3 cm from central position and right image was captured by shifting right camera to right direction by 3.3 cm from central position. Comparing to central images left and right images contain hidden information. Yellow brackets in left and right images show more information compared to central image. Video surveillance system needs to analyze the behaviors of humans in order to prevent the occurrence of the potentially dangerous circumstances [6]. So acquired images must be more informative.

To overcome limitation of two-dimensional cameras new advance method called surveillance using stereoscopic video is highlighted in this paper. The primary goal

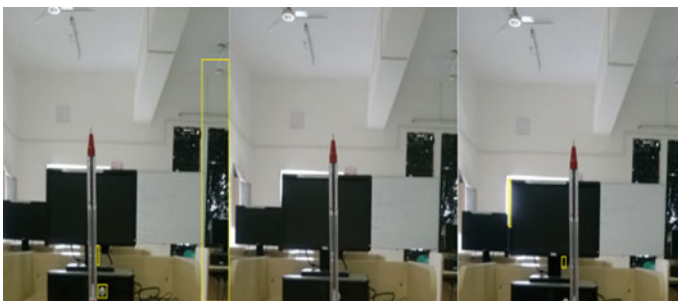


Fig. 1 Images of plane object. **a** Left camera image; **b** center camera image; **c** right camera image



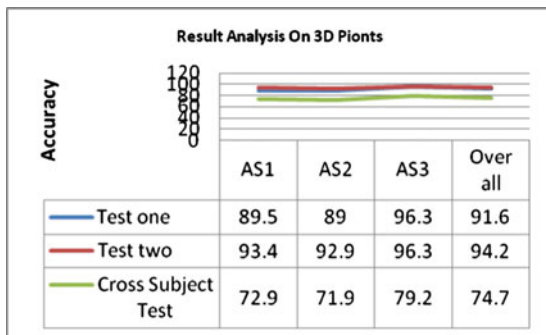
Fig. 2 Images of round object. **a** *Left* camera image; **b** *center* camera image; **c** *right* camera image

is to implement advance inexpensive methodology to control the growing instances of terrorism and socially unsuitable activities. Three-dimensional cameras are suitable choice to improve accuracy of a typical activity surveillance but it is reasonably expensive. This paper expressed two-dimensional camera networks to generate stereoscopic video. Carefully placing two-dimensional cameras in strategic location succeeding human eye visualization concept then accumulates all the video data under observation to generate stereoscopic video. This will increase the data to be processed which will be helpful to propose advance algorithm. This algorithm can be applicable either online or offline applications. In online, real-time numerous applications like person authentication, activity monitoring, and in offline to identify usual events which are common and general or unusual which are strongly restricted in public and secured areas. This will effectively reduce cost of hardware between 70 and 90 %. Our main intention is to use two-dimensional camera networks to generate stereoscopic images and achieve activity recognition in critical situations and to notice very tiny movement of coma patients [7, 6].

Second important method mentioned in this paper is to keep track on mental stress of coma patients. Researcher proves that the coma patients are physically unconscious but mentally conscious. So during medical treatment patient mental stress can be considerable. This is effective for faster recovery and better response for the treatment of the patient.

1.1 Review of Exciting Techniques

In the last decade, researcher in the field of image processing mainly focused on learning and recognizing action set from video sequence captured by two-dimensional cameras. The researchers emphasized on development of image quality by using high-definition cameras and improvement in algorithm of processing the two-dimensional images. It has been observed that two-dimensional cameras are sometimes less sensitive to minute vertical movements to its lenses.

Table 1 Graphical representation of result based on bag of 3D points

Also, it captures two-dimensional images so information collected is less with impact on accuracy.

Table 1 shows the graphical analyzed on bag of 3D points [1]. In this paper author proposed effective projection-based sampling method for sampling the bag of 3D points and he reached to 90 % accuracy by only sampling 1 % of samples. For experimental purpose the author used 80 3D points from downsampled depth map. From the 80 points, 45 % (i.e., 36 points) from xy-projection, 30 % (i.e., 24 points) from zy-projection, and 25 % (i.e., 20 points) from zx-projection form different action set.

In Table 1 AS1 is action set of horizontal arm wave, hammer, forward punch, tennis serves, AS2 is action set of high-arm wave, two-hand wave, forward kick, and AS3 is action set of side kick, jogging, golf swing. This result analysis is based on only 1 % points of three-dimensional images. The graph shows results of different experiments for different actions by different angles for identification of actions set. The result shows that it acquired 90 % accuracies. Experimental results had shown that over 90 % recognition accuracies are achieved by only sampling and processing 1 % of stereo points from the depth maps [1]. So this paper introduces advance techniques called surveillance using stereoscopic video.

2 Implementation Chart

Figure 3 shows implementation of working model. Here two-dimensional cameras network is stereo network as shown in Fig. 1. Process is sequentially described in proposed work. Figure 4 shows effective hardware which follows the nature of human eyes visualization by placing at a distance of 6.5 cm from each other which is strategy for camera network. As shown both camera lenses are perpendicular to the monitoring object and process ideally.

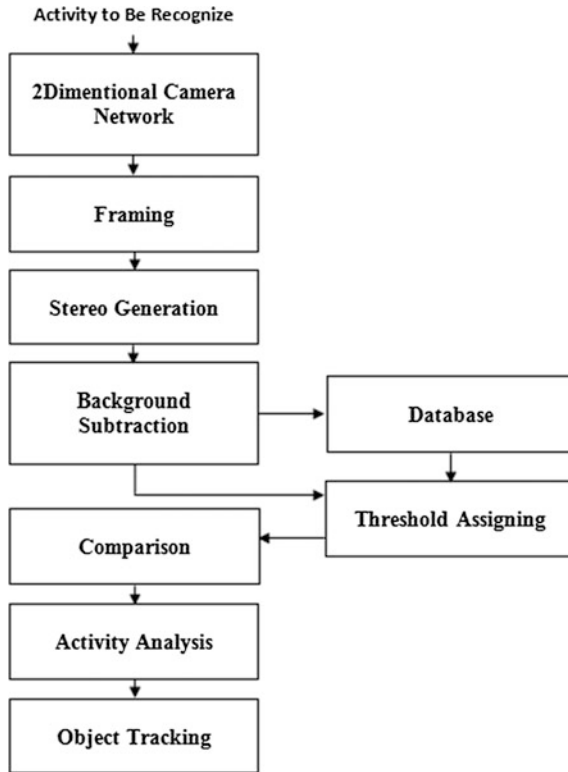


Fig. 3 Implementation chart of activity surveillance

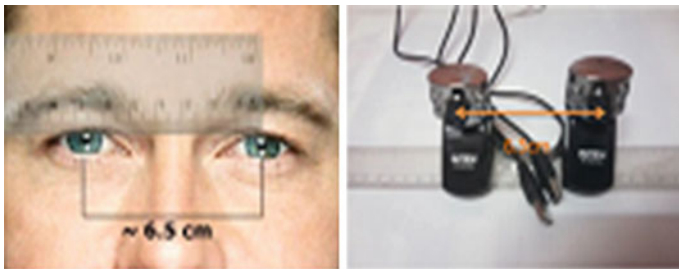


Fig. 4 Two identical two-dimensional cameras placed at distance 6.5 cm. The multicamera network is used [9]

3 Proposed Work

3.1 Framing

Figure 5 shows 16 sequential frames from both left and right cameras for 0.53 s taken using implemented experimental setup. The two cameras identically capture sequential frames which mean first frames of right and left cameras are captured at same movement by keeping distance of 6.5 cm horizontally.

$$\text{frame rate} = \text{mean}(1/\text{diff}(\text{time})) \tag{1}$$

Though camera frame rate is 30 fps, frame rate is calculated for simulation purpose to process next using MATLAB (Fig. 6).

3.2 Stereo Generation

Color image is consisting of three basic colors (red, green, and blue). Direct addition of two frames generates noisy stereoscopic frames. In proposed method red color is extracted from first frame obtained from left camera and cyan color which is



Fig. 5 Frames of left and right camera videos



Fig. 6 Generation of stereo frame

the combination of blue and green is extracted from first frame of right camera. Then frames are added in concatenate arrays along three dimensions.

$$\text{Stereo frame} = \text{cat}(\text{dim}, \text{red}, \text{blue}, \text{green}) \quad (2)$$

where cat = concatenate arrays and dim = dimension. Equation (2) is for stereo image generation. Here cat matrix is open and three frames (i.e., extracted red frame from left camera and extracted blue and green frames from right camera) at same instance are placed in concatenate manner forming single color frame. Same way sequential frames are added to generate stereoscopic video. By applying proposed method, prevention of noise generated due to direct mixing of color information is conceivable.

3.3 Threshold Assigning

Threshold value is specified for classifying moving and nonmoving elements. In case of coma patient moving elements are chest movement due to respiration. Threshold values are calculated on the basis of environmental changes and respiratory movement. By eliminating these values obtaining from experimental results threshold values are calculated. To increase accuracy of proposed method three different threshold values were calculated for red, blue, and green colors and sensitivity of system can be managed by changing threshold value.

$$\text{Threshold} = \sum_{i,j=0}^{m*n} \text{stereo}(i,j) \quad (3)$$

where $m * n$ is rows * columns and (i, j) are pixels values in stereo frames. "Equation (3) is used for threshold calculation to find changes in image and provide artificial intelligence" (Movement detection). The assigned threshold value can be adjusted depending upon sensitivity demands of applications.

3.4 Movement Tracking by Kalman Filter

The Kalman Filter algorithm is an optimized method for determining the best estimation of the moving state [8]. It is used for tracking the moving object. A nonlinear difference (or differential) equation is derived for the covariance matrix of the optimal estimation error. "Figure 7 shows Kalman process flow where square box represents matrix and ellipses represent normal distribution." F_k is the state transition model which is applicable to the previous state u_k , X_{k-1} , H_k is the observation model, B_k is the control input model which is realistic to the control

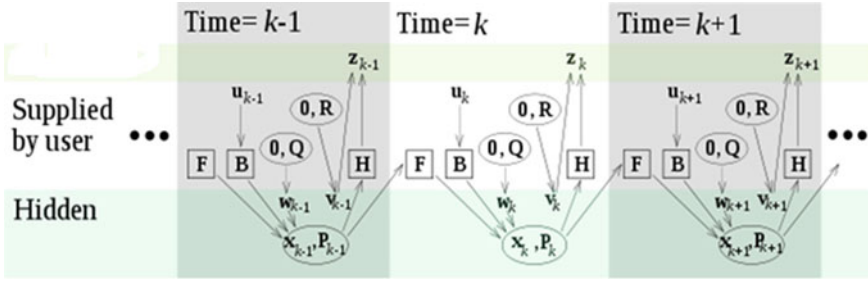


Fig. 7 Kalman filter internal state of a process

vector u_k , Q_k is the covariance of the process noise, and R_k is the covariance of the observation noise.

At time k a measurement Z_k of the true state X_k is defined as

$$Z_k = H_k \cdot X_k \cdot V_k \tag{4}$$

where state X_k is calculated by

$$X_k = F_k X_{k-1} + B_k u_k \tag{5}$$

Process noise W_k is calculated by

$$\text{Process Noise}(W_k) = N(0, Q_k) \tag{6}$$

Equation (6) is calculated by supposing to be drawn from a zero-mean multi-variate normal distribution with covariance Q_k and Gaussian White Noise V_k is

$$\text{Gaussian White Noise}(V_k) = N(0, R_k) \tag{7}$$

Step $X_0, W_1, W_2, \dots, W_3, V_1, V_2, \dots, V_k$, the noise vectors, and the initial state at each are all assumed to be mutually independent.

3.5 Stress Level Measurement

Human skin offers variable resistance to voltage and current. This resistance changes with the emotional stress of the mind. Literture survey shows effect of stress on human [10, 11]. The proposed method measures changes in mental stress resulting in changes in skin resistance. In the relaxed state, human skin offered resistance of 2 MΩ and when the emotional stress is too high it reduces to 500 kΩ or less. The reduction in skin resistance is associated to increased blood flow in nerves and during high stress permeability followed by the physiological changes. This increases the electrical conductivity of the skin. In proposed method, BiMOS

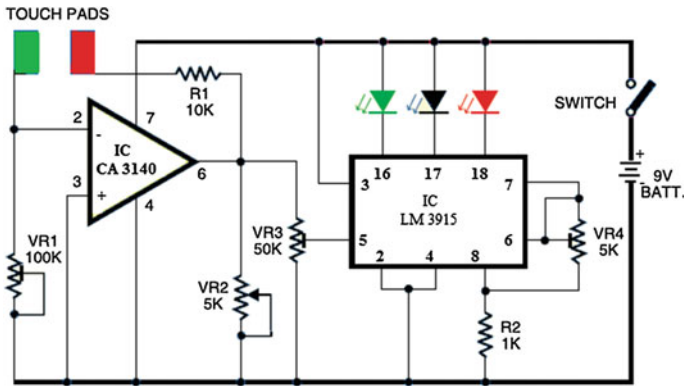


Fig. 8 Proposed hardware circuit diagram for mental stress measurement

operational amplifier with MOSFET inputs and bipolar output IC CA3140 is used. The gate-protected inputs have high impedance and can sense current as low as 10 pA used as a resistance-to-voltage converter whose output voltage changes according to changes in skin’s resistance. For better visual observation, an LED display and audio system are used. IC LM3915 is used to show a logarithmic display through LED indications (Fig. 8).

4 Experimental Result

Simulation Results

Figures 9 and 10 show critical movement recognition at night and day conditions, respectively, from continuous video frames captured on left and right direction. First, second, and third row images look the same but are captured by stereo

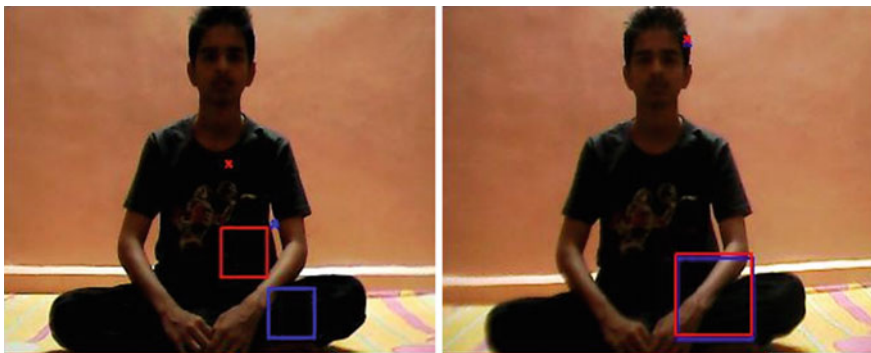


Fig. 9 Tracking minute movement from both angles using Kalman filter at night



Fig. 10 Tracking minute movement from both angles using Kalman filter during day and night circumstances

cameras. Red rectangles and blue crosses spotlight movements for a different camera. It covers and monitors a stereo view of interested object. Stereoscopic image provides additional information of body which is important to distinguish the action.

Red, blue crosses and rectangle in left and right column in some important result are obtained from evaluation of this project and referred papers. Stereoscopic image increases the amount of data to be processed which helps to design new advance algorithm. Compared to two-dimensional silhouette-based recognition error was half [1]. Movement detection accuracy will enhance with reduction in the cost. Optimized movement tracking is done by proposed method of Kalman filter.

Hardware Result

Figure 11 shows different stress conditions. After calibrating resistance of human body the hardware is configured for normal state of human body. The first images show that yellow LED glows showing normal state of body. If mental stress increases resistance between two electrodes decreases in the range of 2 M to 500 k Ω thus current flow between these two electrodes increases. The output voltage of IC 3140 increases. Depending upon stress next LEDs sequentially glow which is shown in second image. The glowing red LED shows high mental stress condition.

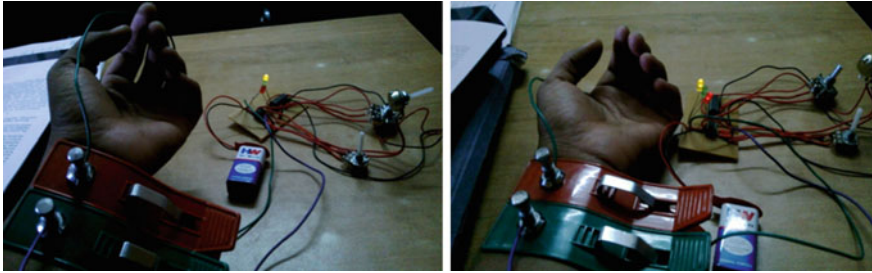


Fig. 11 Normal and high stressed condition of human

5 Future Scope

It has been noticed that one can increase quality of stereoscopic images by using high definition cameras also reduce effects of sensitive environment so as to gain accuracy. Response time can be minimizing for tracking the object in real-time and expand application areas. One can design a simulation based stress analysis system so as to achieve better accuracy.

I believe that two main influences should receive special consideration in future studies on stress in humans. The first influence is sex or gender and the second is age. Sex refers to the biological dissimilarities between males and females, whereas gender refers to the different roles (gender role and gender identity) that men and women may have during their lifespan.

6 Conclusion

A novel technique of movement detection is presented in this paper, which introduces new technique of stereo generation using two-dimensional cameras by following human binocular vision so as to increase amount of data to be processed. The finding highlights that increasing number of cameras not increases simulation cost. During stereo generation I extract color elements (hue and saturation) till luminance values are the same as original; hence this method does not produce any distortion. The said concept can be implemented using three-dimensional cameras but it is not cost-effective. 70–80 % cost reduction is possible using this technique. By proper arrangement one can produce depth in three-dimensional video.

Using this unique technique detection tracking of minute movement is possible using Kalman filter. Obtained result shows different movements tracked by stereo cameras minimizing limitation of two-dimensional cameras. Memory required for recording movement is suitably trimmed down.

The stress response nowadays can have negative consequences for brain development and mental health may have conferred the necessary tools to our

ancestors in prehistorical times for surviving in the presence of predators. A novel technique of stress detection is presented in this paper, which introduces new technique of mental stress measurement. Though coma patient is physically unconscious, he is mentally conscious so during uncomforted condition coma patient responses to stress. Stress may be of uncomforted condition of room, physical stress, medicine effect. A novel technique is really helpful for doctor in patient surveillance and faster recovery.

References

1. Li, W., Zhang, Z., Liu, Z.: Action recognition based on a bag of 3D points. In: IEEE Conference on Advances in Multimedia Research, NSW 2522, Australia (2010)
2. Lekhadass, T.D.V., Nirmal, J.H.: Human detection and tracking using image segmentation and Kalman filter. *Lama* 978-1-4244-4711-4/09, IAMA (2009)
3. Dedeoglu, Y.: Moving object detection, tracking and classification for smart video surveillance. A thesis submitted to the Department of Computer Engineering and the Institute of Engineering and Science of Bilkent University (2004)
4. Li, W., Zhang, Z., Liu, Z.: Expandable data-driven graphical modeling of human actions based on salient postures. *IEEE Trans. Circuits Syst. Video Technol.* **18**, 1499–1510 (2008)
5. Xu, L., Jia, J., Matsushita, Y.: Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 1744–1757 (2012)
6. Zhou, J., Hoang, J.: Real time robust human detection and tracking system. In: Proceedings of the 2005 Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05), 1063-6919/05 (2005)
7. Lin, D.T., Huang, K.-Y.: Collaborative pedestrian tracking and data fusion with multiple cameras. *IEEE Trans. Inf. Forensics Secur.* **6**, 1432–1444 (2011)
8. Kalman, R.E.: A new approach to linear filtering and prediction problems. Research Institute for Advanced Study, Baltimore, Md. *Trans. ASME–J. Basic Eng.* **82**(Series D), 35–45 (1960)
9. Taj, M., Cavallaro, A.: Distributed and decentralized multi-camera tracking. *IEEE Signal Process. Mag.* **28**(3) 46–58 (2011)
10. Ribeiro, P.C., Santos-Victor, J.: Human activity recognition from video: modeling, feature selection and classification architecture. In: HAREM-2005, International Workshop on Human Activity Recognition and Modelling, Oxford University, UK (2005)
11. Gardner, S.: Stress among prospective teachers: a review of the literature. *Australian J. Teacher Educ.* **35**(8), 18–28 (2010) Article 2, Griffith University

High Speed Network Intrusion Detection System Using FPGA

S. Anuraj, P. Premalatha and T. Gireeshkumar

Abstract Modern Network Intrusion detection needs a high-speed interface to analyze the incoming packet. Several network intrusion detection applications detect multiple strings in the payload of a packet by comparing it against predefined pattern set which requires more memory and computation power. To meet this, a dedicated hardware with high processing capacity can be placed at the port of incoming packets. Field Programmable Gate Array (FPGA) is the choice as it can be programmed easily and dynamically for parallel computing. Moreover, FPGA devices support at high-speed interface and are capable of providing better processing capability than other device; also it can be reprogrammed when it is needed. This paper proposes a new alternative approach to leaf attaching algorithm to improve the memory efficiency of algorithm.

Keywords Field programmable gate array (FPGA) · Network intrusion detection system (NIDS) · Networking algorithm

1 Introduction

The Security threats are increasing as the usage of Internet increases. Our system environment protection is becoming a major issue. After the invention of SNORT [1], lot of researches is occurring in network intrusion detection system (NIDS) to

S. Anuraj (✉) · T. Gireeshkumar
TIFAC-CORE in Cyber Security, Amrita Vishwa Vidyapeetham,
Coimbatore, India
e-mail: anurajsreenivasan@gmail.com

T. Gireeshkumar
e-mail: gireeshkumart@gmail.com

P. Premalatha
Department of Electronics and Communications, Amrita Vishwa Vidyapeetham,
Coimbatore, India
e-mail: rppremalathaa@gmail.com

improve the efficiency of NIDS [2]. The speed of the network is the other common factor which affects the performance of NIDS. A software program with strong dataset like SNORT can act as a good software NIDS, but in a network with data rate of gbits/sec, it cannot perform due to the lack of resources in the machine [3]. Delaying the detection of intrusions into the network would be an advantage to attacker to gain access to our system, so the IDS must be realistic and reliable. The intrusion detection is performed by matching the predefined dataset with incoming packets. Each packet contains a header and payload; header contains the information of Internet protocols, address of source, address of the destination, and port number [4] etc., whereas the payload field contains the data. The attacker plays in the payload field where the data lies, so a thorough check has to be done in that field with the predefined dataset. Since the methodology of attack and intrusion are changing day-by-day there is a need of updating the predefined dataset, it also increases the complexity of matching when the size of the predefined dataset increases. The introduction of a separate hardware can sort out problem of complexity of matching in software in heavy network traffic. A Field Programmable Gate array (FPGA) [5] can be used as a separate hardware unit for NIDS, Net-FPGA board [6] will act as PCI card and all incoming traffic will be redirected.

Aho-Corasick [7] algorithm is one of the dictionary-based algorithm [8] used for exact string matching [9]. Modified Aho-Corasick, leaf attaching [10] algorithm becomes a solution for memory inefficiency while implementing Aho-Corasick in a hardware platform [11]. This paper proposes a new alternative approach to Leaf Attaching Algorithm (LAA) to make it more memory efficient while implementing in Net-FPGA 1G board.

The paper is organized as follows. Section 2 contains the related work on intrusion detection system, string matching algorithm, and implementation of it in hardware. Section 3 describes the design and architecture. Results and discussion are briefed in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Related Works

SNORT [1], a packet logger for lightweight NIDS uses some of predefined dataset which is created with some known attack such as buffer overflow, port scans, etc. When new attacks are detected in a system, it is configured with SNORT and logged. Later on, this log will be incorporated in SNORT rule dataset. As a result the rule dataset becomes stronger. The SNORT architecture contains three subsystems, the packet decoder engine, and the detection engine, logging, and alerting system. The incoming packet received at ethernet interface is sent to the decoder that separates the header and payload information in the incoming packet and it also provides the capability of decoding protocols. The detection engine helps to increase the speed of the detection processing. The commonalities in rule dataset are consolidated into a single chain header. Individual detection is performed by separating signatures and kept it in chain option structure. The rule set is maintained

by setting it in two directional rule set. The down system contains list of common attributes in the packet header and up system contains the common payload keyword but when Martin [1] developed the paper on the detection engine was in developing stage. Logging and alerting subsystem is used to create a log of all incoming packets which are passing through the network and it would alert the system administrator when a suspicious activity is taken place. This logging alerting can perform only after the packet is decoded. The log which is created in the system can be read by any user.

A finite field multiple string matching machines [12] can be able to match the all possible combination of incoming string in automated transition state. It consists of two steps: one is to create a go-to function [1] and another is failure function [13]. In go-to function, the searching is done by traversing each and every pattern one by one and if traversing end with is leaf node then function gives a negative output, where as in failure function if traversing cannot find the searching pattern then it will again start with the root node. The searching complexity in finite state string matching is resolved using Binary Searching Tree (BST). Among BST, leaf attaching algorithm [10] helps to achieve it. In leaf attaching algorithm the complete dataset is sorted and saved in the form of binary tree for reducing the complexity of piercing inside the dataset. The pattern which has a value less than the root node is saved in the left node of the tree where as right side contains the value greater then root node. The pattern saving in the intermediate node of BST is pushed into leaf node of the tree using leaf pushing [10]. This procedure helps to reduce the complexity of performing a large dictionary-based pattern matching on hardware. Leaf pushing can disjoint the pattern from set by building a tree. Initially, a pre-prefix tree will be formed from input patterns then build it into the form of full tree by pushing all non-leaf patterns to its par patterns, then it will be merged to its own leaf node and hence, the space for saving non-leaf can be reduced. Lookup latency [14] in string matching does not depend on the length of pattern, but it depends on the number of patterns.

3 Design and Implementation

Hardware memory is a major issue while implementing a dictionary-based algorithm in hardware. To make more memory efficient, the leaf attaching algorithm (LAA) [10] is used. Considering all patterns, pattern with prefix is chosen. The prefix pattern is clubbed with its parent pattern instead of saving them separately. For example, consider two patterns, “net” and “network” from Table 1. The pattern “net” will be present in the parent node of “network”, so as per algorithm the pattern “net” in intermediate node will push toward its child node “network” so there will a free memory in intermediate node. To identify there are two patterns in leaf node the Match Status Vector (MSV) [10] is implemented. Initially MSV will be “00000000” when a pattern “net” of length three arrives in tree then MSV will be changed to “00100000” so while reaching pattern “network,” the corresponding

Table 1 Sample of input patterns

Sl no.	Pattern	Sl no.	Pattern
1	Board	9	Keyword
2	Face	10	Less
3	Faceless	11	Net
4	Inter	12	Network
5	Internet	13	Word
6	Key	14	Wordless
7	Keyboard	15	Work
8	Keyless	16	Workless

Table 2 Sample of output patterns after implementing LAA

Sl no.	Set of patterns	Saved pattern	MSV
1	Board	Board	00001000
2	Face, faceless	Faceless	00010010
3	Inter, internet	Internet	00010010
4	Key, keyboard	Keyboard	00100001
5	Key, keyless	Keyless	00100010
6	Key, keyword	Keyword	00100010
7	Less	Less	00010000
8	Net, network	Network	00100010
9	Word, wordless	Wordless	00010001
10	Work, workless	Workless	00010001

MSV will be “00100010.” The one in the third and seventh positions indicate that there are patterns with length three and seven. Table 2 shows the output after applying leaf attaching algorithm in the input pattern given in Table 1. The leaf attaching algorithm disjoints the pattern by considering patterns with prefix only. To make yet more memory efficient by decreasing the numbers of patterns, the prefixes as well as suffix of the patterns are considered. Table 3 shows the output patterns to be saved. In worst case maximum number of root nodes will be 26 (A–Z). Initially, a counter value is set with string length. Counter value reduces if there is match with each pattern. If search function reaches leaf node and counter value

Table 3 Sample of output patterns after implementing LAA with suffix

Sl no.	Saving pattern	MSV
1	Board	00001001
2	Face	00010000
3	Inter	00001000
4	Key	00100000
5	Less	00010011
6	Net	00100001
7	Word	00010010
8	Work	00010010

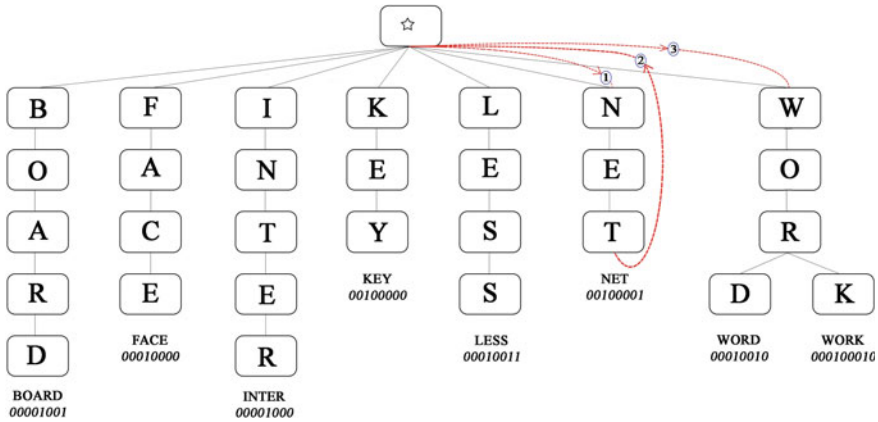


Fig. 1 Binary tree using LAA with prefix and suffix patterns

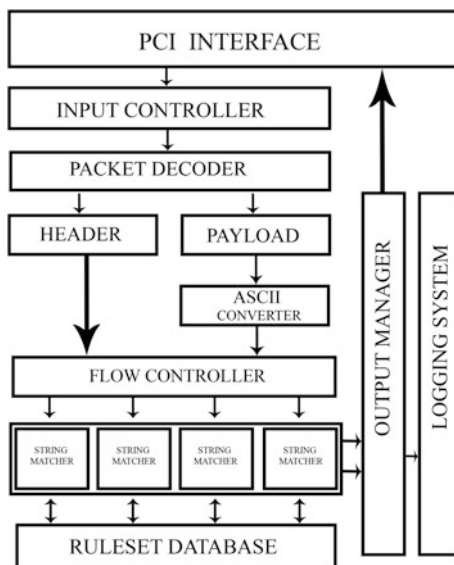
does not reach zero then SME starts to search the next node for the rest of the input pattern. Whenever the MSV is set with one at the position of string length, the algorithm detects it as the matched node.

Suppose we want to search the pattern “network” the counter value will be equal to seven. Figure 1 depicts that searching start with root node, from the root node it goes to the first node which has same ASCII value as the first letter in the pattern. Once the match occurs, the search is continued by checking the next root node. When searching reached at the leaf node and the match does not occur then pointer will point to root node with frailer function. In case of our example it fails with “net” and word “W” is not present in next node. So pointer points to root node and searches for node with “W”. In this case there is a “1” in the seventh position of MSV and counter start with seven we consider that the pattern is present.

3.1 Architecture

The architecture of overall system is explained in Fig. 2. There are four RJ45 interfaces present in Net-FPGA which are used to connect Cat 6 copper network cable to the board. Each port has a wire speed of 1 Gbits. There is eth0 inbuilt interface in every board which is used to connect RJ45 cable so, totally five interfaces are available. The real-time incoming packet reaching at the network interface will be buffered using an input controller. The header field contains the source address, source port; destination address as well as destination port numbers of packet where as payload of a packet contains the data which is taken care to identify the malicious packet. Packet decoder considers each packet from input controller and separates its header and payload information. Packets are fragmented into a fixed length since it varies in its length. The incoming packet is in the form of

Fig. 2 Architecture



hex but rule set are in the form of ASCII so, an ASCII converter single character encoding scheme is used to encode characters of incoming packet to a 7 bit binary form. For example, **0A** is converted to **00110000 01000001** and **123** converted to **00110001 00110010 00110011**. Header information is converted into its corresponding ASCII value and then it is passed to flow controller. The String Matching Engine (SME) [10] will perform a matching of incoming packet with predefined rule set. Rule set database contains SNORT [1] rule dataset in the form of binary search tree that are made from predefined dataset as there is a provision for the user to create his own rule set. The pre-defined dataset have specific rule for each previously known attack that defines what a packet want to do when a match occurs. That is if dataset defines a packet from a predefined data format having a chance for a known attack then it will create an alert to user as well as all the activities can be logged if the user needs. This written log can be read by anyone using that system with proper privileges. The flow controller act as master and distribute the work into different SME. Each SME perform searching in the rule set and if the search pattern is present it will create log to user, otherwise the packet will be passed toward the output manager. The output manager has same functions as an input controller have; it will buffer the output of SME and pass toward the internal network. Logging system is used to log all subspecies activities and all analysis of all packets as per the instruction of user. The log file is then saved in system in the form of tcpdump format which is binary.

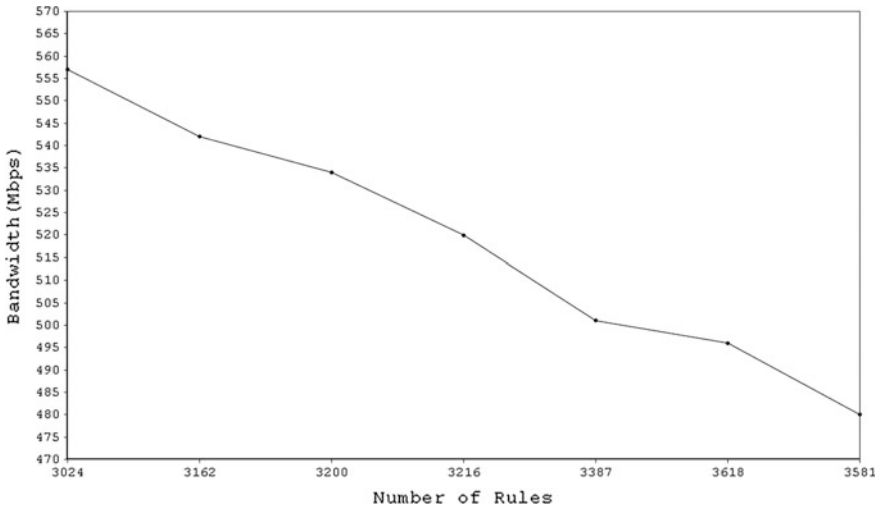


Fig. 3 Bandwidth versus rules set

4 Results and Discussions

A network interface card using Net-FPGA with a throughput of 1 Gbits/s is under development. Experimental results are shown in Fig. 3. It is clearly visible from the graph that when the number of rules increases the bandwidth decreases. The algorithm discussed in the paper helps in reducing the available number of rule set to be handled by Net-FPGA board. Thus, more bandwidth would be available in the network for dataflow. Finally, low-buffer complexity and high throughput is achieved.

5 Conclusion

In real time, huge amount of packet is needed to be compared and analysed during the implementation of an intrusion detection system. The leaf attaching algorithm plays a major role in the comparison of two strings. This paper optimizes leaf attaching algorithm which can disjoint a given set of patterns without increasing the number of patterns to accomplish an enhanced performance in Net-FPGA. The suffix and prefix thread reduces the size of saved dataset which in turn increases the performance of a multi-core processor Net-FPGA.

References

1. Roesh, S.M.: Snort-light weight intrusion detection for networks. In: Proceeding of LISA, vol. 99, pp. 299–238
2. Qingbo Wang and Viktor K Prasanna. Multi-core architecture on fpga for large dictionary string matching. In: FCCM'0, 17th IEEE Symposium on Field Programmable Custom Computing Machines, pp. 96–103, IEEE (2009)
3. NetFPGA-1G website. <http://netfpga.org/2014/#/systems/4netfpga-1g/details/>
4. Song, H., Lockwood, J.W.: Efficient packet classification for network intrusion detection using fpga. In: Proceedings of the 2005 ACM/SIGDA 13th International Symposium on Field-programmable Gate Arrays, pp. 238–245. ACM (2005)
5. Pontarelli, Salvatore, Bianchi, Giuseppe, Teofili, Simone: Traffic-aware design of a high-speed fpga network intrusion detection system. *IEEE Trans. Comput.* **62**(11), 2322–2334 (2013)
6. NetFPGA User Guide. <https://github.com/NetFPGA/netfpga/wiki/Guide>
7. Aho, A.V., Corasick, M.J.: Efficient string matching: an aid to bibliographic search. *Commun. ACM* **18**(6), 333–340 (1975)
8. Varghese, G.: *Network Algorithmics*. Chapman & Hall/CRC (2010)
9. Tuck, N., Sherwood, T., Calder, B., Varghese, G.: Deterministic memory-efficient string matching algorithms for intrusion detection. In: INFOCOM 2004 Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4, pp. 2628–2639, IEEE (2004)
10. Le, H., Prasanna, V.K.: A memory-efficient and modular approach for large-scale string pattern matching. *IEEE Trans. Comput.* **62**(5):844–857 (2013)
11. Arudchutha, S., Nishanth, T., Ragel, R.G.: String matching with multicore cpus: performing better with the aho-corasick algorithm. arXiv preprint [arXiv:1403.1305](https://arxiv.org/abs/1403.1305) (2014)
12. Hasib, S., Motwani, M., Saxena, A.: Importance of aho-corasick string matching algorithm in real world applications (2013)
13. Tumeo, A., Villa, O., Chavarra-Miranda, D.G.: Aho-corasick string matching on shared and distributed-memory parallel architectures. *IEEE Trans. Parallel Distrib. Syst.* **23**(3):436–443 (2012)
14. Rafeeq Ur Rehman: *Intrusion detection systems with Snort: advanced IDS techniques using Snort, Apache, PHP, and ACID*. Prentice Hall Professional, MySQL (2003)

Hawk Eye: A Plagiarism Detection System

Karuna Puri and Preeti Mulay

Abstract College students today use their smartphones or electronic media smartly to capture software code that are part of their curriculum and then circulate same code among the entire batch, leading to plagiarism. To avoid this kind of plagiarism an innovative plagiarism detection mobile system is needed. The proposed system will eventually help students to develop their coding skills and avoid cloning. The system uses Multi-Language OCR-Compiler Engine to convert the clicked snapshot into relevant text file and triggers appropriate compiler to compile the code. Then the system uses plagiarism detection algorithms that use the concept of tokenizing source code. Tokenizing makes difficult to replace the system by unnecessary comments, punctuations, or changing variables–methods names. Mobile applications like Viper, Plagiarisma.Net, Plagiarism CheckerX can be deployed to check the cloned code. Future extension includes applying modified concepts related to (cloned) flowcharts and also have desktop compatible version of this mobile system.

Keywords OCR · TTA · Plagiarism · Detection algorithms · Tokenizing · Viper · Plagiarism CheckerX · Plagiarisma.Net

1 Introduction

Plagiarism is becoming one of the major causes of concerns that have even spread its root across universities especially among students who are the future developers of software industry. One of the most common types of plagiarism encountered

K. Puri (✉) · P. Mulay
Computer Science and Engineering Department, Symbiosis Institute of Technology,
Lavale, Mulshi, Pune, Maharashtra 412115, India
e-mail: karuna.puri@sitpune.edu.in

P. Mulay
e-mail: preeti.mulay@sitpune.edu.in

among students is plagiarism of software codes done either by clicking snapshots of codes using their smartphones or copying already available codes online. So there is a need to devise a check mechanism for preventing such type of code cloning.

Hawk Eye: A Plagiarism Detection System It is an innovative mobile system to detect the code cloning [1]. This system uses Multi-Language OCR-Compiler Engine algorithm [2] to convert the clicked snapshots into the text format. Then the OCR Engine [2] would extract relevant keywords from code and identify appropriate programming language of extracted code, it will be compiled using inbuilt/online compiler [3].

The system uses plagiarism detection algorithms—Karp–Rabin [4] and Greedy String Tiling [5] algorithms that use the concept of tokenizing string to remove unnecessary comments, whitespaces, punctuations, or changing variables/methods names that frequently plagiarizer does in their system. Proposed flowchart for the system: Hawk Eye can be seen in Fig. 1.

The mobile version of plagiarism detection applications like Viper [6] (i.e., for Windows platform), Plagiarisma.Net [7, 8], and Plagiarism Checker [9] (i.e., for Android platform) can be used to detect plagiarism.

2 Multi-Language OCR-Compiler Engine

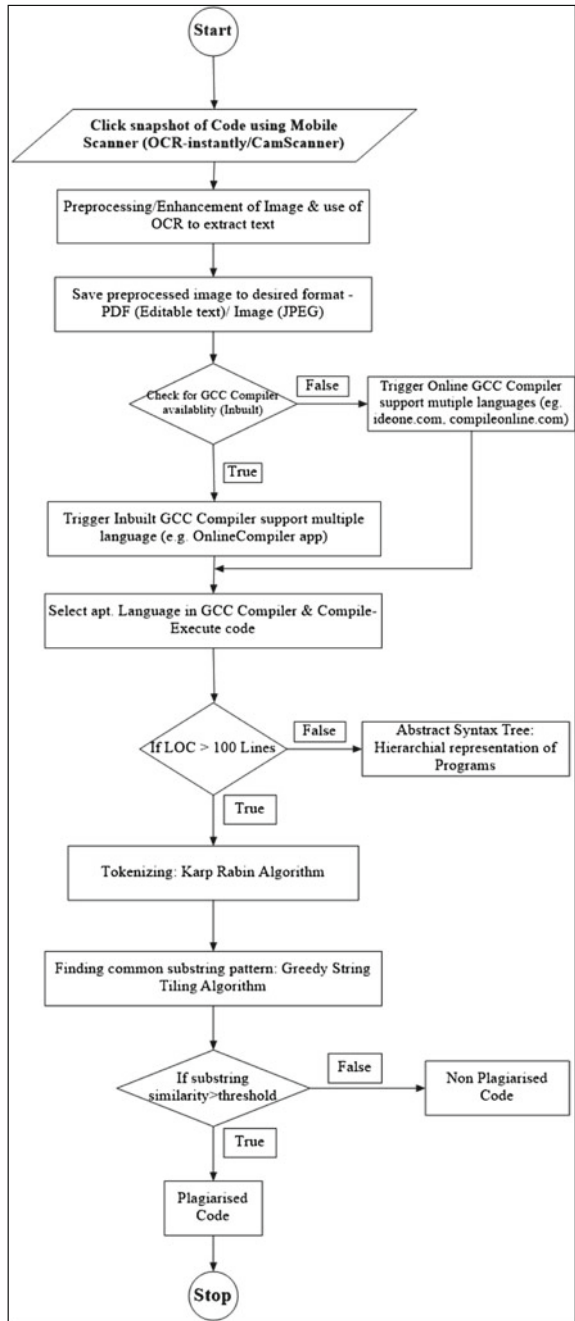
Multi-Language OCR-Compiler can be used to convert the clicked snapshots containing the relevant source code to editable text format (i.e., doc, txt). Then OCR from this converted text extract relevant program keywords removing all unnecessary details. Then an appropriate compiler [3] is triggered to compile the extracted keywords of code based on identified programming language.

2.1 OCR Engine Working

Smartphone's capability of scanning using various scanning apps like OCR Instantly, CamScanner [10], and many more with more advanced features being introduced every day, can be used in the proposed plagiarism detection system for extracting relevant text from clicked snapshot. The complete process can be summarized as:

1. Enhanced image (i.e., Image Preprocessing) for better image quality and reduction of noise as far as possible.
2. Use OCR to extract relevant text in editable form from captured image.
3. Save As (default options available—PDF/JPEG Format) in appropriate format.
4. Share (as available with scanners like OCR Instantly) using various means instantly just with one click.

Fig. 1 Flowchart for proposed system: Hawk Eye



This concept of OCR can be further extended to IWR [11] technology which is used for recognition of handwritten words that can be used for detection of handwritten plagiarized codes by students.

OCR: On Device or in Cloud

Before deciding on an OCR library [12], it needs to be taken care, where the OCR process will take place: on the smartphone/desktop device or in the cloud. Each approach has its advantages.

2.2 Compiler Engine Working (Triggering Appropriate Compiler)

The source code is compiled either using inbuilt system compiler or online compiler [3] using Transport Triggered Architecture [13] concept or GCC compiler [14] (supports multiple languages) to get a compiled source code.

3 Plagiarism Detection Algorithms

3.1 Abstract Syntax Tree

An Abstract Syntax Tree or AST [1, 15] is a hierarchical representation of a program. Each node represents a programming language construct and its children are the parameters of this construct. The nodes of an AST [15] can be mathematical operators, function calls, or other programming structures, the leaves are variables or constants. Compilers perform optimizations on AST before generating lower level code because of this property; AST can be used in plagiarism detection.

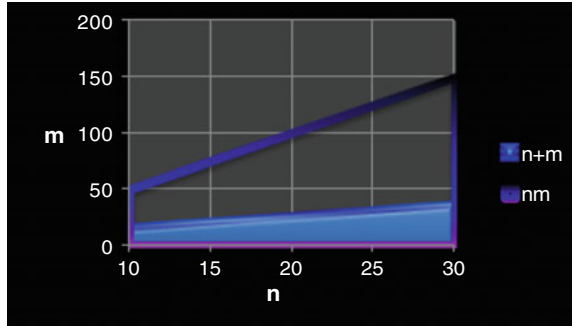
Limitations of AST

AST [15] based systems are often vulnerable to code insertion or reordering. Another difficulty is the complexity of the comparison between the trees.

3.2 Tokenizing String-Based System

Consider the program as a normal text. The preprocessing phase removes all comments, white spaces, and punctuations, and renames variables into a common token. Then a string sequences comparison is performed. It performs a string-based comparison using the Karp–Rabin algorithm [16, 17]. This algorithm uses concept of hash function which can compute hash value of i th k gram from $(i - 1)$ th k gram.

Fig. 2 Complexity of Karp–Rabin algorithm $f(n)$, for $m = 5$



If a k gram $(C_1 \dots \dots C_k)$ is considered as a k -digit number in base b , lets its hash value $H(C_1 \dots \dots C_k)$ be a number:

$$C_1 * b^{k-1} + C_2 * b^{k-2} + \dots + C_{k-1} * b + C_k \tag{1}$$

Thus,

$$H2(C_2 \dots \dots C_{k+1}) = H(C_1 \dots \dots C_k) - C_1 * b^{k-1} * b + C_{k+1} \tag{2}$$

This method helps in computing all hash values in linear time. Hash value comparison is basically used to optimize comparison of two sets of data.

Karp–Rabin algorithm has complexity [17] of $O(nm)$ where n is length of text, while m is length of pattern. However, in practice it is considered that complexity of Karp–Rabin algorithm is $O(n + m)$ which makes it bit faster as can be seen in Fig. 2.

Alternatively systems can also use parsing of source code and transform the elements into tokens. The main advantage of tokens is that they discard all unnecessary information such as variables or methods names. Therefore, token-based systems are insensible to “search and replace” changes. After tokenizing the program, the strings of tokens are compared two-by-two to find common substrings. The most famous token-based softwares are JPlag and YAP3 [18]. They both use the “Greedy String Tiling” algorithm [5]. Due to the properties of this algorithm they are robust against code reordering.

4 Plagiarism Detection Tools

Content analyzed for plagiarism using different plagiarism detection tools: Smartphone’s capability of scanning using various scanning apps like OCR Instantly, CamScanner, and many more with more advanced features being introduced every day, can be used in the proposed plagiarism detection system for extracting relevant text from clicked snapshot. Smartphone scanner scans the

clicked image, preprocesses it to enhance its picture quality using various tools, and removes all noise thereby resulting in enhanced image with a better quality picture. On this enhanced image OCR functionality can be applied to extract relevant text from captured image.

Then the image can be saved in pdf/jpeg format (CamScanner), also many of these apps (like OCR Instantly) allow to share the extracted text using various means like email, notepad (available in smartphone), social networks (WhatsApp Messenger, Face book, Twitter, etc.)

4.1 *Plagiarism CheckerX (Android Platform Mobile) [9]*

See Tables 1, 2 and Fig. 3.

4.2 *Plagiarisma.Net (Android Platform) [7, 8]*

See Tables 3 and 4.

Table 1 Detected results of Plagiarism CheckerX for analyzed content

Item No.	Content	Sources	Duplicate
1.	Smartphones capability of scanning using various scanning apps like OCR Instantly, CamScanner, and many more with more advanced features being introduced every day	Empty	0
2.	Can be used in the proposed plagiarism detection system for extracting relevant text from clicked snapshot	http://www.comp.nus.edu.sg/~tancl/InformationMiningExtraction.htm	25
3.	Smartphone scanner scans the clicked image, preprocesses it to enhance its picture quality using various tools, and removes all noise thereby resulting in enhanced	Empty	0
4.	Image with a better quality picture. On this enhanced image OCR functionality can be applied to extract relevant text from captured image	http://www.seobythesea.com/2008/01/google-on-reading-text-in-images-from-street-views-store-shelves-and-museum-interiors/	18
5.	Then the image can be saved in pdf/jpeg format (CamScanner), also many of these apps (like OCR Instantly) allow to share the extracted text	https://play.google.com/store/apps/details?id=co.kr.generic.ocr	10

Table 2 Plagiarism report of Plagiarism CheckerX

Date	Friday, March 06, 2015
Statistics	12 words plagiarized/128 total words
Sources	More than four sources identified
Remarks	Low plagiarism detected—your document needs optional improvement. Similarity found: 9 %

Fig. 3 Summary report of Plagiarism CheckerX

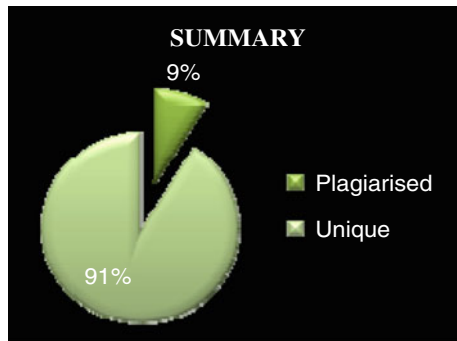


Table 3 Detected results of Plagiarisma.Net for analyzed content

Result	Query	Domains (cached links)
Unique	Smartphone scanner scans the clicked image, preprocesses it to enhance its picture quality using various tools, and removes all noise thereby resulting in enhanced image with a better quality picture	–
Unique	On this enhanced image OCR functionality can be applied to extract relevant text from captured image	–
Unique	Smartphones capability of scanning using various scanning apps like OCR Instantly, CamScanner, and many more with more advanced features being introduced every day, can be used in the proposed plagiarism detection system	–
Unique	Then the image can be saved in pdf/jpeg format (CamScanner), also many of these apps (like OCR Instantly) allow to share the extracted text using various means like email	–

Table 4 Plagiarism report of Plagiarisma.Net for Google–Babylon–Yahoo and Google Scholars–Google Books

Statistics	Total 873 chars, 132 words, 4 unique sentences
Remarks	100 % originality

5 Related Work

Winnowing algorithm [19, 20] is a method of comparison process based on documents fingerprinting. The concept of fingerprinting consists of obtaining a subset of a document in order to process the comparison on a smaller set of data. The fingerprint of a document is composed by hash values of k -grams.

Baxter et al. [1] present a plagiarism detection system using Abstract Syntax Tree designed for large software. It performs comparison using hash values of subtrees based on the assumption that when reusing code, the programmer will do small modifications that appear only near the leaves of the tree. Therefore, only parts of the tree bigger than a specified threshold are considered. It computes the similarity for all subtrees and spots those with a similarity above a specified threshold. This implementation works well for detecting clones in source code.

Viper detection [6] mechanism has very easy interface and can find code cloning and highlights the area of plagiarism.

Plagiarism Checker [9] does scanning of file doc text and tells the area of plagiarism.

6 Conclusions

Hawk-Eye proved to be a valuable software for teachers to evaluate code cloning done by students in lab as part of their curriculum. This mobile system uses a Multi-Language OCR-Compiler Engine algorithm to convert programs from several programming languages into text format and trigger the relevant programming language compiler. The plagiarism detection is then performed on these extracted keywords using plagiarism detection algorithms—Karp–Rabin and Greedy String Tiling algorithms that take into account the concept of tokenizing strings. Also mobile apps like Viper (Windows platform) and Plagiarism Checker (Android platform) can be deployed to check plagiarized codes.

7 Future Works

Future extension to this mobile system can be, applying modified concepts (like performing checks on symbols used, kind of operators, operands, variables, etc. used in flowchart) related to (cloned) flowcharts. A desktop compatible version of this system is very difficult to execute lengthy codes on mobile; also desktop version will be more user friendly for an individual to do evaluation compared to evaluation on mobile. Development of multilingual parser system supporting different programming languages will not only help to distribute the load of compiler but would speed up the compilation process as well. Concept of multithreading can

also be used in comparison job that makes the process relatively faster in terms of computation time, thereby improving the performances of the application, particularly on the computers generally equipped with multi-core processors. Individual software industries can also deploy Hawk-Eye in their respective firm as a check mechanism for code cloning by developers.

References

1. Baxter, I.D., Yahin, A., Moura, L., Sant'Anna, M., Bier, L.: Clone detection using abstract syntax trees. *Proc. Int. Conf. Soft. Maintenance*, **98**, 368–377 (1998)
2. Free Online OCR—Convert scanned PDF and Images to word, JPEG to Word. <http://www.onlineocr.net/>
3. Online Compile. <http://ww38.onlinecompile.com/>
4. Rabin–Karp Algorithm. http://en.wikipedia.org/wiki/Rabin%E2%80%93Karp_algorithm
5. Wise, M.J.: Running Karp-Rabin Matching and Greedy String Tiling. Basser Department of Computer Science University of Sydney NSW 2006 report number 463. <http://sydney.edu.au/engineering/it/research/tr/tr463.pdf> (1993)
6. Viper Plagiarism free Download. <http://en.softonic.com/s/viper-plagiarism>, Viper—Anti Plagiarism Scanner Free Download by ScanMyEasy. http://www.freedomdownloadcenter.com/Utilities/Misc__Utilities/Viper___The_Anti_plagiarism_Scanner.html
7. Plagiarisma.Net: Plagiarism Checker. <http://plagiarisma.net/>
8. Detect Plagiarism Free with Plagiarisma.Net || Free Software. <http://www.ilovefreesoftware.com/26/windows/plagiarisma-net.html>
9. Plagiarism Checker Android (Download). <http://plagiarism-checker.en.softonic.com/android>, Check for Plagiarism on Web for Free. <http://www.plagiarismchecker.com/>
10. Cam Scanner—Phone PDF Creator. <https://play.google.com/store/apps/details?id=com.intsig.camscanner&hl=en>
11. Intelligent Character Recognition Software. <http://www.cvisiontech.com/ocr/text-ocr/intelligent-character-recognition-software.html?lang=eng>
12. Comparison of optical character recognition Software. http://en.wikipedia.org/wiki/Comparison_of_optical_character_recognition_software
13. Transport triggered Architecture. http://en.wikipedia.org/wiki/Transport_triggered_architecture
14. GNU Compiler Collection. http://en.wikipedia.org/wiki/GNU_Compiler_Collection
15. Poongodi, D., Tholkkappia Arasu, G.: An automatic method or statement level plagiarism detection in source code using abstract syntax tree, Research Scholar, Manonmaniam Sundaranar University, Tirunelveli
16. Karp, R.M., Rabin, M.O.: Efficient randomized pattern-matching algorithms. *IBM J. Res. Dev.* **31**(2), 249–260 (1987)
17. Karp-Rabin Algorithm. <http://www-igm.univ-mlv.fr/~lecroq/string/node5.html>
18. Wise, M.J.: YAP 3: improved detection of similarities in computer program and other texts. In: *Proceedings of the Twenty-seventh SIGCSE Technical Symposium on Computer Science Education*. Philadelphia, Pennsylvania, United States, 15(17), pp. 130–134 (1996)
19. Aiken, A., et al.: Moss: A System for Detecting Software Plagiarism. University of California–Berkeley. See www.cs.berkeley.edu/aiken/moss.html (2005)
20. Schleimer, S., Wilkerson, D.S., Aiken, A.: Winnowing: local algorithms for document fingerprinting. In *SIGMOD '03: Proceedings of the 2003 ACM SIGMOD International Conference on Management of data*, pp. 76–85. ACM, New York, NY, USA (2003)

A Novel Action Descriptor to Recognize Actions from Surveillance Videos

T. Pradeepa, S. Abirami, M. Sivarathinabala and S. Murugappan

Abstract Due to the increased application in the area of human action detection, automatic capture and action analysis becomes a great research area. This paper provides a novel method to detect human action by considering features from the positive space and negative space region. Usually, in the literatures features have been considered from the positive space of the subject. Positive space features alone cannot provide solutions to the challenges such as occlusion and boundary variations. Therefore, in this method we have also considered the surrounding regions of the subject along with the positive space features. Initially, the input video has been segmented using background subtraction. Then the features are extracted from both positive and negative space of the subjects. Later, action descriptor has been defined for each pose of an action and proposed a new way for detecting number of cycles required to describe an action. Later, nearest neighbor classifier has been applied to classify the actions. The proposed system is evaluated using Weizmann dataset, KTH dataset and the results seem to be promising.

Keywords Positive space features · Negative space features · Action descriptor · Activity recognition

T. Pradeepa (✉) · S. Abirami · M. Sivarathinabala
Department of Information Science and Technology,
College of Engineering, Anna University, Chennai, India
e-mail: pradeepa026@gmail.com

S. Abirami
e-mail: abirami_mr@yahoo.com

M. Sivarathinabala
e-mail: sivarathinabala@gmail.com

S. Murugappan
School of Computer Science, Tamil Nadu Open University, Chennai, India

1 Introduction

Human activity recognition has been one of the interesting and challenging topics in the field of computer vision and pattern recognition. This research idea has been motivated by many applications such as surveillance systems, video browsing, and human–computer interfaces (HCI) design. In the past two decades, a significant amount of research has been done in the area of human activity recognition [1–6] using a sequence of 2D images. Progress in this field has been driven by a multitude of potential applications which rely on interpreting human motions.

Action recognition is difficult due to the noise in the captured action data, the high-dimensional representation of action sequences, and the existence of highly similar actions. Most of state-of-the-art methods solve these problems in three steps such as feature extraction, feature refinement (e.g., via dimension reduction, clustering or feature selection), and pattern classification. In this paper, we especially focus on the feature extraction process, and proposed a method for extracting features from both the positive and negative space region of actions to achieve high accurate action recognition with lower computational complexity.

The paper has been organized as follows: Sect. 2 gives the detailed description of related works. Section 3 gives the overview of our proposed action descriptor. Section 4 explains our preprocessing method whereas Sect. 5 describes the feature extraction techniques. Section 6 details about classification and Sect. 7 illustrates results and discussion. Finally, Sect. 8 concludes the paper.

2 Related Works

For the past two decades, a significant body of research literature has been contributed, proposing and/or investigating various methodologies for human activity recognition [1–6] from video sequences. Bobick and Davis [7] proposed temporal templates such as motion energy images and motion history images to recognize actions in terms of human movement. These approaches have limitations such as occlusion and high complex articulated shapes. Chaaaraoui et al. [8] proposed an adaptive Human Action Recognition with a bag of key poses by means of an evolutionary optimization method. The adaptive model learns the human actions over time through a bag of key poses. Detecting the action using silhouette-based method [9] is popular because it is robust to noise and is easy to extract the features. Silhouette-based method is of two types [10]: implicit and explicit methods, in which explicit method action has sequence of poses and implicit method recognizes actions simply and efficiently without any prior knowledge.

Ikizler and Duygulu [11] proposed a novel pose descriptor, in which Histogram-of-Oriented-Rectangles (HOR) has been used for representing and recognizing human actions in videos and this HOR is represented by rectangular patches that are extracted from the human silhouette. The drawback of this method occurs because of the imperfect silhouettes, the human actions have not been

identified properly. Kushwaha et al. [12] proposed a rule-based approach to recognize human activities. In this, rule-based classifier is applied to classify the human activities such as running, walking, bending, boxing, jogging, etc. This method is slow due to the computational complexity. Blank et al. [13] proposed a method to deal with volumetric space-time shapes that are induced in the human actions. This method provides the solution to the Poisson equation to extract various shape properties. The drawback of the system is that it considers only the space-time shape and ignores the intensity information inside the shape.

The motion-based features can predict the significant moving direction of the body. On the other hand, shape-based features can depict the pose information of the body but failed to detect the movement of the body parts from the silhouette region. To solve this problem, a novel action descriptor has been proposed here with improved motion features fused with the shape information that are extracted from negative space of the silhouettes. This new action descriptor has been provided as input to the nearest neighbor classifier to recognize the actions successfully. The main contribution of this work lies in the feature extraction technique with enhanced motion features and negative shape features, which results in a new action descriptor to describe the action in a better way.

3 Novel Action Descriptor

In this work, features have been extracted from the negative space as well as from the positive space which are used for action detection. The original silhouette of the person has been considered as positive space and negative region [14] means the space between the canvas and the silhouette. Canvas is defined using the bounding box and the space between the silhouette and the bounding box is termed as the negative space. The negative space can be described by simple shapes, e.g., triangles and quadrangles. Our system describes poses by low-dimensional and computationally efficient feature vectors. Each action descriptor (AD) describes an action with low-dimensional features. This implies that AD does not require dimensionality reduction, whereas most of the other implicit methods require a significant amount of processing time for dimensionality reduction. Additionally, we propose a method based on signal processing technique to determine the cycle length.

Our proposed system has been summarized in the Fig. 1. The input to our system consists of sequence of images of a particular action. The input image is segmented using background subtraction. Then the speed feature for each action is determined by taking the displacement of silhouettes of poses frame by frame. Optical flow algorithm is applied on the positive space in order to determine the motion of the silhouette.

Next to this, the negative space region is captured and subsequently features are extracted using the following steps: 1. Partitioning the negative space into smaller regions, 2. Extracting the positional features, 3. Extracting the shape-based features for each region, 4. Framing the new action descriptor using extracted features, and 5. Recognizing the actions under different datasets.

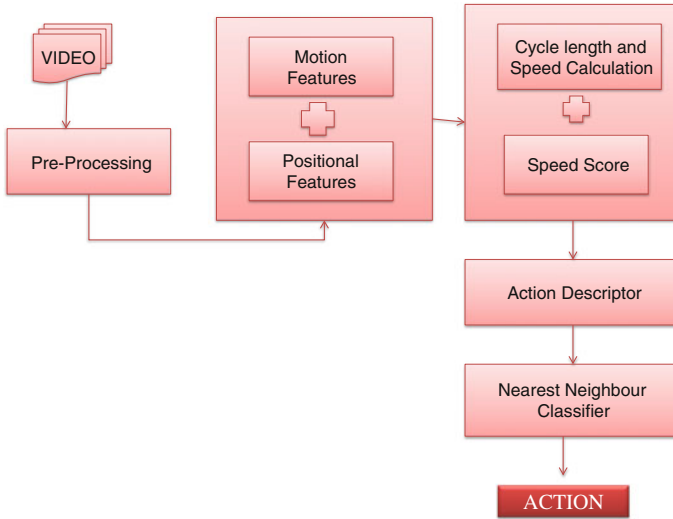


Fig. 1 Proposed action descriptor

4 Preprocessing

The input video frames are segmented using background subtraction [15] which is a common approach that identifies the moving objects from the portion of the video frame that differs significantly from the background model. Consider an image as background and take the frames obtained at the time t , compare with the background image. Here segmentation of the objects has been carried out by using image subtraction technique for each pixel obtained at the time t , and subtract it with the corresponding pixels at the same position on the background image.

5 Feature Extraction

Here, features are extracted from positive and negative spaces.

5.1 Motion Features (Positive Space)

Here, motion features are extracted from the whole silhouette of the person. Optical flow [16] is a technique to find the apparent motion of objects, surfaces, and edges in a visual scene. Predicting motions from one frame to next frame is an important part in recognizing human actions. Motion estimation is a difficult step when the

object is shapeless, in the presence of shadow of the object and motion blur. The target person has been represented in contour and their motion has been estimated in each and every frame. The motion of a person has been computed here using optical flow technique [16]. The optical flow vectors u and v are calculated when computing the motion of the person in each and every frame.

$$u = \frac{dx}{dt} \text{ and } v = \frac{dy}{dt} \tag{1}$$

Speed [14] is calculated by taking the displacement of the centroid of the silhouette of the person in each frame. The speed is calculated by taking the displacement of the centroid of a silhouette along the X -axis of consecutive frames divided by the time required for that displacement. Fuzzy membership functions are generated which are then employed to calculate speed scores.

5.2 Positional Features (Negative Space)

Here, features are extracted for each pose from its negative region [10, 14, 17] (surrounding to the subject) by partitioning the negative space into small regions. The negative space and positive space are shown in Fig. 2. The region partitioning process has been performed, where location and shape-based information from negative space region of the poses are extracted to describe each pose. This negative space is divided into small regions by using line growing process. Region partitioning has been shown in Fig. 3.

Feature vector of each pose is divided into 14 blocks which represent the 14 anchoring points. Each block is divided into six cells which represent the six shape-based features. In a negative space region these six features are assigned to a block corresponding to the anchoring point of that region. For example, anchoring point for region “A” is 1 so that the six features are assigned to block 1. Gray scale color of each cell corresponds to the feature value (1 for white and 0 for black). Dimensionality of the feature space is $14 * 6 = 84$. Cycle length of an action describes the minimum number of frames required to complete an action. Cycle length is calculated from the feature vector by using the frequency tracking algorithm.

Fig. 2 Negative space region

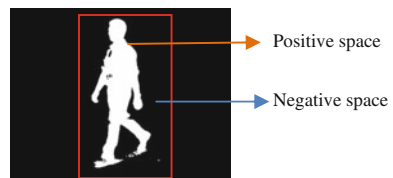
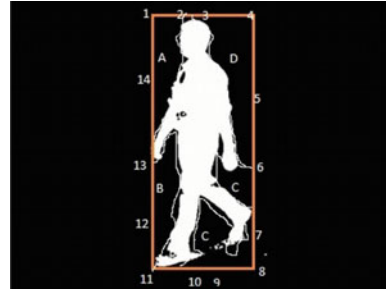


Fig. 3 Region partitioning



Positional features [14] of the negative space are obtained by taking 14 anchoring points in the bounding box of each pose. For example, anchoring points for regions A, B, C, D, and E are points 1, 12, 9, 5, and 7, respectively. Region-based features are extracted to describe the shape of the negative region such as area, eccentricity, orientation, rectangularity, horizontal and vertical lengths of the bounding box that are considered in this work. These six features are used to describe a region, because these features are effective in approximating triangles and quadrangles. The action descriptor is generated by combining the feature vector of each pose and the cycle length of the action which is computed from the feature vector. Here, positional and region-based features are combined together to form the feature vector.

6 Classification

Here, nearest neighbor classifier is used to identify the input action type [18]. From the input action sequence, we form the ADs for the sequences. Later, a distance metric such as Euclidean distance has been used to measure the distance of the training sets (generated in training phase) with training set.

7 Results and Discussion

The implementation of this novel action descriptor has been done using MATLAB (Version2013a). MATLAB is a high performance language for technical computing. The input videos are taken from Weizmann dataset [19] and KTH datasets [20]. The proposed algorithm has applied and tested over many different test cases.

Figure 4 shows the waving and running silhouette and Fig. 5 shows the optical flow vector for the bending action. The accuracy of different types of ADs for Weizzman dataset and KTH dataset is shown in Fig. 6. Here, AD type 1, action

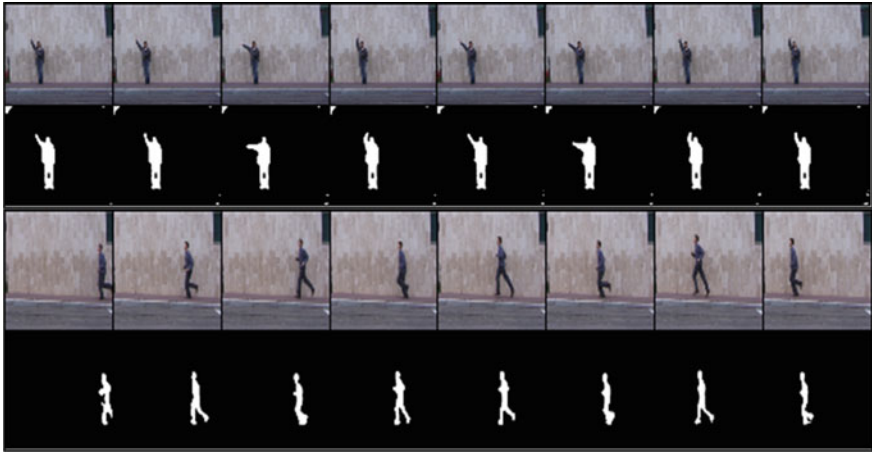


Fig. 4 Waving and running silhouette

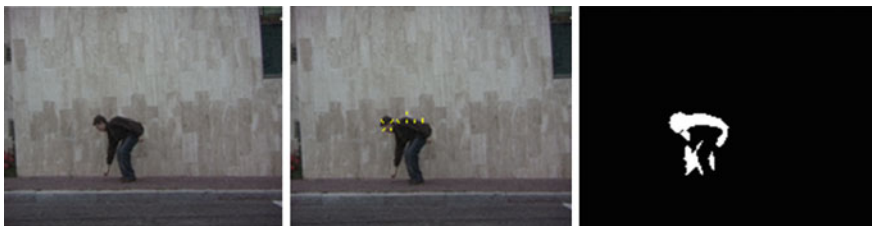
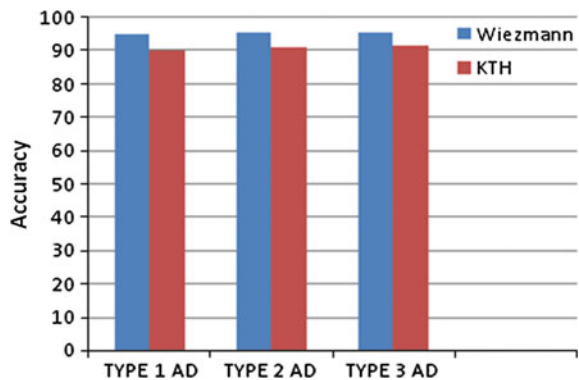


Fig. 5 Optical flow vector for bending action

Fig. 6 Classification accuracy



descriptor has been framed only for motion feature, i.e., Positive space. AD type 2, action descriptor has been framed only for positional features, i.e., Negative space. AD type 3, which has been framed by the combination of motion and positional

Table 1 Comparison of proposed method with other similar works

Method	Accuracy (%)	Feature extraction time	Classification time (s)
Proposed method	95.56	4.5 s	1.3
Fathi and Mori [21]	98	–	39.5
Rahman et al. [10]	98.2	4.6 s	35
Bregonzio et al. [22]	95.2	12.8 s	0.37
Liu and Yuen [23]	97.2	7.45 s	628

features, gives best result. The proposed method has been compared with other similar works in terms of accuracy, feature extraction time, and classification time that have been shown in Table 1.

8 Conclusion

In this paper, we have proposed a novel action description framed by the combination of positive space and negative space features. In this method the shape-based action recognition features have been considered from the surrounding region of the subject. The feature vector combines the motion feature and positional feature. The feature vector, motion information, and cycle length are altogether combined to form the action descriptor for an action. The proposed system shows high recognition accuracy rate. The feature extraction method is highly efficient and is faster compared to other feature extraction techniques. The classification using KNN makes the classification more accurate and the time also gets reduced. In future work, this system can be extended to recognize human interactions also.

References

1. Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review. *ACM Comput. Surv.* **43**, 1–43 (2011)
2. Sadek, S., Al-Hamadi, A., Michaelis, B., Sayed, U.: An efficient method for real-time activity recognition. In: *Proceedings of the International Conference on Soft Computing and Pattern Recognition*, Paris, pp. 7–10 (2010)
3. Poppe, R.: A survey on vision-based human action recognition. *Image Vis. Comput.* **28**, 976–990 (2010)
4. Gowsikhaa, D., Manjunath, Abirami, S.: Suspicious Human activity detection from Surveillance videos. *Int. J. Internet Distrib. Comput. Syst.* **2**(2), 141–149 (2012)
5. Gowshikaa, D., Abirami, S., Baskaran, R.: Automated human behaviour analysis from surveillance videos: a survey artificial intelligence review. doi:[10.1007/s10462-012-9341-3](https://doi.org/10.1007/s10462-012-9341-3) (2012)
6. Gowsikhaa, D., Abirami, S., Baskaran, R.: Construction of image ontology using low level features for image retrieval. In: *Proceedings of the International Conference on Computer Communication and Informatics*, pp. 129–134 (2012)

7. Bobick, A.F, Davis, J.W.: The recognition of human movement using temporal templates. *Proc. IEEE Trans. Pattern Anal. Mach. Intell.* **23**(3), 257–267 (2001)
8. Chaaraoui, A.A., Climent-Pérez, P., Flórez-Revuelta, F.: Silhouette-based human action recognition using sequences of key poses. *Pattern Recogn. Lett.* **34**(15), 1799–1807 (2013)
9. Sivarathinabala, M., Abirami, S.: Motion Tracking of Humans under Occlusion using Blobs. In: *Advanced Computing, Networking and Informatics*, vol. 1. Smart Innovation, Systems and Technologies, vol. 27, pp. 251–258 (2014)
10. Rahman, S.A., Leung, M., Cho, S.-Y.: Human action recognition by extracting features from negative space. In: Maino, G., Foresti, G. (eds.) *International Conference on Image Analysis and Processing*, pp. 29–39 (2011)
11. Ikizler, N., Duygulu, P.: Histogram of oriented rectangles: a new pose descriptor for human action recognition. *Proc. Int. Conf. Image Vis. Comput.* **27**(10), 1515–1526 (2009)
12. Kushwaha, A.K.S., Prakash, O., Khare, A., Kolekar, M.H.: Rule based human activity recognition for surveillance system. In: *Proceedings of International Conference on Intelligent Human Computer Interaction*, pp. 1–6 (2012)
13. Blank, M., Gorelick, L., Shechtman, E., Irani, M., Basri, Ronen: Actions as space-time shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12), 2247–2253 (2007)
14. Rahman, S.A., Song, I., Leung, M.K.H.: Negative space template: a novel feature to describe activities in video. In: *IEEE International Joint Conference on Neural Network (IJCNN)*, pp. 197–213 (2012)
15. Rakibe, R.S., Patil, B.D.: Human motion detection using background subtraction algorithm. *Int. J. Adv. Res. Comput. Sci. Soft. Eng.* **4**(2) (2014)
16. Sun, D., Roth, S., Black, M.J.: Secrets of optical flow estimation and their principles. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010)
17. Rahman, S.A., Cho, S.-Y., Leung, M.K.H.: Recognizing human actions by analyzing negative spaces. *IET Comput. Vis.* **6**, 197–213 (2012)
18. Chua, T.W., Leman, K., Pham, N.T.: Human action recognition via sum-rule fusion of fuzzy K-Nearest Neighbor classifiers. In: *International Conference on Fuzzy Systems (FUZZ)*, pp. 484–489 (2011)
19. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space-time shapes. In: *Tenth IEEE International Conference on Computer Vision (ICCV)* (2005)
20. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: *International Conference on Pattern Recognition*, pp. 32–36 (2004)
21. Fathi, A., Mori, G.: Action recognition by learning mid-level motion features. *Computer Vision and Pattern Recognition*. 1–8 (2008)
22. Bregonzio, M., Xiang, T., Gong, S.: Fusing appearance and distribution information of interest points for action recognition. *Pattern Recognition*. **45**, 1220–1234 (2012)
23. Liu, C., Yuen, P.C.: Human action recognition using boosted eigen actions. *Image and Vision Computing*. **28**, 825–835 (2010)

Protecting the Augmented Browser Extension from Mutation Cross-Site Scripting

S. Remya and K. Praveen

Abstract Browser extensions have a great role in bringing changes to the web browser behavior and improving browser performance. However, nowadays, many browser extensions fail to meet the security requirements. Due to this, they become a medium for various attackers to steal the credential information of the users. The proposal in this paper makes a pitch for the protection of an augmented browser extension against mutation-based cross-site scripting attack. A method is introduced for hardening the browser extension script along with dead code injection. A mutation-based cross-site scripting identifier to identify the attacks that affect the extensions is also discussed in this paper. These methods will protect browser extensions from various malicious script injection attacks.

Keywords Augmented browser · Mutation cross-site scripting · Dead code injection

1 Introduction

Web browsers connect users to the World Wide Web. A browser helps the user to visit various web sites and performs various activities such as sending mails, receiving mails, and also connects to various links. Most commonly used browsers are Mozilla Firefox, Google Chrome, and Internet Explorer. Using uniform resource locator (URL) one can browse through a particular link and reach the particular page they want to visit. Web browser extensions [1] are softwares which help to enhance a browser's functionality. Browser extensions are typically written in HTML, JavaScript, and CSS languages. Nowadays, new extensions are created

S. Remya (✉) · K. Praveen

TIFAC CORE in Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
e-mail: remyasomanadhan@gmail.com

K. Praveen

e-mail: praveen.cys@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_22

215

very frequently by developers. Some of them are useful and add value to the browser while others may harm the browser, thus compromising the security of the system. Augmented browser extensions are those extensions which come under augmented browsing. Examples for such browser extensions are Grease Monkey, Grease Kit, Tamper Monkey, etc. This type of extensions helps the end users to bring changes to their web pages and also allows personalization of their web pages. For example, in order to get rid of any advertisement displayed in one's Facebook page, one can download an add-blocker script from script market and install it into the Grease Monkey browser extension if the browser is Mozilla Firefox or install it into the Tamper Monkey in case of Google Chrome. Users can install any suitable scripts according to their purpose. But there is no guarantee that these scripts are free from vulnerabilities. An attacker can get into the user's system through some malicious JS scripts and steal the user's valuable information.

There are many kinds of cross-site scripting attacks that target browser extensions to steal session cookies, passwords, and various credential information. Cross-site scripting [2] are those types of attacks that inject malicious scripts into benign or some trusted websites. The attacker will send some malicious script to the targeted end user's system to exploit its security and functioning. There are different kinds of cross-site scripting attacks such as stored XSS attack, reflected XSS attack, and DOM-based XSS attack. Stored XSS attack is one which will store the malicious code inside the targeted server such as in some database, in command-line or in open forums. When the victim requests for some function, the malicious script will start injecting and will break the security. Reflected XSS is another type of XSS attack which mostly happens through web mails, on clicking a link in the email. Spam mails typically contain many kinds of malicious links. DOM-based XSS attacks affect the DOM properties such as `document.write` and `document.read`. The attacker brings changes to its entire behavior and can make the particular site or browser malicious. A mutation cross-site scripting attack affects the inner HTML of the JavaScript that is being installed into the browser extension. In such an attack, the attacker creates an HTML or an XML formatted string and injects the same into the browser extension. When the string is inserted into the browser's DOM with the help of inner HTML property, the string will start to mutate. In such cases, a server or a client-side filter will not be able to detect these attacks. Because of these kinds of XSS attacks, the augmented browser extensions are losing their security which in turn affects the browsers. Indeed, a more secure environment is necessary to work with these extensions. In this paper we propose a system which protects the augmented browser extensions from the mutation XSS attack (mXSS attack).

In Sect. 2, the previous works found in the literature, relevant to the current proposal are discussed. Section 3 describes the architecture of the proposed system. This is followed by results and discussions in Sect. 4. Section 5 concludes the paper.

2 Related Work

Heiderich et al. demonstrated in 2013 that mutation cross-site scripting mainly affects three browsers, Internet Explorer, Google Chrome, and Mozilla Firefox. The mXSS [3] mainly affects the inner HTML and its related properties. Heiderich and his coworkers pointed out that it is difficult to identify the mXSS by the normal XSS filters and HTML purifiers. Hence mXSS attacks are more vulnerable and may tamper the security of the system. By installing certain JS scripts in an augmented browser extension like Grease Monkey [4], one can bring changes to the appearance of the browser. These scripts can be obtained from script markets such as *script.org*, *Greasy Fork*, etc. But the scripts that are obtained from these script markets may not be free from vulnerabilities. This is because script markets may contain thousands of scripts which may not be written by experts. Because of this, security of both the browsers and the browser extensions could be compromised. Grease Monkey extension is more vulnerable to DOM-based XSS attacks enabling the attacker to steal credential information [5] of end user. XSS attacks [6] can happen due to malformed *href* tags, image tags, and also due to default source tag by leaving it empty [7]. This allows the attacker to easily inject their malicious code into the browser. In another study, Heidan et al. focused on JS information flow analysis [8] and tracking using an interpreter which is written in JavaScript. The study also provides information about the data flow in JavaScript along with third party script. Because of this, there is a possibility of knowing the changes that happened after the malicious code injection.

According to Athanasopoulos et al. the XSS attacks can be prevented by the isolation of client-side codes. Here they bring up a concept called XJS [9] which is performed based on instruction set randomization. This will protect the browser from malicious script injection attacks. Protecting JavaScript in the browser is the most important requirement for keeping the browser secure. As part of the present study, script analysis has been conducted on a few scripts which were collected from various script markets. This analysis has been performed using a framework called Vega [10] which is a vulnerability scanner and testing platform, designed using Java language. With the help of this tool one can find the script attacks easily. The results are given in Table 1. Also presented is a graphical analysis of the JS script based on the attack XSS and SQL injection as shown in Fig. 1. The performance is evaluated using Common Vulnerability Scoring System (CVSS). There are some score fixed by the CVSS. Based on that score, a JavaScript is considered as exploitable or not. The National Vulnerability Database [11] have fixed some CVSS score, based on that Fig. 1 is designed. Barua et al. in their study, focused on the protection of browser extension [12] from injection attacks such as XSS attacks, SQL attacks, etc. They put forward code randomization technique with static analysis. In their work, the code randomization is performed using encryption. In this paper, similar this concept is analyzed and a new pattern is arrived upon; wherein, the focus is on mXSS attack which was not addressed in previous works.

Table 1 JS script attacks

Scripts in script market	XSS attack	SQL attack
Redirect Userscripts.org to Userscripts-mirror.org	No	Yes
MoreManga	No	Yes
ZShowPass	No	Yes
WME Toolbox	No	Yes
uso—Count Issues	No	Yes
Old Twitter (black)	Yes	No
PublicHD—Remove News	No	Yes
uso—Count Issues	No	No
Larger Radio Buttons	Yes	No
YouTube—Better Watch Page	Yes	No
Clucker v5.61 SPOCKLET	No	Yes
Facebook Autosubscriber	Yes	No
IS Colorful Timetable	No	No

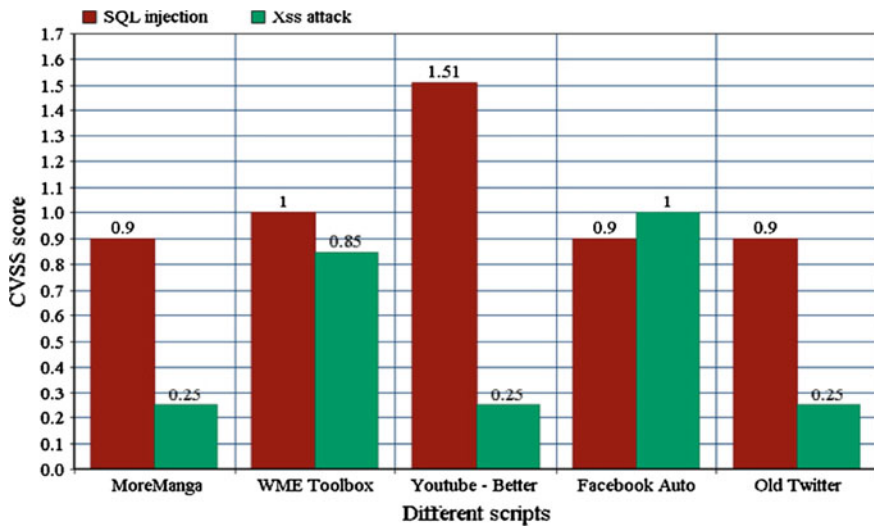


Fig. 1 Analysis of malicious JS file using Vega

In this paper, obfuscation of JS file in the extension is achieved using one-time pad approach along with dead code injection. The code in Fig. 2 gives a piece of information of the JS file from the augmented browser extension, Tamper Monkey [13]. Here the malicious code is injected into a benign JavaScript which allows an attacker to steal session cookies.

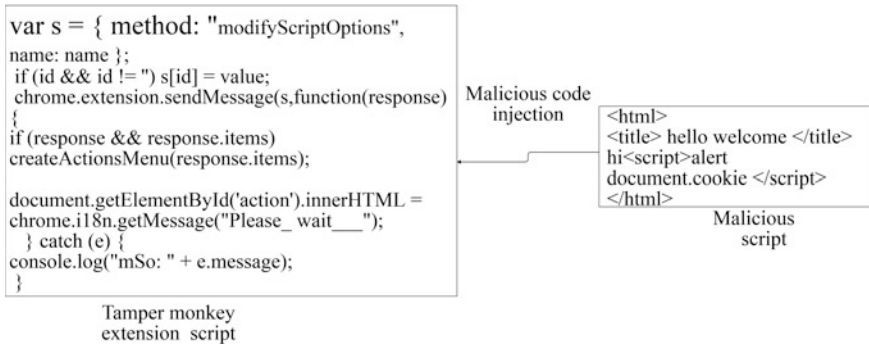


Fig. 2 Example of malicious script injection (Source Tamper monkey script)

3 Proposed System

3.1 Protection Mechanism from mXSS

The current proposed system is evaluated by performing experiments over multiple augmented browser extensions of Google Chrome. The details of the same are described in the following sections. In this paper the extension script is protected with Obfuscating module by using one-time pad, as described in Fig. 3. Using this one-time pad method the code will be hardened and it will be difficult for the attacker to inject the malicious code and bring changes to the browser extension. By this method of one-time padding the JS file is hardened with random private keys generated using pseudorandom number generator. This randomness of key provides more security, so that the attacker will have difficulty in breaking the code. A browser extension usually consists of several files and is usually available as a complete package in compressed format. As the target is to prevent malicious JavaScript injection in an extension, we focus on the file that contains executable JavaScript code. Executable JavaScript code can be found in stand-alone JavaScript code files (*.JS files). One can extract all the JS code files from the extension and feed these files to the code transformer for obfuscation. Deobfuscation is performed by the same keys which are generated at the time of one-time padding that will be given to the trusted user so that they can deobfuscate by XOR'ing with those keys.

If an attacker attempts to inject a regular, non-obfuscated attack script into an obfuscated extension script, attacker will have difficulty in understanding the hardened code to bring changes in the scripts. Thus, the information remains safe within the browser. An attempt to inject malicious code to an obfuscated extension scripts can be detected using mXSS identifier. A detailed explanation of the logic on which the mXSS identifier works is mentioned in the following section. After identification, the obfuscated extension scripts are converted into its normal form, assigned the appropriate security principal, compiled into the valid JavaScript byte

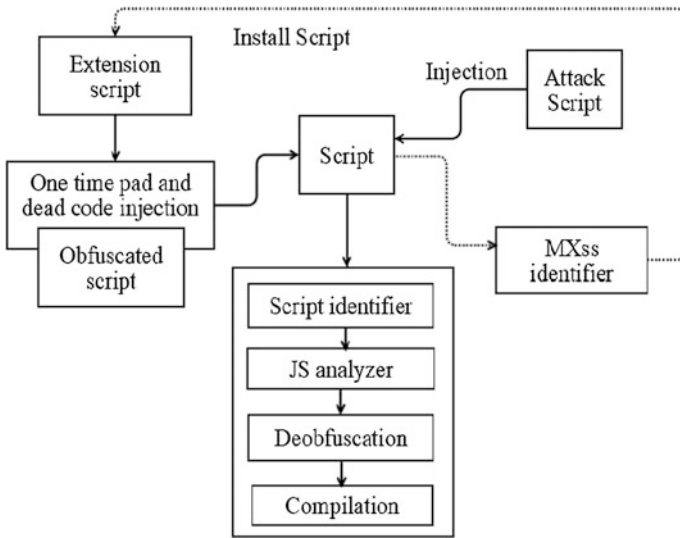


Fig. 3 Protection mechanism from mXss

code, and is executed properly. However, the regular JavaScript code of the attacked script is deobfuscated and a garbled output is obtained.

3.2 Obfuscation of Extension and Dead Code Injection

Obfuscation is a method of converting a program in such a way that it will be more difficult for an adversary to understand or bring any changes to the original code. Some private keys are used to protect the JS scripts. It is performed using one-time padding, so that the key will be generated randomly. After the obfuscation, it requires more money and time for reaching back to the original program. The one-time padding method generates the keys using a pseudorandom generator.

To protect the browser extensions from mXSS attack another method has been used in this work; dead code injection. Here the code will be designed in such a way that it will be in the format of inner HTML so that the adversary will think that it is the place where they can perform the attack. Dead code will not be executed while the source code is running, so this will have no impact on the security of the script files. Thus one can maintain security of the system. By such methods, security of the browser extension is ensured.

The code given here is a piece of example to show code hardening, this code is taken from Grease Monkey extension script [4]. The sample code shown below is not obfuscated. After obfuscation the code will not be in a normal readable format. Here the hardening of the objects and functions are performed using one-time pad.


```

init function () {
}
onLoad function () {
var cb = Registry.callbacks.pop();
}

```

Shown below is the JavaScript hardened using randomized private key to make the code more difficult for an attacker to understand. In the code, function () and pop () are hardened using one-time padding. By this, one can protect one's code from malicious injection attacks. Here obfuscation is performed using XOR operation and the key generated here will be in form of ASCII.

```

init XHce3tavi {
}
onLoad Svlu6rlo {
var cb = Registry.callbacks.uaow7jtxa;
}

```

The mXSS identifier used here will identify whether any mutation XSS attack has happened in the JavaScript files. This can be detected using the method called taint analysis. Taint analysis provides an attack alert before the particular JavaScript code is going to execute. If the attacker initiates any malicious code injection, by performing taint analysis one can understand whether any attacks have happened or not. Taint analysis is performed by focusing on malicious script injection attacks which affect the cookies, sessions, and passwords. These types of scripts are analyzed for attacks and a database of malicious script is generated. This database is typically known as trained data set. While performing a taint analysis, if any part of the script matches with the identified malicious script in the trained data set then that particular script is considered to be malicious. If a script is found malicious during the taint analysis, mXSS identifier will not allow the JavaScript to be installed. Otherwise, it will proceed with further installation.

4 Results and Discussions

This section explains about the experimental results. The experiments are performed on different JavaScript files that are going to be installed in an augmented browser extension. These scripts are collected from different script markets such as *script.org*, greasy fork, etc. They are installed to Tamper Monkey extension, which is an augmented browser extension from Google Chrome. Each script, before being installed, is checked using the mXSS identifier to see whether the script is trying to inject any attack into the particular browser extension. From the analysis it was observed that the code injection attacks mainly affect the areas such as: Onclick (), Onload(), URL, document.url, Inner.html and Eval () function. The malicious scripts are mainly added by the attacker through the location listed in

Table 2 Locations of malicious script attacks

Location	Malicious scripts
<code>OnClick()</code>	Auto like Facebook
<code>Onload()</code>	Facebook New look
<code>url</code>	Youtube video viewer
<code>document.urlunencoded</code>	Twitter version
<code>Inner.html</code>	WME Toolbox

Table 2, which was identified through the analysis performed. The framework that we built will help to protect the augmented browser extension from mutation cross-site scripting attacks. It acts as a barrier between the extensions and brings privacy to the browser extension and its performance, so that one can protect the end user system from malicious attacks. Each script that is collected from the script market was checked properly and then allowed to be installed. Table 2 gives a short description about a few selected scripts that are affected by improper data sanitization. Data sanitization is very much essential to avoid mXSS and also XSS attacks.

5 Conclusion

Mutation XSS is one of the main attacks in cross-site scripting. It has an impact on all three browsers such as Internet Explorer, Mozilla Firefox, Google Chrome. In this work, our main contribution is a methodology to protect the augmented browsing from mutation cross-site scripting attack that affects the extensions. Code hardening along with dead code injection is performed to protect the browser extension scripts, so that the vulnerable scripts cannot be installed into the extensions. It is a run time protection mechanism which enables the end user from malicious script attacks. This safeguards both the browser extensions and the browser from mXSS attack. This proposal will provide a new vision to the developers and security analysts.

References

1. Guarnieri, S., Pistoia, M., Tripp, O., Dolby, J., Teilhet, S., Berg, R.: Saving the world wide web from vulnerable javascript. In: Proceedings of the 2011 International Symposium on Software Testing and Analysis, pp. 177–187. ACM (2011)
2. Sundareswaran, S., Squicciarini, A.C.: Xss-dec: A hybrid solution to mitigate cross-site scripting attacks. In: Data and Applications Security and Privacy XXVI, pp. 223–238. Springer (2012)
3. Heiderich, M., Schwenk, J., Frosch, T., Magazinius, J., Yang, E.Z.: mxss attacks: Attacking well-secured web-applications by using innerhtml mutations. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, pp. 777–788. ACM (2013)

4. Van Acker, S., Nikiforakis, N., Desmet, L., Piessens, F., Joosen, W.: Monkey-in-the-browser: malware and vulnerabilities in augmented browsing script markets. In: Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security, pp. 525–530. ACM (2014)
5. Magazinius, J., Hedin, D., Sabelfeld, A.: Architectures for inlining security monitors in web applications. In: Engineering Secure Software and Systems, pp. 141–160. Springer (2014)
6. Manico, J.: Xss filter evasion cheat sheet. <https://www.owasp.org/index.php/XSSFilterEvasionCheatSheet>
7. Wang, Y.H., Mao, C.H., Lee, H.M.: Structural learning of attack vectors for generating mutated xss attacks (2010)
8. Hedin, D., Birgisson, A., Bello, L., Sabelfeld, A.: Jsflow: tracking information flow in javascript and its apis. In: Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 1663–1671. ACM (2014)
9. Athanasopoulos, E., Pappas, V., Krithinakis, A., Ligouras, S., Markatos, E.P., Karagiannis, T.: xjs: practical xss prevention for web application development. In: Proceedings of the 2010 USENIX Conference on Web Application Development, pp. 13–13. USENIX Association (2010)
10. Zalewski, M.: Vega vulnerability scanner. <https://subgraph.com/vega/documentation/index.en.html>
11. NIST: National vulnerability database. <https://nvd.nist.gov/CVSS-v2-Calculator?vector>
12. Barua, A., Zulkernine, M., Weldemariam, K.: Protecting web browser extensions from javascript injection attacks. In: 2013 18th International Conference on Engineering of Complex Computer Systems (ICECCS), pp. 188–197. IEEE (2013)
13. Schneider, J.: The extrasolar planets encyclopaedia. <http://exoplanet.eu/catalog.php>

Large-Scale Data Management System Using Data De-duplication System

S. Abirami, Rashmi Vikraman and S. Murugappan

Abstract Data de-duplication is the process of finding duplicates and eliminating it from the storage environment. There are various levels where the data de-duplication can be performed, such as file level, where the entire file as a whole is considered for the purpose of duplicate detection. Second is chunk level, where the file is split into small units called chunks and those chunks are used for the duplicate detection. Third is byte level, where the comparisons take byte-level comparison. The fingerprint of the chunks is the main parameter for the duplicate detection. These fingerprints are placed inside the chunk index. As the chunk index size increases, the chunk index needs to be placed in the disk. Searching for the fingerprint in the chunk index placed in the disk will consume a lot of time which will lead to a problem known as chunk lookup disk bottleneck problem. This paper eliminates that problem to some extent by placing a bloom filter in the cache as a probabilistic summary of all the fingerprints in the chunk index placed in the disk. This paper uses the backup data sets obtained from the university labs. The performance is measured with respect to the data de-duplication ratio.

Keywords Data de-duplication • Storage • Compression

S. Abirami (✉) · R. Vikraman

Department of Information Science and Technology, College of Engineering,
Anna University, Chennai, India
e-mail: abirami_mr@yahoo.com

R. Vikraman

e-mail: vikramanrashmi@gmail.com

S. Murugappan

School of Computer Science, Tamilnadu Open University, Chennai, India

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_23

1 Introduction

For any organization, its information is the most important component. Hence it is competitive to manage the data efficiently. As the amount of information in organization increases, managing the same becomes a very complicated task due to high amount of redundancies occurring in the backup. Data de-duplication [1, 2] comes as a rescue for such a problem by eliminating the redundancies occurring in the backup and keeping only one copy of the redundant data and all others as a pointer to this unique copy. The data de-duplication divides the stream of data into variable or fixed size chunks [3]. These chunks undergo a hashing algorithm like Rabin fingerprint or SHA-1 to produce a unique fingerprint which is placed in the chunk index [4]. Every incoming fingerprint is searched in the chunk index, if it exists, then the chunk is duplicate else it is unique and stored in the chunk index. A pointer to the unique data is stored for the duplicate chunk is placed instead of the actual data. All the information related to the file, chunks, etc., is placed inside the metadata of the file which is used for the reconstruction of the file.

The nature of the data in the backup workloads is mostly redundant and is often stored in similar order as in the week before. Hence a large amount of duplicates can be found when 2 consecutive backup workloads are given to the duplicate detection engine. An LRU cache memory is simulated to keep the bloom filter to avoid the chunk lookup disk bottleneck problem to some extent. This is helpful in eliminating few searches for the unique chunks in the chunk index placed in the disk.

2 Literature Survey

Various existing works based on different technologies of data de-duplication has been discussed below. Chang et al. discussed the Two Thresholds Two Divisors (TTTD) algorithm used for the CDC. To eliminate the problem of a chunk size being too small or big in variable length chunking, two thresholds were used for setting the minimum and maximum size of a chunk. Two divisors are also used, i.e., the main divisor used for finding a breakpoint and a second divisor to find a backup breakpoint if a breakpoint is not found by using the main divisor till the maximum chunk size is reached. The second divisor leads to breakpoints which are near to the maximum threshold thus leading to large-sized chunks. This is the limitation of TTTD algorithm.

He et al. [5] had given a brief study on the various data de-duplication techniques. Their work deals with the various data de-duplication strategies, processes, and implementation. The various detection strategies are alone at file level, byte level, and block level. Here, the algorithm used by them for fingerprint generation is MD-5 or SHA-1 which are less likely to produce collisions. Tarkoma et al. [6] discussed the bloom filter algorithm in detail. Burton H. Bloom [7] proposed the

concept of bloom filter at first. They conclude that the bloom filter supporting a set of membership queries. It can result in false positives in worst cases, but never lead to false negatives.

Zhu et al. presented a combination of three approaches, used in the EMC data domain file system (DDFS), to overcome the chunk lookup disk bottleneck. As a result in this paper, motivated by the various factors discussed above, we propose a data de-duplication system which can promote a successful de-duplication in this system.

3 Data De-duplication System

In this research, the overall architecture is shown in Fig. 1. The files like word, text docs, etc., in the backup stream enter into the de-duplication engine. Here the files are processed one at a time. Initially, the file enters into the file chunker where the file is segmented into chunks of fixed size or of variable size. The chunks enter into the fingerprint generator where Rabin fingerprint algorithm [8] is performed to obtain fingerprint of the chunks. The fingerprints play a major role in identifying the duplicates in the system. The fingerprints enter into the de-duplication logic where

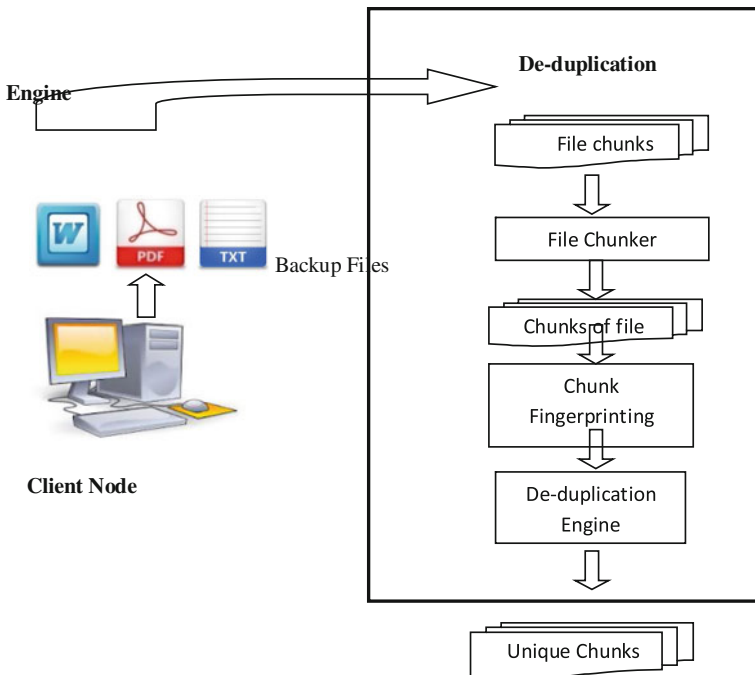


Fig. 1 Overall architecture

it is found whether the chunk is a duplicate or unique. This would identify the redundant chunks to remove it and place a pointer in the place of it which will be pointing to the previously stored unique chunk. The unique chunks coming out of this logic are stored inside the storage pool.

Module Description

As mentioned in above section, the detailed module description has been given below.

Files are segmented into chunks using fixed-sized chunking or variable-sized chunking. For variable-sized chunking Rabin fingerprinting algorithm can be used. The output of the file chunker is the chunks. In our approach we are making use of Two Thresholds Two Divisors (TTTD) variable-sized chunking algorithm [9] for chunking. The TTTD algorithm uses four parameters, the maximum threshold, the minimum threshold, the main divisor, and the second divisor.

The chunks from the file chunking are given to the fingerprint generator which will undergo Rabin fingerprint algorithm to produce a 4-byte fingerprint for the input chunk. The fingerprint is the main component for finding the redundancies. All the fingerprints obtained from here enter into the bloom filter.

3.1 Bloom Filter

The Bloom filter, a space-efficient data structure supporting a good amount of set membership queries can be used to check the presence of fingerprint in the chunk index [7]. This filter can lead to false positives in worst cases, but never end toward false negatives.

3.2 Duplicate Detection Using Chunk Index

This module involves finding the duplicates based on the fingerprints obtained from previous module. The presence of all fingerprints needs to be checked in the chunk index [10–13]. Because the limitation of the bloom filter is that it results in false positives. The fingerprint is searched in the main chunk index based on the result from the bloom filter. The duplicate fingerprint implies that the chunk is duplicate and the file recipe of the file is updated with the pointer to the unique chunk store. If the fingerprint is not a duplicate, then the chunk is stored and the file recipe of the file is updated with the chunk details.

3.3 *Chunk Mapping and Store*

The unique chunks are inserted into the chunk index stored inside the block. Chunk mapping also involves creation of a file recipe which consists of the file metadata information such as the filename, file type, backup time, and file size. Along with this, the details of the chunks within this file are also included. File recipe is helpful in the reconstruction of the file later. The block containing the unique chunks is compressed using Ziv Lempel algorithm and a compressed block is obtained at the end. The block recipe contains the details of the fingerprints in the blocks. The block number needs to be updated inside the file recipe of the file for the purpose of reconstruction of the file.

3.4 *Block Manager*

The Block Manager deals with keeping an index of all the blocks along with their address and the fingerprint details. This creates a block recipe which is essential to reconstruct a block using all its fingerprints. The block recipe consists of the block number along with the fingerprints in the block.

3.5 *Reconstruction of File*

When the user makes a request for a backup stream, then the file recipe of the particular file is loaded from the disk. The fingerprints obtained from the file recipe are searched in the block recipe to get the blocks. The compressed block is retrieved, uncompressed, and the chunks are obtained from the offset given in the file recipe. The file is reconstructed from all the chunks and it is given back to the user.

4 *Implementation*

We have evaluated the throughput and scalability of our de-duplication system using a simulated memory. It was carried out in a 4 GB RAM memory using the Microsoft visual studio 2010 as the debugging tool. For the variable-sized chunking, TTTD algorithm has been used which uses four parameters for an average chunk size of 1000 (bytes) and are initialized with the following fields (Table 1).

Table 1 Parameters used in the TTTD algorithm

Parameter	Purpose	Optimal value
Maximum threshold	To reduce very large chunks	2816 (bytes)
Minimum threshold	To reduce very small chunks	448 (bytes)
Main divisor	To determine breakpoint	540
Second divisor	To determine a backup breakpoint	260

Pseudo Code for TTTD Variable Length Chunking Algorithm

Overview: using variable size chunking algorithm, the incoming files is split into various chunks of varying sizes

Steps:

Initialize **current pointer** to zero

Initialize **last pointer** to zero

Initialize **Backup breakpoint** to zero

for not **EOF**

set **c** to the next byte of the input

compute the **hash key** of the input

assign **hash** to hash key obtained

if difference in the current pointer and last pointer is less than **min threshold**
then

continue

if hash mod second divisor is equal to **second divisor -1** then

set the **backup breakpoint** at the current pointer

if hash mod main divisor is equal to **main divisor -1** THEN

add **breakpoint** at the current pointer

set **backup breakpoint** to zero

set **last pointer** to current pointer

continue

if the difference in the current pointer and the last pointer is less than **max threshold** then

continue

if **backup breakpoint** is not equal to zero THEN

add **breakpoint** at the backup breakpoint

set the **last pointer** to the backup breakpoint

set **backup breakpoint** to zero

else

add **breakpoint** at the current pointer

set **last pointer** to the current pointer

set **backup breakpoint** to zero

increment the **current pointer**

The chunks thus obtained from the file chunker undergoes Rabin fingerprint algorithm to produce a fingerprint for every chunk with a length of 4 bytes each.

Variable-sized chunks obtained from chunking algorithm have been considered as input and the fingerprints of chunks have been obtained as output.

Input to the duplicate detection engine is the fingerprints obtained from the fingerprint generator. First the fingerprints enter into the Bloom filter to check whether the fingerprint is present, a member of the chunk index. If it is not, then a check in the chunk index is eliminated. Else a search has to be made in the chunk index to ensure the presence of the fingerprint as Bloom filter [7] can lead to false positives but never a false negative. The Eq. (4.1) shows how to find the de-duplication percentage.

$$\text{Data de-duplication ratio} = (\text{bytes in} - \text{bytes out}) / \text{bytes in} \quad (4.1)$$

4.1 Fixed-Sized Chunking Versus Variable-Sized Chunking

Table 2 shows the performance evaluation for fixed-sized chunking along with the corresponding de-duplication savings percentage. It shows the various test data sets on which the test is performed and their corresponding results based on it.

Table 3 shows the performance evaluation for fixed-sized chunking with respect to the de-duplication saving percentage. It shows the various test data sets on which the test is performed and their corresponding results generated based on it.

Table 2 Performance evaluation in fixed-size chunking

File type	Number of files	Average file size (KB)	Total chunks	Duplicate chunks	Total size (KB)	De-duped size (KB)	De-dupe saving (%)
*.txt	103	58	10,673	150	10,618	10,472	1.37
*.rar	5	450	2674	0	2672	2672	0
*.bin	50	256	12,887	0	12,845	12,845	0
All	158	100	30	0	27	27	0

Table 3 Performance evaluation in variable size chunking

Type	number of files	Average file size (KB)	Total chunk	Duplicate chunks	Total size (KB)	De-duped size (KB)	De-dupe saving (%)
*.txt	103	58	7043	5631	5900	836	86
*.rar	5	450	1947	158	2200	1990	8.4
*.bin	50	256	8327	5855	7500	3050	60
All	158	100	17,317	11,645	15,800	5950	62

Table 4 Accuracy estimation of different data sets

Data sets	Total chunk	True positive	True negative	False positive	False negative	Precision (%)	Recall (%)	Accuracy (%)
Set 1	12	6	2	1	4	85.7	60	61.5
Set 2	15	7	3	1	5	87.5	58.33	62.5
Set 3	35	16	6	2	3	84.21	88.8	81.4

4.2 Accuracy Estimation

Table 4 represents the data sets with the corresponding precision and recall values to find the accuracy of the system. Set 1, 2, and 3 are the set containing txt files, binary files, and cpp files respectively. It shows the total number of chunks present along with the TP, FP, TN, FN values calculated for each data set to get the precision and recall values. Precision can be found using the Eq. (4.2), recall is found using the Eq. (4.3), and accuracy is found using the Eq. (4.4).

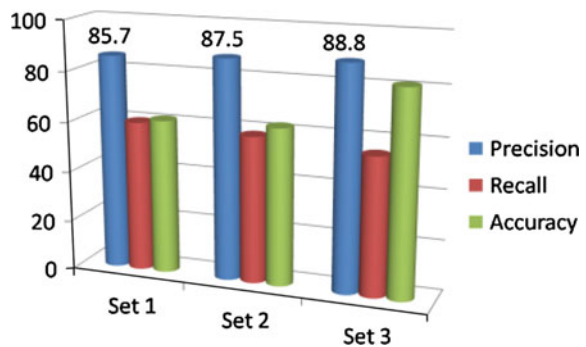
$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \tag{4.2}$$

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \tag{4.3}$$

$$\text{Accuracy} = \text{TP} + \text{TN}/(\text{TP} + \text{TN} + \text{FP} + \text{FN}) \tag{4.4}$$

Figure 2 shown below depicts the graphical representation of precision and recall and accuracy. The horizontal axis represents the recall values and the vertical axis represents the precision values. An average of 68.46 % accuracy has been produced using this data de-duplication system. This may be due to the fingerprint algorithm used, hence more advanced hashing algorithm can be used for this purpose but it may consume a lot of time. Figure 2 shows the relationship between precision, recall, and accuracy.

Fig. 2 Performance evaluation across different datasets



5 Conclusion and Future Work

This paper presents the evaluation of a de-duplication system using variable-sized chunking algorithm. The chunks undergo fingerprint algorithm to generate a hash key for making duplicate detection. The duplicate detection also uses a Bloom filter for set membership query. By keeping the Bloom filter in the cache, a few unnecessary searches in the chunk index placed in the disk can be eliminated. The chunking algorithm can be improved in order to produce better chunks for making better chunk duplicate detection. This can be done by managing the chunk size distribution where maximum chunks can be found in the average size instead of having chunks which are too big in size or too small in size. Further, the similarity or locality of the backup data set can be considered to reduce the chunk lookup disk bottleneck problem to some further extent.

References

1. Mark, R.C., Whitner, S.: *Data De-duplication for Dummies*. Wiley Publishing, Inc (2008)
2. Vikraman, Rashmi, Abirami, S.: A study on various data de-duplication systems. *Int. J. Comput. Appl.* **94**(4), 35–40 (2014)
3. Bhagwat, D., Eshghi, K., Lillibridge, M., Long, D.D.E.: Extreme binning: scalable, parallel de-duplication for chunk-based file backup. In: *Proceedings of the IEEE International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems*, pp. 1–9 (2009)
4. Thein, N.L., Thwel T.T.: An efficient indexing mechanism for data de-duplication. In: *Proceedings of the International Conference on the Current Trends in Information Technology (CTIT)*, pp. 1–5 (2012)
5. He, Q., Zhang, X., Li, Z.: Data de-duplication techniques. In: *Proceedings of the International Conference on Future Information Technology and Management Engineering*, pp. 430–433 (2010)
6. Rothenberg, C.E., Lagerspetz, E., Tarkoma, S.: Theory and practice of bloom filters for distributed systems. Published in *IEEE Communications Surveys and Tutorials*, pp. 131–155 (2012)
7. Zhu, B., Patterson, H., Li, K.: Avoiding the disk bottleneck in the data domain de-duplication file system. In: *Proceedings of the Sixth USENIX Conference on File and Storage Technologies*, pp. 269–282 (2008)
8. Rabin, M.O.: Fingerprinting by random polynomials. Technical Report TR-15-81, Center for Research in Computing Technology, Harvard University (1981)
9. Chang, B., Moh, T.: A running time improvement for two thresholds two divisors algorithm. In: *Proceedings of the ACM Southeast Regional Conference*, pp. 69–107 (2010)
10. Mishra, M., Sengar, S.S.: E-DAID: an efficient distributed architecture for in-line data de-duplication. In: *Proceedings of the International Conference on Communication Systems and Network Technologies*, pp. 438–442 (2012)
11. Wang, C., Wan, J., Yang, L., Qin, Z.-G.: A fast duplicate chunk identifying method based on hierarchical indexing structure. In: *Proceedings of the International Conference on Industrial Control and Electronics Engineering (ICICEE)*, pp. 624–627 (2012)

12. Gadan, A., Miller, E., Rodeh, O.: HANDS: a heuristically arranged non-backup in-line de-duplication system. In: Proceedings of the 29th IEEE International Conference on Data Engineering (ICDE), pp. 446–457 (2013)
13. Feng, D., Sha, E.H., Ge, X., Tan, Y., Yan, Z.: Reducing the de-linearization of data placement to improve de-duplication performance. In: Proceedings of the 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, pp. 796–800 (2012)

Movement Detection Using LabVIEW by Analysis of Real-Time Video

D.V. Sumalatha and A.V. Ravi Kumar

Abstract Movement detection is the key to solving many simple and complex problems in the real world. In its simplistic form, movement detection involves capturing the subject in question and comparing it with previous knowledge of shape, size, and displacement of subject from the captured snapshot/reference. This paper explores different noise filtering techniques *Averaging/Mean filter*, *Median filter*, *Gaussian smoothing* and object detection methods such as *Background subtraction*, *Optical flow method*, *Temporal differencing*, *Sum of Absolute difference* along with its advantages and disadvantages. The paper describes the *Gaussian smoothing* and *absolute difference* method that was used to detect movement in real-time video using LabVIEW. The approach involved processing set of consecutive video frames, extracting absolute difference of each other to detect foreground and background objects and its relative displacement from previous position. Subsequent to movement detection, the method also aims to highlight the region of object movement along with a Boolean indicator to visually inform the end user about movement detection.

Keywords LabVIEW · Absolute difference · Object movement detect

D.V. Sumalatha (✉) · A.V.R. Kumar
Electronics Communication Engineering Department, V.T.U, Sri Jagadguru
Balagangadharanatha Swamiji Institute of Technology, Bangalore, Karnataka, India
e-mail: suma.dv18@gmail.com

A.V.R. Kumar
e-mail: ravi_sjb@rediffmail.com

1 Introduction

1.1 *Laboratory Virtual Instrument Engineering Workbench (LabVIEW)*

Laboratory Virtual Instrument Engineering Workbench (LabVIEW) is a visual programming language for system-design and platform development offered by National Instruments Corporation. LabVIEW program language is known as “G”. It is graphical in nature and supports dataflow. In LabVIEW, graphical block diagram are referred as source code, user interfaces are called as front panels, while subroutines/programs are called as Virtual Instruments (VIs). Block diagram, front panel, and connector pane are the three important components of VI. User interfaces or front panels are developed from Controls and Indicators. Any information that user passes as input are developed using controls. Indicators act as output which shows the result depending on the evaluation of input passed to the VI through controls. Back panel of the VI encapsulates itself the source code. Terminals in back panel depict all controls and indicators of front panel. The graphical block diagram decides the execution pattern or path of the executable VI. Wires are used as bridge connectors between various functions, i.e., output of one node is connected as input of other node or vice versa. Even variables in a LabVIEW program are moved using wires. Unless all input data is available, a node cannot be executed. The term Virtual Instruments is used for LabVIEW programs as its mimics physical devices. LabVIEW is supported on various operating system flavors of Microsoft Windows, Mac OS, Linux, and UNIX. LabVIEW finds its application in the areas of automation for industries, acquisition of data from I/O devices, instrument control, etc. [1].

1.2 *Noise Filtering*

Noise filtering is the technique of eliminating undesired noise from the image by applying image processing filters. In image processing, smoothing refers to the process of eliminating pixels which change swiftly over a given area. Edge detection is the process of improving pixels which change slowly over a given area [2]. Since object detection using real-time video involves pixels that change rapidly, image smoothing techniques are examined further.

Averaging/Mean filter. The process of decreasing the strength of adjacent pixels is known as averaging or mean filter. In this procedure, target pixel is replaced with calculated mean value of the pixels adjoining the target. This will remove pixels that have no similarity between its neighbors. Disadvantages of mean filter are that (A) if a pixel has significant deviation from its neighbors, resultant mean value would affect the adjacent pixels to a large extent. (B) When the averaging method is applied on pixels representing edges, this may obscure the edges of an image. This affects applications that rely heavily on sharp edges [2].

Median filter. This filter differs from mean filter in the sense that instead of mean, target pixel is replaced with calculated median value of the pixels adjoining the target. First, the chosen pixel and its neighbors are sorted in ascending order. The value of middle pixel in sorted list is chosen and target pixel is reinstated with this value. If the chosen image has large number of pixels, it is time-consuming to sort the pixel value. This is one of the disadvantages of median filter method [2].

Gaussian smoothing. Gaussian smoothing is a mathematical operation on two-dimension input whose output is overlap of inputs. Gaussian smoothing uses a kernel whose graph when plotted resembles a symmetric bell curve. The Gaussian distribution in 2D has the form as below [2]:

$$h(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{\{(x - x_0)^2 + (y - y_0)^2\}}{2\sigma^2}} \quad (1)$$

x and y represent horizontal and vertical axes of graph respectively, while σ denotes standard deviation. The distribution when applied results in a convoluted matrix. The convoluted matrix is applied to the image where target pixel is replaced with the highest Gaussian value while adjacent pixels are given lower value based on the distance from target pixel. Gaussian distribution in theory will result in nonzero values and would need very large matrix to apply transformations. In practice however, if transformations are applied more than thrice, the resultant value is small enough to be replaced with zero [2].

1.3 Object Movement Detection

It is a fundamental requirement of many visual-based applications such as collision detection of vehicles, trajectory tracking of a moving particle, intruder detection systems, head counting systems in places such as banks, airports, etc. The process of ascertaining an object that is of concern, its shape, features in an image or video sequence is known as object detection [3]. Once the objects of interest are determined, the next step is to identify if the objects have changed its locality with respect to its surroundings. There are many methods to detect objects and its movement which are explored in the below sections.

Background subtraction. This is one of the well-recognized means of detecting objects that are under movement. The idea behind this approach is to estimate an appropriate representation of the background such that object(s) of interest in the current frame can be identified. This is achieved by subtracting the chosen frame with the background image(s) [4]. In other words, classify those regions of the image that are moving relative to the background [5]. If the pixels have not changed between the images, its value is near to zero and is considered as background. On the other hand, the pixels with bigger value are treated as foreground [6]. Though background subtraction method is flexible and reliable, the disadvantage is that

(A) it is time-consuming to determine the background region [4] and (B) this method is not ideal when the image source such as camera is in movement and background changes continuously [7].

Optical flow method. This is another approach to identify movement detection. This method calculates the velocity of a pixel in image frames by measuring the number of vectors starting or ending at a pixel [4]. The calculated velocity of pixel specifies how much the target pixel has deviated between adjacent pixels. If the velocity value of a pixel between two images is higher, then it is considered as movement [7]. An advantage of this system is that it can be used even when background and camera are not stationary. Disadvantages are (A) the method is complex (B) changes in background should be relatively less and camera capturing the images must not move too fast [7] and (C) suffers from the illumination problem.

Temporal differencing. It is based on frame difference and often uses more than two consecutive frames to determine objects and movement. Temporal differencing tries to find the disparity between successive frames to detect areas that are moving. The advantage of this method is its flexibility in rapidly changing environments, while the disadvantage is that it does not derive the entire shape of the moving object [7].

Sum of Absolute difference (SAD). SAD is the popular method to detect movement changes amid blocks of images. This method calculates absolute difference among every pixel in the selected image block and the interrelated pixel in the block similarity. In this technique, first color transformation is applied on the frames of video as part of the preprocessing step. This deducts foreground object from background which is followed by background subtraction. The frames where intensity changed suddenly are instantly identified by the actions mentioned above. While the above process is in progress, the boundary of the object where intensity changed is determined by applying edge detection mechanism [8]. Advantages of the SAD method are (A) due to non-complex nature of the method, it is quite fast. (B) Since every pixel can be processed independently, the function can be run concurrently [9]. SAD is represented by the below formula [10]

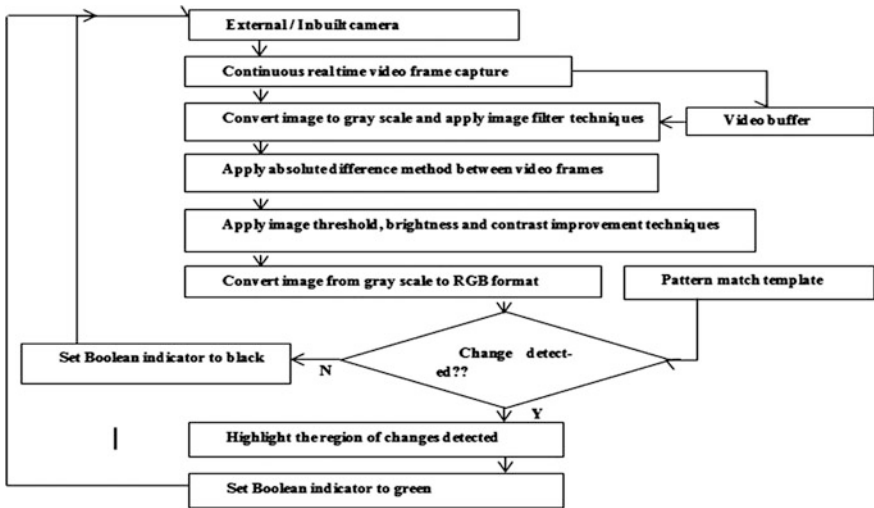
$$SAD = \sum_i \sum_j |I_k(i,j) - I_{k-1}(i,j)| \quad (2)$$

where: $I_k(i, j)$ = Current K th Frame and $I_{k-1}(i, j)$ = Previous $(K - 1)$ th Frame

Image Thresholding. The purpose of thresholding operation is to convert an image into binary format. The method sorts each pixel in the image as either dark (black) or bright (white). The simplest form of thresholding analyzes every pixel in the image or frame to a chosen definite value (threshold). If the pixel is greater than the threshold value, it is marked as foreground (bright), else the pixels are marked as background (dark) [11]. If only one value is used for the whole of image/frame, it is known as global thresholding. If multiple values are used for several pixels in an image, it is called local thresholding.

2 Movement Detection Work Flow

The first step is the selection of video source. Video source can be in-built or an external camera mounted on the computer system. The second step is to continuously capture real-time video from camera chosen in step 1. A buffer is set aside to select multiple frames for later use. The next step converts frames into grayscale since it is easier to process single color dimension than red blue green alpha (RGBA). RGBA requires processing in four color dimensions and then combining them for further processing. Unwanted noise from resultant frames are eliminated using filtering techniques (1.2) [12]. The frames K and $K - 1$ are then subjected to absolute difference method (2) to eliminate background and focus on objects that have moved. The resultant frame quality which is made visible to end user is further improved using image improvement methods. Global thresholding technique is used in the current paper whose value is a user configurable parameter. The frames are then converted back to RGB format to compare with pattern. The pattern is either white or gray color. White indicates significant change between the compared images while gray represents marginal changes in the intensity of pixels. The LED is set to Green if there is a pattern match, i.e., movement detection else LED is reset to black. The region of movement detection is also highlighted to indicate to end user the areas where changes were detected. The whole process runs in continuous loop until end user terminates the program.



3 Critical Implementation Methods

A. Image conversion and filter

Images/frames captured from camera in LabVIEW are in binary format and requires conversion to grayscale. External noise should also be removed for further processing as noted in Sect. 2. Figure 1 represents the module which converts binary image into grayscale format (Color plane extraction). “HSL Luminance plane” is the color plane extraction method used in this paper. Upon format conversion, image filtering techniques must be applied. “Smoothing Gaussian” is the image filtering method used in this paper (1). Kernel is a 3X3 convolution matrix that is subjected on the selected frame/image.

B. Absolute difference

Figure 2 shows the important method in the program design where the VI subtracts one image from another or a constant from an image and returns the absolute value of the difference (2). At times, a constant factor is also provided as input based on type of image processing method or requirement. In this paper, no constants were used. Successive buffered images were classified as even and odd numbered images. Even numbered images were used as Image Src A (Fig. 2) and odd numbered images were used as Image Src B (Fig. 2). If the images are similar, i.e., if there is no difference between Image Src A and Image Src B, the resultant computed image Image Dst out (Fig. 2) is empty or blank. If Image Src A and Image Src B are not the same, the difference is seen in the output, i.e., Image Dst Out. This difference in video frames indicates object movement changes.

C. Post-processing

Post-processing of images/frames is crucial as the resultant image representing the difference may not be distinguishable or easily noticeable to end user. This is

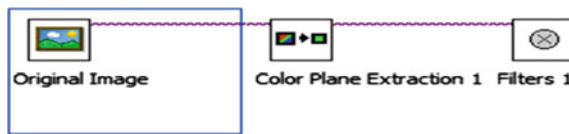


Fig. 1 Image conversion and filter method [1]

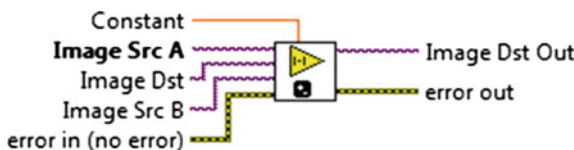
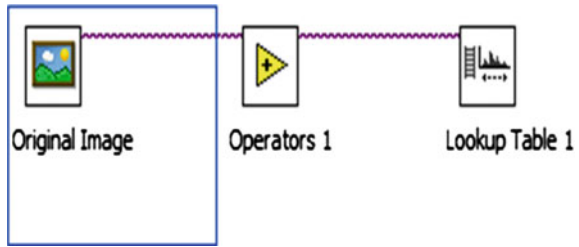


Fig. 2 Absolute difference method [1]

Fig. 3 Post-processing method [1]



more so in situations where the changes are very small or quick to notice by the naked eye. To enhance user visual experience, the images/frames are subjected to image enhancement methods. Figure 3 represents the post-processing module. The output image derived out of absolute difference method is enhanced by applying the “Multiply” by a factor of two and sent as input to lookup table module. The resultant image is further enhanced with “Power 1/x” by a factor of 1.5. This results in processed image quality enhancement by which end user has better clarity visually on even the minute regions that have changed.

4 Experiments and Results

The implemented project output has two screens, the first (Screen 1) shows the real-time video being captured from in-built camera, while the second screen (Screen 2) shows object movement detection. There is an LED indicator as well to inform end user whenever an object is in movement or displaces from its current position.

4.1 Experiment 1—Human Object Movement Detection

In first experiment, human hand is used as subject and detects its movement.

Figure 4 output shows no displacement of hand in Screen 1, i.e., static position; hence the LED and Screen 2 are black indicating no movement detection.

Fig. 4 Human hand without movement

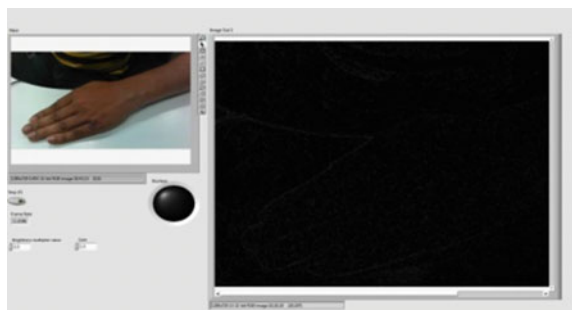
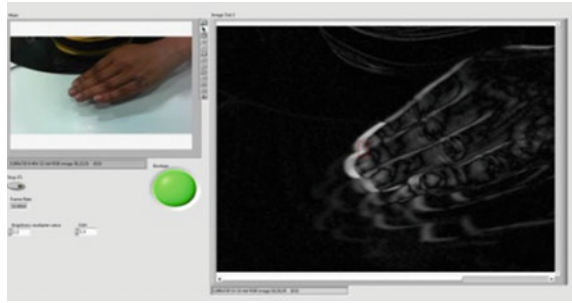


Fig. 5 Human hand with movement



The captured snapshot in Fig. 5 shows movement of hand and the change detection is indicated by LED turning green. The red square in Fig. 5 highlights the region of movement on Screen 2.

4.2 Experiment 2—Multiple Object Movement Detection

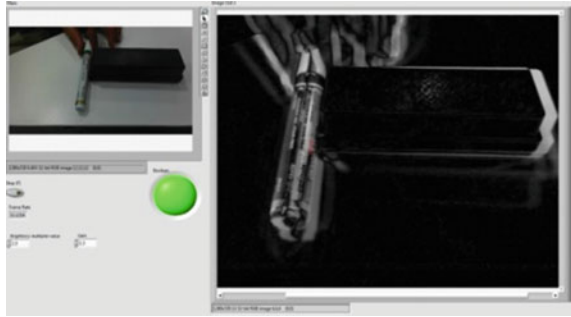
In the second experiment, two objects, marker and duster are used as test subjects.

With no movement of the white board marker and duster as observed in Fig. 6, LED and Screen 2 are black indicating that no objects in Screen 1 are displaced from its current position. The screenshot in Fig. 7 shows both marker and duster in movement simultaneously. Screen 2 white and gray areas indicate the displaced regions while red square highlights the region where the change has been detected. The LED turning green is again an indication to end user of the objects movement. The object that moved first is highlighted by red square and later the second object.

Fig. 6 Marker and duster without movement



Fig. 7 Marker and duster with movement



5 Conclusion

LabVIEW method is cost-effective as it does not involve pricey movement detection sensors/devices. It is also low on maintenance since LabVIEW program is easy to save and port to different computers/places compared to physical devices. LabVIEW along with its library modules provides simple and easy to use interface for image and video processing. Despite detecting multiple objects movement, the region of change is highlighted one after another and not simultaneously. Although Gaussian filter and absolute difference method has been used for the experiments, it may not be ideal for all situations. The end result may vary depending on the light intensity, quality of the camera/video, type of noise in the frames, etc. It is hence suggested to try other forms of object detection and noise reduction methods which fit best depending on the application that is being developed. Future enhancement could be adding multiple pattern recognition, counting systems, path tracking, and speed calculation. The program can be interfaced with different mechanical or electrical equipment to be used in real applications such as monitoring baby during its sleep, animal sighting and capturing its frame in forest, setting off alarm in intruder detection system, automatic door lock/unlock, water dispenser turn on/off system, etc.

References

1. National Instruments Corporation.: IMAQ vision for LabVIEW user manual. Part Number 322917B-01, June 2003
2. Qidwai, U., Chen, C.J.: Digital Image Processing: An Algorithmic Approach with MATLAB®, pp. 104–109. Chapman and Hall/CRC, Boca Raton (2009)
3. Wikipedia® Outline of object recognition. [Online] (2015)
4. Liua, C., Yuena, P.C., Qiu, G.: Object Motion Detection Using Information Theoretic Spatio-Temporal Saliency. Elsevier Ltd., New York (2009)
5. Swati, C., Vishal, J.: Human activity tracking and analysis for surveillance. *Int. J. Comput. Sci. Inf. Technol* **5**(5) 6707–6711 (2014)

6. Moore, D.: A Real-World System for Human Motion Detection and Tracking. California Institute of Technology, California (2003)
7. Lu, N., Wang, J., Wu, Q.H., Yang, L.: An improved motion detection method for real-time surveillance. *IAENG Int. J. Comput. Sci.* **35**, 119 (2008)
8. Sreedevi, M., Yaswanth Kumar, A., Anjan Babu, G., Sendhil Kumar, R.: Real time movement detection for human recognition. In: *Proceedings of the World Congress on Engineering and Computer Science vol. I* (2012)
9. Wikipedia® Sum of absolute differences. [Online] (2015)
10. Bhat, M., Kapoor, P., Raina, B.L.: Application of SAD algorithm in image processing for motion detection and Simulink block sets for object tracking. *Int. J. Eng. Sci. Adv. Technol.* **2**(3), 731–736 (2012)
11. Bradley, D., Roth, G.: Adaptive thresholding using integral image. *J. Graph. Tools* **12**(2), 13–21 (2007)
12. Tripty, S., Sanju, S., Bichu, V.: A new algorithm designing for detection of moving objects in video. *Int. J. Comput. Appl.* **96**(2), 4–11 (2014)

Connectivity Model for Molecular Communication-Based Nanomachines Network in Normal and Sub-diffusive Regimes

Prachi Raut and Nisha Sarwade

Abstract Nanomachines network is an interconnection of nanomachines (NMs) capable of communicating with each other. NMs networks are expected to provide an intelligent alternative to contemporary wireless sensor networks due to their biocompatibility, pervasiveness, and energy efficiency. However, connectivity issues of NMs networks are yet to be explored fully. This paper presents a probabilistic connectivity model for molecular communication-based NMs network which involves transmission of a message via diffusion of messenger molecules. This model has been developed through signal to interference and noise ratio (SINR) analysis considering effects of co-channel interference (CCI) and inter-symbol interference (ISI). It is found that ISI is the dominating factor in degrading the network connectivity than CCI. Also, results have shown that selection of symbol time is crucial and should depend on internode distance (or transmission range), for higher network connectivity. Physical obstructions in transmission media lead to anomalous diffusive behavior. This paper has investigated effects of sub-diffusion on the connectivity of NMs network to reveal that presence of physical obstructions can be a favorable condition in MC-based NMs networks if symbol time is adjusted accordingly.

Keywords Nanomachines network · Molecular communication · Network connectivity · Anomalous diffusion

P. Raut (✉) · N. Sarwade
Veermata Jijabai Technological Institute, Mumbai 19, India
e-mail: rautps@gmail.com

N. Sarwade
e-mail: nishasarwade@vjti.org.in

1 Introduction

Nanomachines (NMs) network is a futuristic networking paradigm consisting of interlinked nanoscale devices capable of sensing, computing, and communicating [1]. NMs may communicate with each other using molecular communication (MC) which takes place in fluidic media such as air, water, blood, etc. Transport in MC is classified as passive or active. In passive transport, messenger molecules diffuse through the medium to reach the receiver while in active transport molecules are carried from one point to another by molecular motors. Small size of NMs provides biocompatibility and pervasiveness. NMs are expected to harvest energy from their surroundings [2]. Energy efficiency and the ability to work in aqueous mediums are the unique features of NM networks. Although such NMs and NM networks have not been developed yet, breakthroughs like liposomes embedded with Gemini peptide lipids (GPL) capable of transmitting/receiving molecules [3], text message transfer by molecular communication [4], and wet lab experimentation on bacterial communication in multi-hop nanonetworks [5] suggest successful realization of nanoscale networks in near future. With prospective applications in biomedical, environmental, industrial, military, IoT, etc. [1, 6], NMs networks are set to become an intelligent alternative for electromagnetic energy-based WSN.

Connectivity of a network governs accessibility of its nodes by every other node, either directly or via relay nodes. A connected network guarantees reliable data transmission. Connectivity of conventional wireless sensor networks has been studied extensively. However, the problem of network connectivity appears to be more demanding in NM networks due to stochastic nature and large delay involved in molecular diffusion process. Also MC gets affected by physical obstructions present in the channel. Hence it is vital to analyze network connectivity of MC-based NMs networks.

In this paper, a probabilistic connectivity model for NMs network is developed. This network uses MC by diffusion of messenger molecules. The model evolves through the notion of node connectivity which is defined as the probability with which two arbitrary nodes are connected to each other. Node connectivity has been analytically derived using signal to interference and noise ratio (SINR). SINR is evaluated considering the Gaussian noise affecting the reception process and the interference. Co-channel interference (CCI) and intersymbol interference (ISI) are the two components of the interference considered here. In the context of MC by diffusion, ISI means number of molecules from previous transmission present in the receiver space and CCI stands for number of molecules received from another transmission occurring concurrently. If the SINR of the link between two nodes exceeds a threshold value, the two nodes are assumed to be connected to each other. Further, network connectivity is estimated from node connectivity between transmitter and neighboring nodes. Using MATLAB simulations, effect of internode distance (d) and symbol time (T_s) on network connectivity is observed. It is found that ISI is the dominating factor in degrading the network connectivity than the CCI. Also, results have shown that selection of T_s is crucial and should depend on

d , for higher network connectivity. Also, effect of sub-diffusive transport on the node connectivity which stems from presence of physical obstructions is investigated. The sub-diffusion is modeled as per master equation approach presented by Pagnini et al. [7]. Network connectivity for various internode distances is studied under sub-diffusive regime. It is revealed that presence of physical obstructions can be a favorable condition in MC-based NM networks if symbol time is adjusted accordingly.

Remaining part of the paper is structured as follows : Sect.. 2 presents a review of related previous work whereas Sect. 3 is a network overview. Development of connectivity model is described in Sect. 4. Anomalous diffusion is briefly introduced in Sect.. 5. Section 6 is dedicated for numerical results and discussions followed by conclusion in Sect.. 7.

2 Previous Work

Network connectivity is a relatively unexplored topic in nanoscale networks research scenario. However, a small body of the literature is dedicated to the study of reception probability in MC and the variation it undergoes due to change in transmission range and symbol timings. Kuran et al. [8] have studied effect of CCI on the probability of the transmitted molecules being received by correct receiver (P_{Rhit}) and the probability of the transmitted molecules being received by incorrect receiver (P_{Whit}). The authors have shown, through simulation, that as the separation between adjacent transmitters increases, P_{Whit} reduces and eventually attains zero value whereas P_{Rhit} increases moderately. Similarly, increasing transmission range has a negative effect on P_{Rhit} while P_{Whit} remains unaffected. If the transmission range is increased further P_{Whit} starts increasing as the messenger molecules move freely in the space increasing chances of being absorbed by wrong receiver.

In another publication, Kuran et al. [9] have analyzed probability of reception of molecules under different transmission ranges and symbol timings and also investigated effect of transmission range on data rates for various energy budgets. Due to slowness of diffusion process, molecular wave front reaches slowly to the receivers situated at longer distances. Hence the authors have proposed that communicating pair should negotiate the symbol time value according to separation between them prior to actual transfer of message. Simulation results in this work indicate that P_{hit} decreases with increasing distance between transmitter and receiver. Also as symbol time increases, P_{hit} increases. For the higher transmission ranges, increase in time symbol leads to smaller improvements in P_{hit} .

Walsh and Balasubramaniam [10] have analyzed reliability and delay of virus-based nanonetworks for single hop and multi-hop topologies. Findings of this paper show that in a virus-based nanonetworks particle dilution and diffusion speed can have a major effect on transmission delay and reliability. Single path with acknowledgement and random multipath topologies have large retransmission

delay which introduce more uncertainty for source to destination communication time compared to defined multipath topologies.

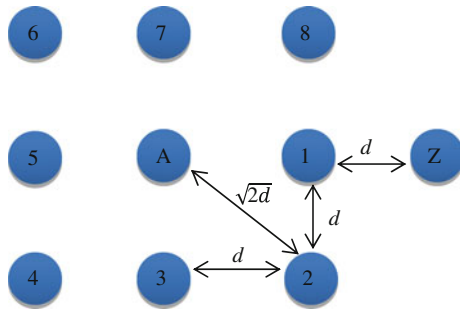
At the outset, [8, 9] have worked on single link connectivity while [10] considers a multi-node environment with various topologies. However, effects of noise and interference on network reliability are not considered in [10]. This paper considers node connectivity as well as network connectivity in presence of noise and interference. Moreover, environmental conditions like physical obstructions are taken into consideration and the effect they have on network connectivity are studied.

3 Network Overview

NM network considered in this work is deployed in an unbounded two-dimensional environment. The nodes of the network are NMs organized in a uniform square grid topology. The size of a square in the grid is d cm \times d cm as shown in Fig. 1. Although it is possible to have NMs which are mobile, for this analysis they are considered to be fixed at the given position. For computation of node connectivity, node A is considered to be the transmitter and nodes 1–8 are the receivers. Node Z is transmitting simultaneously causing CCI to the links between A and nodes 1 through 8.

The nodes transmit/receive messages in the form of concentration of messenger molecules. It is assumed that transmitter can generate required number of messenger molecules for transmission. A simple binary modulation scheme is considered in which transmitter emits Q number of molecules in a single puff to indicate symbol ‘1’ while no molecules are transmitted for symbol ‘0’. Transmitted molecules diffuse through the medium until some of them reach in the vicinity of a receiver. It is assumed that receiver is able to sense the presence of messenger molecules and counts them at the end of every symbol time T_S without changing them chemically. Mean value of molecular concentration present at the distance ‘ x ’ and at the time ‘ t ’ is given as shown in (1).

Fig. 1 Network topology



$$C(x, t) = \frac{Q}{4\pi Dt} \exp\left(\frac{-x^2}{4Dt}\right). \quad (1)$$

where D is the diffusion coefficient.

Since the receiver makes a decision at the end of the symbol time T_s , the molecular concentration available at the end of symbol time at receiver location is

$$C_{T_s}(x, T_s) = \frac{Q}{4\pi DT_s} \exp\left(\frac{-x^2}{4DT_s}\right). \quad (2)$$

$S(x, T_s)$ denotes the molecules present in the reception area A_r around the receiver surface as indicated in (3). It can also be viewed as strength of the received signal.

$$S(x, T_s) = C_{T_s} \cdot A_r. \quad (3)$$

Along with the actual signal strength, receiver is likely to pick up noise and interference signals as well. Due to the slow diffusion process MC suffers from ISI. For simplicity, ISI caused by only the previous symbol transmitted by the same transmitter is considered. If the previous symbol was '0', no intersymbol interference will be present in current symbol duration since no molecules have been emitted. Considering the probability of previously transmitted symbol '1' to be p , the amount of intersymbol interference present in symbol duration is quantified as:

$$ISI(x, 2T_s) = \frac{pQ}{4\pi D(2T_s)} \left(\frac{-x^2}{4D(2T_s)}\right). \quad (4)$$

CCI can be quantified as per (5) if the interfering source is located at the distance x_1 . Again, the probability of interfering source transmitting symbol '1' during the reception period is considered to be p .

$$CCI(x_1, T_s) = \frac{pQ}{4\pi D(T_s)} \left(\frac{-x_1^2}{4D(T_s)}\right). \quad (5)$$

As shown by ShahMohamaddian et al. [11] the noise affecting the received signal in MC has a Gaussian distribution with zero mean and variance equal to mean of the received signal strength when $Q \gg 1$.

$$N_0 \sim \text{Gaussian}(0, \langle S(d, t) \rangle). \quad (6)$$

If both symbols, i.e., '1' and '0' are considered to be equally likely, then $p = 0.5$. Also, large Q is considered in simulations (described in Sect. 6) to make a symbol detectable even at higher distances. Due to these conditions, noise in (6) and

interferences in (4) and (5) can be characterized as Gaussian processes. Also, it is instinctively clear that $S(x, T_S)$ should be a Gaussian process. It is well known that linear operations can be performed on a Gaussian process and the result is always another Gaussian process.

4 Connectivity

It is customary to measure communication link quality in terms of SINR which is the ratio of received signal strength to noise plus interference strength. To ensure sufficient link quality SINR should be greater than a threshold value β . Two nodes are said to be connected if the SINR for the link between them satisfies the condition

$$\frac{S}{N_0 + I} \geq \beta. \quad (7)$$

where S = signal strength, N_0 = received noise component, I = Interference component (sum of CCI and ISI). The probability of the two nodes being connected to each other can be expressed as

$$P_{ij} = \text{Prob} \left[\frac{S}{N_0 + I} \geq \beta \right]. \quad (8)$$

Since S , N_0 , and I are independent Gaussian processes, it is possible to compute their linear combinations. Hence (8) can be rearranged as

$$P_{ij} = \text{Prob}[S \geq \beta(N_0 + I)]. \quad (9)$$

Substituting $\beta(N_0 + I) = X$ and rearranging the terms in (9)

$$P_{ij} = \text{Prob}[S - X \geq 0]. \quad (10)$$

Equation (10) can also be expressed as

$$P_{ij} = 1 - \text{Prob}[S - X < 0]. \quad (11)$$

If μ and σ are mean and standard derivation of Gaussian process $(S - X)$, then

$$P_{ij} = 1 - \text{Prob} \left[\frac{S - X - \mu}{\sigma} < \frac{0 - \mu}{\sigma} \right]. \quad (12)$$

Equation (12) can be expressed in terms of moment generating function $\phi(\cdot)$

$$P_{ij} = 1 - \phi\left(\frac{-\mu}{\sigma}\right). \quad (13)$$

Following the definition of P_{ij} , $(1 - P_{ij})$ should be a probability of two nodes NOT being connected to each other. Further, in a network consisting N nodes, with uniform internode connectivity P_x , probability of the node NOT being connected to ANY of the nodes is given as

$$P_{\text{iso}} = (1 - P_x)^{N-1} \quad (14)$$

Complement of P_{iso} is probability of the node being connected to at least one node, that is

$$P_{\text{one}} = 1 - P_{\text{iso}} = 1 - (1 - P_x)^{N-1} \quad (15)$$

Finally, the connectivity of all the nodes in the network being connected to at least one other node or the network connectivity is given as

$$P_{\text{nw}} = (1 - (1 - P_x)^{N-1})^N \quad (16)$$

5 Anomalous Diffusion

The time required for a diffusing particle to cover mean-squared displacement $\langle x^2 \rangle$ is a linear function of time. In anomalous diffusion, mean-squared displacement is a fractional power of time not equal to one [12].

$$\langle x^2 \rangle \sim t^\alpha \quad (17)$$

α is an anomalous diffusion exponent and its value determines the type and extent of anomaly in the diffusion process. When $\alpha < 1$, the effective process is called sub-diffusion, while when $\alpha > 1$ it is referred to as super-diffusion. As $\alpha \rightarrow 1$, diffusion becomes normal. Due to small size of the messenger molecules and the heterogeneous environments in which they have to operate, it is expected that MC-based NMs networks will get affected by sub-diffusion. Therefore, it is essential to model sub-diffusive behavior and to test the network connectivity under it.

Pagnini et al. [7] have employed master equation approach to find the solution of anomalous diffusion equation in complex media. The process of anomalous diffusion is characterized as a super position mechanism representing inhomogenous and nonstationary nature of the medium. The solution emerges as a natural generalization of Gaussian distribution as shown below:

$$C(x, t) = \frac{1}{\sqrt{4\pi D_\alpha^{1-\alpha/2} t^\alpha}} \exp\left[\frac{-x^2}{4D_\alpha^{1-\alpha/2} t^\alpha}\right]. \quad (18)$$

where α = Anomalous diffusion exponent, D_α = Time-dependent diffusion coefficient. Equation (18) represents modification to (4) after taking sub-diffusion into account.

6 Results

To represent various scenarios on the presence of interference, four cases are considered which are shown in Table 1. This approach allows us to observe effects of CCI and ISI on the connectivity separately. Equation (13) is used to determine P_{ij} . Received signal (S) and noise (N_0) are computed using (3) and (6), respectively.

As proposed in [1], pheromones can be used as messenger molecules for long range (transmission range from few millimeters to few kilometers) MC. For simulation, *Pogonomyrmex Badius* harvester ant's alarm pheromone ($D = 0.43 \text{ cm}^2/\text{s}$) is considered. Bossert and Wilson [13] reported that the ratio of pheromonal concentration released to minimum concentration required at the receptors for initiation of a response is approximately 1391.11. In a laboratory experiment conducted by them the maximum communication radius observed was 6 cm. For the transmission range, 1–10 cm, $Q = 10^{10}$ molecules and $A_r = 10^{-6} \text{ m}^2$ satisfy minimum concentration for response condition. To characterize anomalous diffusion, $\alpha = 0.7$ is considered. Also, it is assumed that concentration of these physical obstructions is uniform throughout. Signal molecules counted at the receiver should be at least twice in number as that of noise and interference molecules. Hence the β (or the minimum SINR required) value is chosen to be 2.

First, P_{ij} is determined by varying d and keeping symbol time T_s constant at 200 s, between transmitter A and nodes 1–8 shown in Fig. 1. For further analysis average node connectivity P_x is obtained by taking average of P_{A1} to P_{A8} . Result is as shown in Fig. 2. It is expected that the P_x should be higher for lower values of d , however, as seen from Fig. 2, contradictory results are obtained. For all the cases, smaller internode distances ($d < 6 \text{ cm}$) exhibit lower connectivity due to the choice of symbol time T_s . For smaller internode distances ($d < 6 \text{ cm}$), small T_s is sufficient

Table 1 Various scenarios on the presence of interference (I)

Case	Interference present	Computation of I
(i)	Only CCI present	By Eq. (4)
(ii)	Only ISI present	By Eq. (5)
(iii)	Both CCI and ISI present	By addition of Eqs. (4) and (5)
(iv)	No interference	$I = 0$

Fig. 2 Average node connectivity versus internode distance

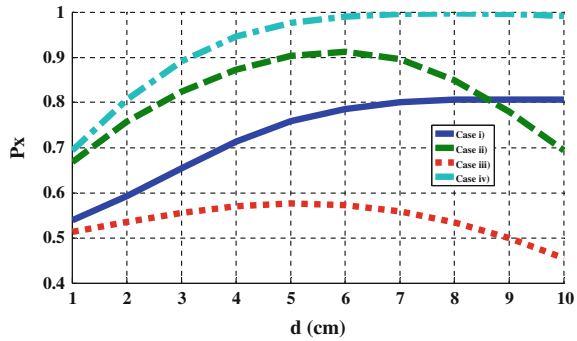
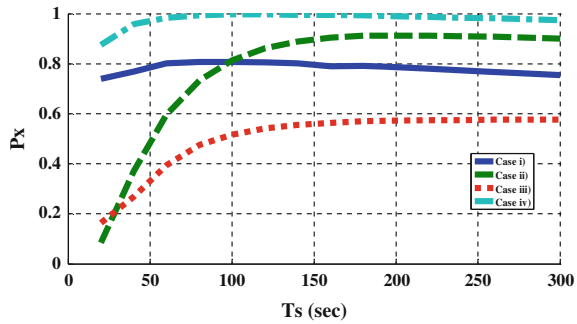


Fig. 3 Average node connectivity versus symbol time



since the message reaches receivers faster. However, increasing T_s further deteriorates performance since actual signal reception occurs only in the initial period and for the rest of the time, interference and noise are being received.

All the plots in Fig. 2 show similar trends, however, maximum P_x is 0.9999 in case (iv) and 0.57 in case (iii). Higher P_x is obtained in case (ii) is 0.91 compared to case (i) where it is 0.80, however, in case (ii) connectivity deteriorates at the faster rate as d increases beyond 7 cm. Average node connectivity in case (ii) falls below that of case (i) when $d > 8$ cm. These results indicate that presence of CCI and ISI degrades the network connectivity by 42.9 %. Hence it becomes necessary to mitigate the interferences. However, case (ii) exhibits connectivity of 0.9 when d is in the range of 5–8 cm.

Next, P_{ij} is determined by varying T_s (20–300 s) and keeping d at 6 cm. P_x is an average of P_{A1} to P_{A8} . As seen from Fig. 3, in all the four cases P_x increases as T_s increases, reaches a maximum value, and starts reducing slowly. The receiver has to wait for sufficient time till the required molecular concentration is built. The amount of waiting time depends on receivers distance from transmitter. At $d = 6$ cm, T_s of 80–200 s is required for all the cases. Once the molecular pulse has passed receiver location, $C(x,t)$ reduces. Due to which reduction in P_x is observed for higher values of T_s . Case (ii) shows rapid improvement in average node connectivity as T_s increases from 20 to 150 s. Since ISI arises due to small T_s , it is clear that the

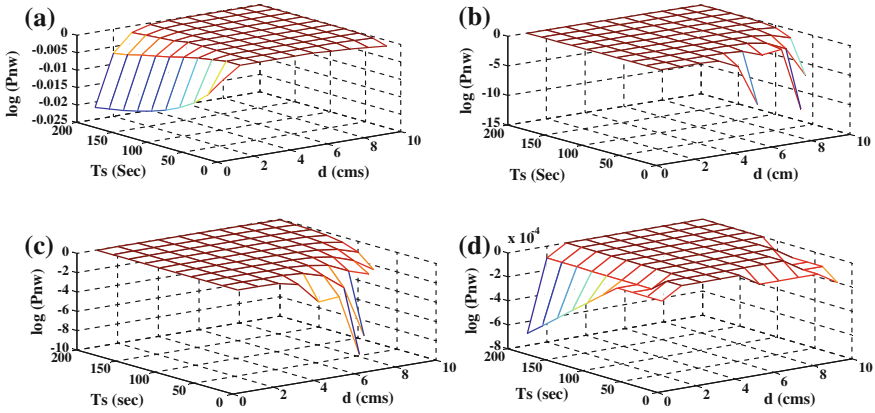


Fig. 4 P_{nw} for various d and T_s values **a** case (i), **b** case (ii), **c** case (iii), **d** case (iv)

performance in this case should improve with increasing T_s . Average node connectivity of case (iii) enhances with T_s so much so that it surpasses case (i) and reaches the maximum value of 0.95 (only 4 % less than that of case (iv)) which accounts for no interference condition. Since case (iv) is very ideal case, maximum node connectivity can be obtained by case (ii) with $d = 6$ cm and $T_s = 200$ s. Results of P_x are used to obtain P_{nw} using (16). Figure 4a–d are $\log(P_x)$ values obtained from varying d and T_s . All the four figures show a region in which P_{nw} is the maximum. In Case (i), P_{nw} shows little variation except for the region where $d < 2$ cm. As d increases further, P_{nw} remains more or less constant since relative distance of the nodes from transmitter and interfering source remains constant.

For $d < 2$ cm, degrading effect of CCI becomes prominent with increasing T_s . Figure 4b shows that the ISI affects P_{nw} for higher values of d in which T_s is small. Therefore, in case (ii) for increased transmission range, T_s should be greater than 100 s. Case (iii) is a combination of case (i) and case (ii), however, from Fig. 4c it is clear that ISI is the dominant interference than the CCI. Case (iii) represents realistic operating conditions hence it can be safely assumed that mitigating only ISI should be sufficient to obtain P_{nw} of 0.9999. Moreover, transmission range greater than 6 cm can be obtained if $T_s > 100$ s. Figure 4d P_{nw} is maximum except for the regions ($d = 1$ cm, $T_s > 100$ s) and ($d > 10$ cm, $T_s = 20$ s) where it is slightly less than 0.9999. This reduction is due to presence of noise in the reception process. Noise will have a degrading effect on P_{nw} when signal strength is low. It is clear enough that in the region where d is very small and d is very large, signal strength must be very low.

Figure 5 shows the plots of P_x versus T_s for $d = 1, 6,$ and 10 cm under normal and sub-diffusive regimes. Since sub-diffusion reduces the velocity, molecules dwell for a longer time at each point of their passage. This improves required molecular built up in the vicinity of the receivers. As the symbol time increases, crowding of molecules increases, increasing P_x further. However, for larger d ,

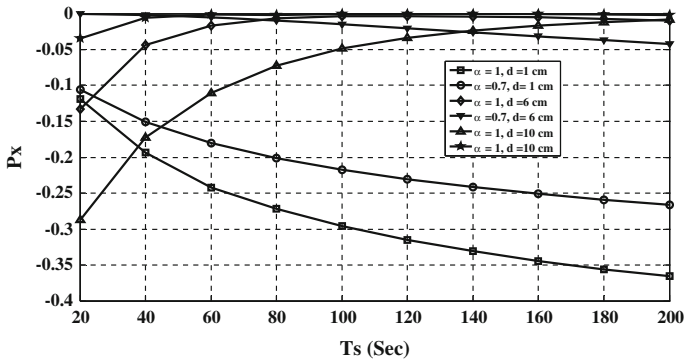


Fig. 5 P_x versus T_s under normal and sub-diffusion regime

sub-diffusion performs better than normal diffusion only up to certain T_s . After this value, normal diffusion performs better than sub-diffusion. Presence of physical obstructions can be a favorable condition in MC-based NM networks if symbol time is adjusted accordingly.

7 Conclusions

MC-based NM networks are expected to work in aqueous media sometimes crowded with physical obstacles. It is important to investigate connectivity of such networks in order to have reliable communication. In this paper, a probabilistic connectivity model for MC-based NM network has been developed through SINR analysis. Some of the previous papers have worked on single link connectivity but have not considered multi-node environment, while the one which considers multi-node environment has not considered effect of noise and interference.

This paper has modeled network connectivity in the presence of noise, interference, and environmental conditions like physical obstructions. Using MATLAB simulations effect of internode distance (d) and symbol time (T_s) on network connectivity is observed. It is found that ISI is the dominating factor degrading the network connectivity than the CCI. Also, results have shown that selection of T_s is crucial and should depend on d , for higher network connectivity. Network connectivity for various internode distances is studied under sub-diffusive regime. It is revealed that presence of physical obstructions can be helpful for improving network connectivity in MC-based NM networks if symbol time is adjusted accordingly.

At the outset, consideration of interference in SINR analysis and study of sub-diffusive behavior of an NMs network are the major contributions of this work. However, there is still a scope to extend this work further. In many practical scenarios, orderly deployment of NMs cannot be guaranteed. Hence network

connectivity in randomly deployed NMs network needs to be investigated. Geographical coverage obtained by a network is also important. Parameters like internode distance and symbol time need to be optimized for required coverage and connectivity.

References

1. Akyildiz, I.F., Brunetti, F., Blázquez, C.: Nanonetworks: a new communication paradigm. *Comput. Netw.* **52**(12), 2260–2279 (2008). Elsevier
2. Nakano, T., Moore, M., Enomoto, A., Suda, T.: Molecular communication as biological ICT. In: Sawai, H. (ed.) *Biological Functions for Information and Communication Technologies*, vol. 320, pp. 49–86. Springer (2011)
3. Hiyama, S., Moritani, Y.: Harnessing biochemical materials to engineer biomimetic communication systems. *Nano Commun. Netw.* **1**(1), 20–30 (2010)
4. Farsad, N., Guo, W., Eckford, A.: Tabletop molecular communication: text messages through chemical signals. *PLoS ONE* **8**(12), e82935 (2013). doi:[10.1371/journal.pone.0082935](https://doi.org/10.1371/journal.pone.0082935)
5. Balasubramaniam, S., Lyamin, N., Kleyko, D., Skurnik, M., Vinel, A., Koucheryavy, Y.: Exploiting bacterial properties for multi-hop nanonetworks, *IEEE Commun. Mag.* 184–191 (2014)
6. Nakano, T., Moore, M., Wei, F., Vasilakos, A., Shuai, J.: Molecular communication and networking: opportunities and challenges. *IEEE Trans. Nanobiosci.* **11**(2), 135–148 (2012)
7. Pagnini, G., Mura, A., Mainardi, F.: Generalized fractional master equation for self-similar stochastic processes modelling anomalous diffusion. *Int. J. Stochast. Anal.* **2012**, (2012). doi:[10.1155/2012/427383](https://doi.org/10.1155/2012/427383)
8. Kuran, M., Yilmaz, H., Tugcu, T., Akyildiz, I.: Interference effects on modulation techniques in diffusion based nanonetworks. *Nano Commun. Netw.* **3**(1), 65–73 (2012)
9. Kuran, M., Yilmaz, H., Tugcu, T., Ozerman, B.: Energy model for communication via diffusion in nanonetworks. *Nano Commun. Netw.* **1**(2), 86–95 (2010)
10. Walsh, F., Balasubramaniam, S.: Reliability and delay analysis of multihop virus-based nanonetworks. *IEEE Trans. Nanotechnol.* **12**(5), 674–684 (2013)
11. ShahMohammadian, H., Messier, G.G., Magierowski, S.: Optimum receiver for molecule shift keying modulation in diffusion-based molecular communication channels. *Nano Commun. Netw.* **3**(3), 183–195 (2012)
12. Sokolov, I.: Models of anomalous diffusion in crowded environments. *Soft Matter* **8**, 9043 (2012). doi:[10.1039/c2sm25701g](https://doi.org/10.1039/c2sm25701g)
13. Bossert, W., Wilson, E.: The analysis of olfactory communication among animals. *J. Theor. Biol.* **5**, 443–469 (1963)

Application of Locally Weighted Regression for Predicting Faults Using Software Entropy Metrics

Arvinder Kaur, Kamaldeep Kaur and Deepti Chopra

Abstract There are numerous approaches for predicting faults in the software engineering research field. Software entropy metrics introduced by Hassan (Predicting faults using the complexity of code changes, 78–88, 2009) [1] are also popularly used for fault prediction. In previous studies, statistical linear regression (SLR) and support vector regression (SVR) for predicting faults using software entropy metrics have been validated. However, other machine learning approaches have not yet been explored. This study explores the applicability of locally weighted regression (LWR) approach for predicting faults using the software entropy metrics and compares it with SVR. It is noticed that the LWR performs better than SVR in most of the cases.

Keywords Software fault prediction · Locally weighted regression · Software entropy

1 Introduction

Fault prediction is an important research field. There is no software that is free from faults. It is essential to predict these faults or determine the fault-prone files/subsystems in the software system so as to reduce the testing time and resources. The fault-prone subsystems should be tested thoroughly and more carefully. More testing efforts should be spent on faulty subsystems.

A. Kaur (✉) · K. Kaur · D. Chopra
University School of Information and Communication Technology (U.S.I.C.T),
Guru Gobind Singh Indraprastha University (G.G.S.I.P.U.), New Delhi, India
e-mail: arvinderkaurtakkar@yahoo.com

K. Kaur
e-mail: researchkdk2@gmail.com

D. Chopra
e-mail: dchopra27@gmail.com

Like faults, changes to the software are also inevitable. The software may be changed to add functionality or to remove bugs. The complexity of these changes is used to determine the fault-prone files or subsystems. In general, the greater the complexity of change, the more faulty is the file/subsystem likely to be. This complexity of changes is quantified using software entropy metrics as described by Hassan [1]. The entropy-based metrics are then used to predict faults.

Hassan [1] has proposed a method to quantify the complexity of changes using entropy in his study. He has also validated the approach using simple linear regression (SLR). It was found that the software entropy metrics were better predictor of bugs than prior faults or modifications. D'Ambros et al. [2] evaluated various fault prediction approaches including the software entropy approach, software code metrics, process metrics, and previous faults. A benchmark dataset was also released by them that consist of faults, change, and version information of five software systems: Eclipse JDT Core, Eclipse PDE UI, Equinox framework, Mylyn, and Apache Lucene.

Singh and Chaturvedi [3] used the software entropy metrics proposed by Hassan [1] to predict bugs using support vector regression (SVR). They showed that SVR performs better than the simple linear regression (SLR). Other machine learning techniques were not explored by them.

Our study tries to explore the performance of locally weighted regression (LWR) approach given by Atkeson et al. [4] for predicting faults using software entropy metrics. This paper compares LWR approach with the SVR approach.

The rest of this paper is structured into the following sections: Sect. 2 describes the software entropy metrics that are used for predicting faults. Section 3 explains the prediction process and analyzes the results. Section 4 lists some threats to validity of the results. The study is concluded along with directions for future work in Sect. 5.

2 Software Entropy Metrics

Software entropy metrics [1] are used to quantify the complexity of code change process. A file is changed when a fault is to be rectified, a new functionality is to be added, or when there are any changes in formatting and coding styles. The most important and complex of these are the changes that take place when a new functionality is to be added. The complexity of such changes is measured in terms of entropy and metrics and are derived based on it.

2.1 Entropy of Changes

Entropy for a period of changes in a system/subsystem is calculated using the Shannon's entropy formula as given in (1). The period is taken as one year for this study.

$$E_n(P) = - \sum_{k=1}^n (P_k * \log_2 P_k). \tag{1}$$

where $P_k \geq 0$ and $\sum_{k=1}^n P_k = 1$

P_k is the probability that the k th file is changed in the specified period, i.e., the number of times k th file is changed divided by the total number of changes. The entropy is normalized using (2), so that it is possible to compare the entropy of subsystems that contain different number of files.

$$E(P) = \frac{1}{\text{Maximum Entropy}} * E_n(P) = - \frac{1}{\log_2 n} * \sum_{k=1}^n (P_k * \log_2 P_k). \tag{2}$$

such that, $0 \leq E \leq 1$. E is the value of normalized entropy.

2.2 History Complexity Metric

History complexity metric (HCM) is derived from the entropy of the subsystem/system. It is the effect of complexity of changes as assigned to each file in the subsystem/system. The history complexity period factor (HCPF_{*i*}(*j*)) for a file *j* during period *i* is derived using (3).

$$\text{HCPF}_i(j) = \begin{cases} C_{ij} * E_i, & j \in F_i \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where E_i is the entropy of the system/subsystem for period i and C_{ij} is the contribution of E_i that is assigned to a file j that is modified in period i . Three variants of HCPF used for calculating HCM are defined by varying the definition of C_{ij} . Table 1 describes the three variants of HCM.

The HCM for a file j for the evolution period set $\{s, \dots, t\}$ is defined as

$$\text{HCM}_{\{s, \dots, t\}}(j) = \sum_{i \in \{s, \dots, t\}} \text{HCPF}_i(j). \tag{4}$$

Table 1 HCM variants

HCM metric	Description
HCM1	$C_{ij} = 1$, for all files that are modified in period i
HCM2	$C_{ij} = P_j$, where P_j is the probability of change for file j in period i
HCM3	$C_{ij} = 1/F_i$, for all files that are modified in period i where F_i is the number of files that are modified in period i

Similarly, HCM for a system/subsystem S for the evolution period set $\{s, \dots, t\}$ is defined as the summation of HCMs for all files in that subsystem as in (5).

$$HCM_{\{s, \dots, t\}}(S) = \sum_j \in SHCM_{\{s, \dots, t\}}(j). \tag{5}$$

3 Prediction Process

The process to be followed for predicting faults using software entropy metrics is listed in this section. Figure 1 illustrates the prediction process. The three major phases are data extraction, metrics calculation, and regression. Each phase is described in detail in the following subsections:

3.1 Phase I—Data Extraction

The first step of this phase is to select the software system/subsystem on which the prediction results will be validated. The “mozilla/layout/generic” subsystem of the Mozilla project is selected to carry out this study. Mozilla is a popular web browser written mostly in C++. This subsystem of Mozilla contains 132 files.

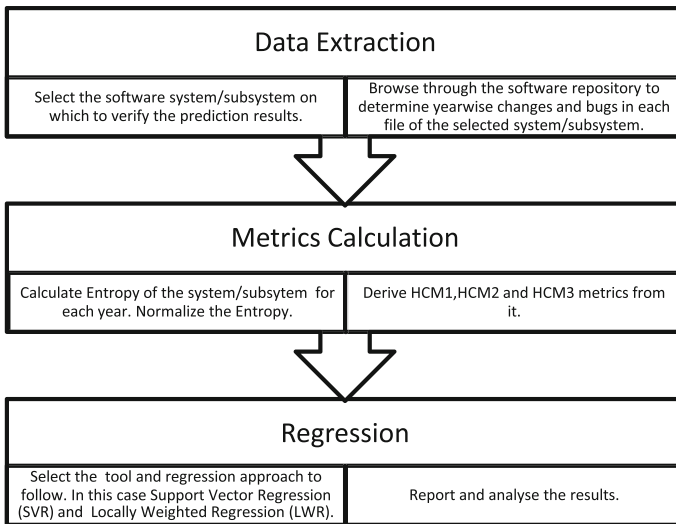


Fig. 1 Prediction process

The next step is to browse through the repository to determine the number of bugs and number of changes per year for each file in this subsystem. The data is collected for the years 2007–2014. Mercurial repository for Mozilla source code, i.e., Mozilla-central [5] is browsed for extracting this information.

3.2 Phase II—Metrics Calculation

In this phase, first, entropy is calculated for the selected subsystem for each year using (1). This entropy is normalized as specified in (2). Then HCM metrics, i.e., HCM1, HCM2, and HCM3 are calculated using the formulas specified in Eqs. (3), (4) and (5). The normalized entropy and number of changes for each year are specified in Table 2.

3.3 Phase III—Regression

The first step in this phase is the selection of tool and regression approaches. Weka 3.6 [6] is used to perform regression on the calculated metrics for bug prediction. Weka is a popular data mining tool and provides a suite of machine learning algorithms. The intention is to compare the support vector regression (SVR) approach with locally weighted regression (LWR) approach. These two regression approaches are described and the results are analyzed in the following subsections.

Support Vector Regression (SVR): Weka uses SMOReg to implement SVR, the detailed explanation of it is given by Shevade et al. [7]. The aim of SVM is the estimation of function $f(x)$ for the training dataset $\{x_i, d_i\}$ where x_i is the i th input and d_i is the i th target, such that it is as close as possible to the target value and is also maximally flat to provide good generalization.

$$f(x) = w \cdot \varphi(x) + b. \quad (6)$$

Table 2 Normalized entropy and number of changes

Year	Changes	Normalized entropy
2007	138	0.8493
2008	73	0.6334
2009	58	0.5583
2010	61	0.5556
2011	504	0.8500
2012	1005	0.8207
2013	22	0.5459

where, b is the bias and w is the coefficient vector. Also, $z = \varphi(x)$ is the feature space vector. All computations are done by the kernel function which is defined as

$$K(x, \hat{x}) = \varphi(x) \cdot \varphi(\hat{x}). \tag{7}$$

where, \cdot denotes the dot product in feature space. The primal optimization problem for SVR assuming ε -intensive loss function is to minimize

$$\frac{1}{2} \|w\|^2 + C \sum (\xi_i + \widehat{\xi}_i) \tag{8}$$

such that the following constraints are met: $d_i - w \cdot z_i - b \leq \varepsilon + \xi_i$, $w \cdot z_i + b - d_i \leq \varepsilon + \widehat{\xi}_i$ and $\xi_i, \widehat{\xi}_i \geq 0$.

The primal problem is hard to solve due to infinite-dimensional z and w . Therefore, a finite dimensional optimization called dual problem is defined in (9) using Lagrange multipliers $(\alpha_i, \widehat{\alpha}_i)$.

$$\text{Maximise } \sum_i d_i(\alpha_i - \widehat{\alpha}_i) - \varepsilon \sum_i \alpha_i + \widehat{\alpha}_i - \frac{1}{2} \|w(\alpha, \widehat{\alpha})\|^2 \tag{9}$$

where, $w(\alpha, \widehat{\alpha}) = \sum_i (\alpha_i - \widehat{\alpha}_i) z_i$. The following constraints should be met: $\sum_i (\alpha_i - \widehat{\alpha}_i) = 0$ and $\alpha_i, \widehat{\alpha}_i \in [0, C]$ for every i . C is the coefficient of smoothness.

Weka provides four different kernels for numeric data: PolyKernel, NormalizedPolyKernel, Puk, and RBFKernel and these have been used for our analysis.

Locally Weighted Regression (LWR). Weka also implements LWR that performs regression using locally weighted learning (LWL) as described by Atkeson et al. [4]. LWL is a lazy learning approach since it defers in processing unless a query is to be answered. It is a local model that tries to fit the training instances only around the query point. The weightage to training points is calculated using distance function with nearby points getting more weightage. LWR tries to fit a surface to neighboring points using distance weighted regression. The weighting function is known as kernel function (K).

There are two approaches to weighting:

- *Weighting the error criterion*: The error criterion is minimized using (10).

$$C(q) = \sum_i (\widehat{y}_i - y_i)^2 K(d(x_i, q)) \tag{10}$$

- *Direct Data Weighting*: The training data is directly assigned weights using the kernel function as specified in (11).

$$\hat{y}(q) = \frac{\sum y_i K(d(x_i, q))}{\sum K(d(x_i, q))} \tag{11}$$

where, x_i is the i th input vector, y_i is the i th training data, and $d(x_i, q)$ is the distance function.

Weka provides six kernel functions for LWR: Linear, Epanechnikov, Tricube, Inverse distance, Gaussian, and Constant weighting which are used for our analysis.

Result Analysis: The results of regression are measured in terms of mean absolute error (MAE) and root-mean-square error (RMSE) according to the definition in Witten et al. [8].

- Mean Absolute Error (MAE): $\frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}$ (12)

- Root Mean Square Error (RMSE): $\sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}$ (13)

The MAE and RMSE for SVR are specified in Table 3 and those for LWR are specified in Table 4.

Table 3 SVR results

Metrics	Kernel	Poly	NormalizedPoly	Puk	RBF
HCM1	MAE	125.09	687.69	172.61	577.47
	RMSE	173.62	719.97	196.12	604.44
HCM2	MAE	186.81	591.69	228.22	579.86
	RMSE	199.65	656.85	260.06	606.20
HCM3	MAE	186.71	682.93	227.81	579.85
	RMSE	199.51	718.99	259.58	606.19

Table 4 LWR results

Metrics	HCM1		HCM2		HCM3	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Linear	135.91	177.01	136.63	171.99	136.62	171.97
Epane-chnikov	136.80	178.25	135.15	172.05	135.14	172.02
Tricube	136.29	178.07	132.31	166.87	132.31	166.85
Inverse-distance	137.01	177.97	138.71	177.56	138.71	177.55
Gaussian	136.96	178.34	136.79	174.94	136.79	174.93
Constant weighting	137.95	179.34	137.95	179.35	137.95	179.34

The locally weighted regression (LWR) approach performs better than the support vector regression (SVR) approach [9]. The MAE and RMSE in case of LWR is less than that for SVR for all cases except when the PolyKernel is used with HCM1 as predictor. Therefore, it can be concluded that LWR approach can be used for predicting faults using software entropy metrics. These values are plotted in Figs. 2, 3, 4, 5, 6, and 7.

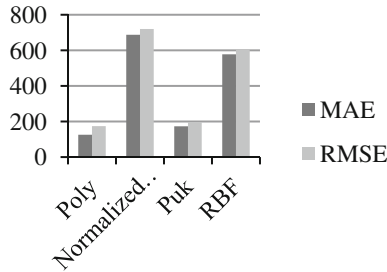


Fig. 2 MAE and RMSE for SVR with HCM1

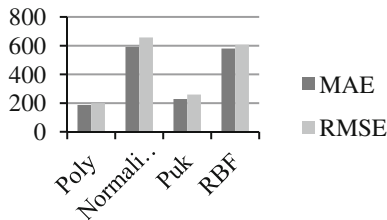


Fig. 3 MAE and RMSE for SVR with HCM2

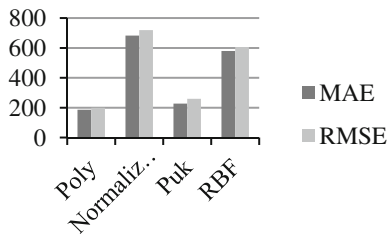


Fig. 4 MAE and RMSE for SVR with HCM3

Fig. 5 MAE and RMSE for LWR with HCM1

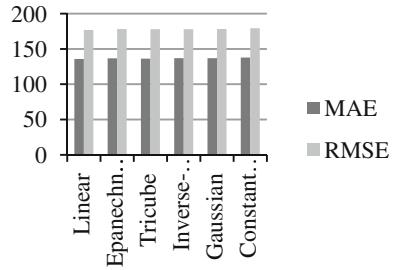


Fig. 6 MAE and RMSE for LWR with HCM2

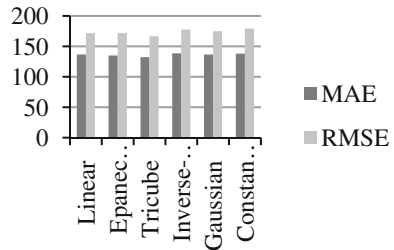
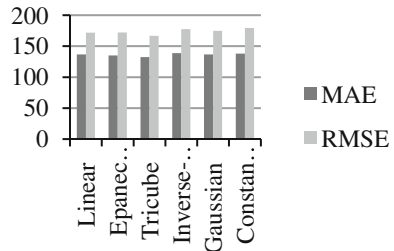


Fig. 7 MAE and RMSE for LWR with HCM3



4 Threats to Validity

The limitations that can affect the validity of the study include the fact that the repository of changes for the Mozilla project, i.e., Mozilla-central is browsed manually and changes are classified based on the keyword. The keywords ‘Bug’, ‘Fault’, ‘fix,’ and ‘Defect’ are used to classify the change as one which repairs fault. Similarly keywords like ‘add’, ‘function’, ‘enhance’, ‘implement’ are used for classifying a change as one that adds functionality.

Also, the study excludes any bugs that were reported but never fixed from consideration while calculating the frequency of faults per year. But since the chance of this happening is almost negligible, the results of the study are justifiable.

5 Conclusion and Future Work

The study introduces locally weighted regression (LWR) approach for predicting faults using software entropy metrics as predictor. It is observed that LWR performs better than support vector regression (SVR) in all cases except one. This concludes LWR approach is a powerful regression approach for predicting faults using software entropy metrics.

It is suggested that LWR approach should be validated for other open source software systems/subsystems. Also, other regression approaches should be compared with it.

References

1. Hassan, A.E.: Predicting faults using the complexity of code changes. In: 31st International Conference on Software Engineering, pp. 78–88. IEEE Computer Society (2009)
2. D'Ambros, M., Lanza, M., Robbes, R.: Evaluating defect prediction approaches: a benchmark and an extensive comparison. *Empirical Softw. Eng.* **17**(4–5), 531–577 (2012)
3. Singh, V.B., Chaturvedi, K.K.: Entropy based bug prediction using support vector regression. In: ISDA, pp. 746–751 (2012)
4. Atkeson, C.G., Moore, A.W., Schaal, S.A.: Locally weighted learning. *AI Rev.* **11**, 75–113 (1997)
5. The Mozilla-Central repository. <http://hg.mozilla.org/mozilla-central/file/9ee9e193fc48/layout/generic>
6. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. *ACM SIGKDD Explor. Newsl.* **11**(1), 10–18 (2009)
7. Shevade, S.K., Keerthi, S.S., Bhattacharyya, C., Murthy, K.R.K.: Improvements to the SMO algorithm for SVM regression. *IEEE Trans. Neural Netw.* **11**(5), 1188–1193 (2000)
8. Witten, I.H., Eibe, F.: *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco (2005)
9. D'Ambros, M., Lanza, M., Robbes, R.: An extensive comparison of bug prediction approaches. In: 7th IEEE Working Conference on Mining Software Repositories (MSR), pp. 31–41. IEEE (2010)

Effect of Defects on Current-Voltage Characteristics of a Silicene ZNR-Based Field Effect Transistor

E. Meher Abhinav, M. Chandra Mohan, A. Suresh Reddy, Vemana Chary and Maragani Thirupathi

Abstract In this paper, we investigated the behavior of negative differential resistance (NDR) and analysis on various deformations like twist, wrap, and ripple/buckler and defects like vacancy and rough edge on short channel bilayer silicene zigzag nanoribbon (ZNR). Effects are caused by deformations like wrap with 5° and by rippling the channel by 0.5 \AA amplitude on 6 nm silicene. FET is evaluated by density functional theory (DFT) and by nonequilibrium green's function (NEGF) approach. We studied the I–V characteristics of deformations and defects. These characteristics of device with different conditions and mainly negative differential resistance (NDR) behavior are studied.

Keywords Germanene · Zigzag nanoribbon (ZNR) · Negative differential resistance (NDR) · Metal-oxide-semiconductor field-effect transistor (MOSFET) · Non-equilibrium green's function (NEGF) · Density functional theory (DFT)

1 Introduction

On creating different defects in silicene zigzag nanoribbon (ZNR), values of the maximum current decrease. For rough edge defects, the decrement is maximum followed by vacancy and then by ripple defect. The current for twist and wrap defects exceed the current at normal operation for a short range of voltages in the linear region of operations. For ripple defect, the current decreases for each value of drain voltage. Transmission pathway shows the movement of electrons in the silicene ZNR after the introduction of vacancy defect in it. Some electrons rebound back after hitting the vacancy and others change their direction.

E.M. Abhinav (✉) · M.C. Mohan · A.S. Reddy · V. Chary · M. Thirupathi
Department of Electronic and Communications, Malla Reddy College of Engineering,
Dhulapally, Hyderabad 500014, Telangana, India
e-mail: abhi4abhi09@gmail.com

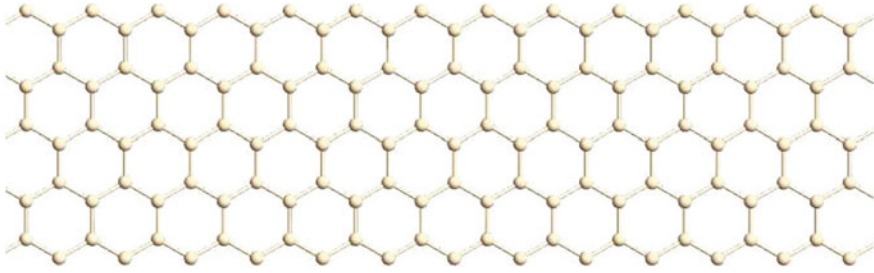


Fig. 1 Top view of the silicene ZNR

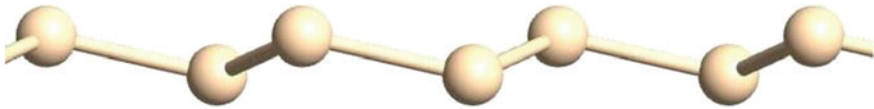


Fig. 2 Edge view of the silicene ZNR

Researchers moved from BJT to MOSFET because it had less power dissipation. Now, as the size of transistor is decreasing rapidly, many problems like leakage power dissipation, DIBL have become prominent [1]. As a solution to these problems some devices were proposed, i.e., HEMT, FINFET, SPINFET, etc. [2–4]. It was also suggested to use a different material instead of silicon in devices and then materials like graphene, silicene, and germanene came into the picture because of their ultrathin structure [5–7]. Graphene, silicene, and germanene are the monolayers of the atoms of graphite, silicon, and germanium, respectively. Silicene has advantages over graphene and germanium in that its fabrication would be consistent with the modern fabrication techniques. Unlike graphene, structure of silicene and germanene is buckled; Figs. 1 and 2, and their band gap can be altered by applying an electric field perpendicular to their plane [8–10]. Buckling is more in case of germanene [11, 12]. After optimization of the silicene ZNR, using ATK, the buckling was 0.5 \AA , which is also the value given in most works [13, 14]. The band gap of silicene is 100 meV.

In this work, we have considered a field effect transistor (FET) with a silicene ZNR as a material in the channel region, Fig. 3. We have used density functional

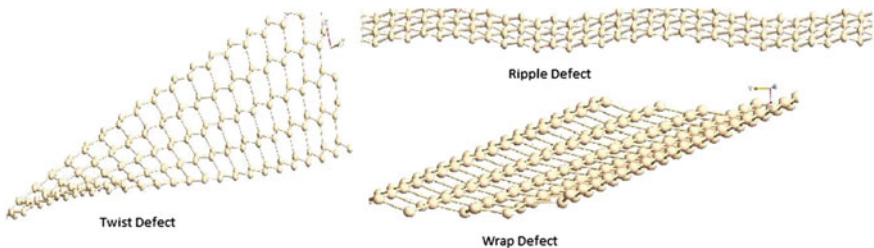


Fig. 3 Twist, ripple, and wrap defects

theory calculations, using ATK, to study the current-voltage characteristics of this transistor and the effect of defects like twist, ripple, and wrap in silicene ZNR on it [15]. Variation of current with drain voltage in silicene with no defects has been shown [16].

2 Model

Structure of the FET is like a traditional MOSFET. It consists of source, drain, gate, and substrate. Since it is a monolayer material device, the length of the material, silicene, is equal to the channel length, i.e., 6 nm, oxide thickness is 6 Å and the oxide used is hafnium oxide (HfO_2). The width of the silicene taken was 1.5 nm. The left box is a source electrode which was set to 0 V; box at the right end is a drain whose voltage was varied from 0 to 0.2 V. Normal operation simulations were done for two gate voltages (V_g), i.e., 0.1 and 0.15 V, but for defects only 0.1 V, gate voltage was used. Figure 3 shows the defects which were created for the study (Fig. 4).

For creating ripple defect, maximum displacement and number of periods of ripple was taken to be 0.5 Å and 2, respectively, for wrap, an angle of 5° was used, and for twist also the angle was 5° .

3 Results and Discussion

Figure 5 shows the current-voltage characteristics for the silicene-based FET. The characteristics can be studied by dividing the whole region into three parts. First region, as the drain voltage is increased from 0 to 0.15 V, current starts increasing with $I_D = 0$ at $V_D = 0$. Second region, for the values of drain voltage between 0.15 and 0.175 V current still increases but its rate of increment decreases and after reaching a peak value of 34 μA , at a drain voltage of 0.18 V, it starts decreasing.

Figure 6 shows that in ripple, twist, and wrap defects decrement in current is maximum for ripple. In the first or linear region, the difference between the current



Fig. 4 FET using a channel of silicene ZNR

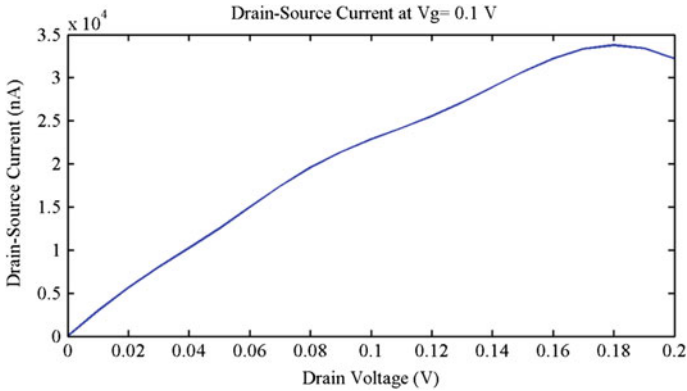


Fig. 5 Variation of drain current with increasing drain voltage for a silicene-based

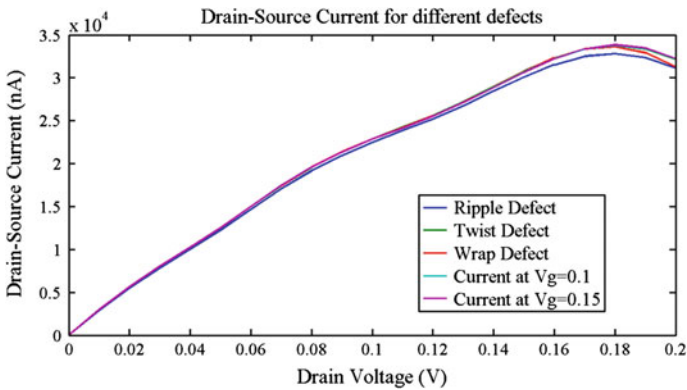


Fig. 6 Current (nA) for different drain voltages (V) for different gate voltages and defects

at normal operation and for all defects is almost negligible. The difference increases in the second and third regions. Ripple current is least in all the three regions of the plot. Even the slope of ripple part is lower than NR for the first and second regions. But in the third region, its slope again becomes equal to the current of normal operation and hence the gap between these two currents remains constant in the third region. Zoomed plot of the third and second regions is shown in Fig. 7.

Twist current is the largest having a drain voltage up to 0.15 V and then its slope decreases relative to the slope of the current at normal operation. At the onset of the third region, twist current at normal operation becomes lesser. The difference between them increases with increasing drain voltage up to the mid of the third region, i.e., 0.18 V and then again the value of the twist current starts to approach the value of current at normal operation.

Wrap current is lesser than the twist current but greater than the current at normal operation till the drain voltage of 0.15 V is at normal operation. After that it

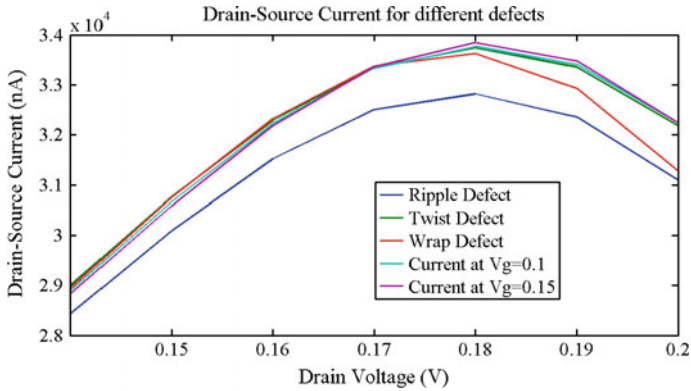


Fig. 7 Current (nA) for a short range of drain voltage (V)

becomes the largest of all current up to the end of the first region. At the onset of the second region, its slope decreases faster and after a drain voltage of 0.18 V its value starts decreasing and at the maximum drain voltage its value is comparable to the ripple current.

Increasing the gate voltage to 0.15 V does not increase the current up to a drain voltage of 0.17 V, but after that in the second region, value of the current for $V_g = 0.15$ is greater than the value of the current for $V_g = 0.1$. Results show that on increasing the gate voltage, the fall in value of the current magnitude for drain voltages above 0.18 V is larger. The current in silicene-based FET decreased significantly for vacancy and rough edge defects, Fig. 8. The maximum values of currents are 3.9 and 2.98 μA for vacancy and rough edge defects. For these cases, currents are significantly different than the currents for other defects and the current at normal operation. Two negative differential resistance (NDR) regions can be seen for these cases. NDR regions are observed whenever there is a quantum confinement

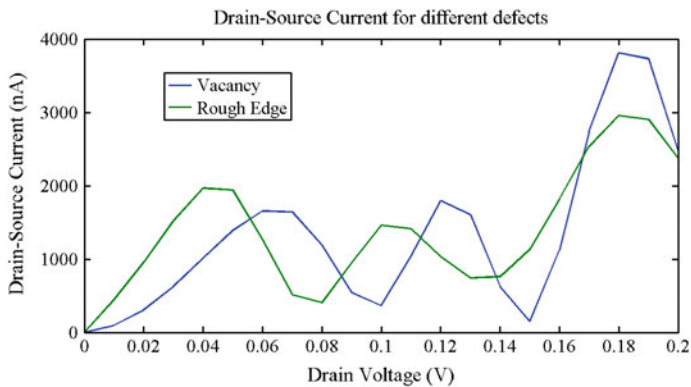


Fig. 8 Variation of current (nA) with drain voltage (V) for vacancy and rough edge defects

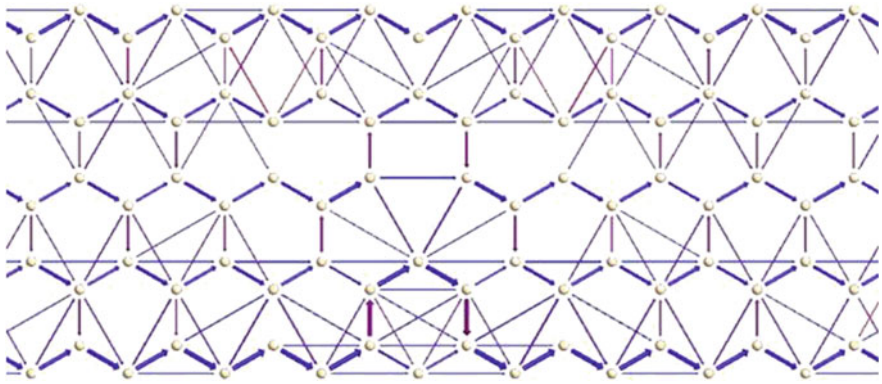


Fig. 9 Path of the movement of electrons in a silicene nanoribbon having vacancy defect

of electrons. Since the nanoribbon is already short, the introduction of vacancy increases the confinement of the electrons, thereby bringing NDR regions in the picture.

Figure 9 shows the transmission pathway of the device with a vacancy defect in the silicene ZNR. An atom is missing from the center due to vacancy. The direction of the arrows shows the direction of the movement of the electrons when a bias is applied across the ends of the nanoribbon.

Since the arrows on the edge of the nanoribbon are thicker, electron current density is maximum at edges. When the electrons moving from source to drain arrive at the vacancy point, a few of them rebound and others shift the direction of their motion which is shown in purple.

4 Conclusions

Results show that the peak value of the drain current decreases with the introduction of defects in silicene ZNR; but in the voltage range of 0–0.17 V, the current is greater than the current at normal operation for twist and wrap defect. Only in cases of vacancy and rough edge defect, two NDR regions were obtained. NDR regions are shifted to left in case of rough edge. Current at high voltages, from 0.18 to 0.2 V, decreases most if the silicene ZNR used has a wrap defect, for other defects current decreases with almost the same slope.

Acknowledgments The authors thank the department of science and technology of the government of India for partially funding this work.

References

1. Ferain, I., Colinge, C.A., Colinge, J.-P.: Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature* **479**(7373), 310–316 (2011)
2. Eom, B.H., Day, P.K., Zmuidzinas, J.: A wideband, low-noise superconducting amplifier with high dynamic range. *Nature* **8**(8), 623–627 (2012)
3. Yamashita, T., Basker, V.S., Standaert, T.: Sub-25 nm FinFET with advanced fin formation and short channel effect engineering. In: 2011 Symposium on VLSI Technology (VLSIT), Honolulu, HI (2011)
4. Geim, A.K., Novoselov, K.S.: The rise of graphene. *Nat. Mater.* **6**(3), 183–191 (2003)
5. Voon, L.C.L., Guzmán-Verri, G.G.: Electronic structure of silicon-based nanostructures. *Phys. Rev. B* **76**(7), 075131 (2007)
6. Kara, A., Enriquez, H., Seitsonen, A.P., Lew Yan Voon, L.C., Vizzini, S., Aufray, B., Oughaddou, H.: A review on silicene—new candidate for electronics. *Surf. Sci. Rep.* **67**(1), 1–18 (2012)
7. Drummond, N.D., Zólyomi, V., Fal'ko, V.I.: Electrically tunable band gap in silicene. *Phys. Rev. B* **85**(7), 075423 (2012)
8. Abhinav, E.M., Chary, D.V.: Strain-induced on germanene monolayer 6 nm short channel FET from first-principle study. In: 2014 International Conference on Circuits, Communication, Control and Computing (I4C), pp. 1, 4, 21–22 Nov 2014. doi:[10.1109/CIMCA.2014.7057743](https://doi.org/10.1109/CIMCA.2014.7057743)
9. Pan, L., Liu, H.J., Tan, X.J., Lv, H.Y., Shi, J., Tangb, X.F., Zhengc, G.: Thermoelectric properties of armchair and zigzag silicene nanoribbons. *Phys. Chem. Chem. Phys.* **14**(39), 13588–13593 (2012)
10. Fang, D.-Q., Zhang, S.-L., Xu, H.: Tuning the electronic and magnetic properties of zigzag silicene nanoribbons by edge hydrogenation and doping. *RSC Adv.* **3**, 24075–24080 (2013)
11. Song, Y.-L., Zhang, S., Lu, D.-B., Xu, H., Wang, Z., Zhang, Y., Lu, Z.-W.: Band-gap modulations of armchair silicene nanoribbons by transverse electric fields. *Eur. Phys. J. B* **86**, 488 (2013)
12. Ni, Z., Liu, Q., Tang, K., Zheng, J., Zhou, J., Qin, R., Gao, Z., Dapeng, Y., Jing, L.: Tunable bandgap in silicene and germanene. *Nano Lett.* **12**(1), 113–118 (2012)
13. Kamal, C.: Controlling band gap in silicene monolayer using external electric field. ArXiv e-prints, no. 1202.2636 (2012)
14. Cahangirov, S., Topsakal, M., Akturk, E., Sahin, H., Ciraci, S.: Monolayer honeycomb structures of group-IV elements and III-V binary compounds: first-principles calculations. *Phys. Rev. B* **80**(15), 155453 (2009)
15. Eom, B.M., Day, P.K., Zmuidzinas, J.: A wideband, low-noise superconducting amplifier with high dynamic range. *Nature* **8**(8), 623–627 (2012)
16. Sengupta, A., Mahapatra, S.: Negative differential resistance and effect of defects and deformations in MoS₂ armchair nanoribbon metal-oxide-semiconductor field effect transistor. *J. Appl. Phys.* **114**(19), 194513 (2013)

Performance Improvement of Read Operations in Distributed File System Through Anticipated Parallel Processing

B. Rangaswamy, N. Geethanjali, T. Rangunathan
and B. Sudheer Kumar

Abstract A distributed system (DS) consists of a set of computers networked together and gives an impression to the users that it is a single system. The main storage component to be used in a DS is distributed file system (DFS). The DFS is used as the main file storage system in the cloud computing systems. Most of the cloud-based applications require a good performance from the DFS as for as read operations are concerned. Many caching and pre-fetching techniques have been proposed in the literature for improving the performance of the DFS. But all of these techniques use synchronous approach which increases read access time. In the literature, a speculation-based technique has been proposed for improving the performance of read access in the DFS which implements asynchronous reading using client-side caching technique. In this paper, we have proposed a new read algorithm for the DFS based on asynchronous reading, client-side caching technique, and anticipated parallel processing. We have carried out performance analysis of the speculation-based algorithm and our proposed algorithm based on anticipated parallel processing. The results of analysis indicate that our proposed algorithm performs better than the earlier speculation-based algorithm.

B. Rangaswamy (✉)

Prosoft Resources Pvt. Ltd, Hyderabad, India

e-mail: burujula1971@gmail.com

N. Geethanjali

Sri Krishnadevaraya University, Anantapur, India

e-mail: geethanjali.sku@gmail.com

T. Rangunathan · B.S. Kumar

ACE Engineering College, Kancheepuram, India

e-mail: ragu_savi@yahoo.com

B.S. Kumar

e-mail: sudheer.itdict@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_28

1 Introduction

A distributed system (DS) consists of a set of computers networked together and gives an impression to the users that it is a single system. Users of the DS are able to share the storage, computing, and other resources of the system. The main storage component to be used in a DS is distributed file system (DFS) which is used for storing the users and system information, and also a DFS can act as the common storage system for the users of the DS and can provide file sharing facility to the users of the DS.

In the emerging cloud computing scenario, a large number of IT organizations are deploying their applications in the cloud computing systems. The DFS is used as the main storage system in such cloud computing systems. Most of the cloud-based applications require a good performance from the DFS as for as file read operations are concerned. This is because most of the web applications deployed in the cloud computing systems perform file read operations on the DFS more frequently, and less frequently these applications carry out file write operations on the DFS. So improving the performance of file read operations in DFS is an important research issue.

Over the years, processing speed has been increased, but disk I/O speed has not increased that level. So the disk I/O is the bottleneck for improving the file read or file write performance of the DFS. Many caching and pre-fetching techniques have been proposed in the literature for improving the file read performance of the DFS. These techniques reduce the disk I/O traffic by allowing the client programs to read the requested data from the cache if it is available. Two types of caching techniques namely server-side caching and client-side caching are discussed in the literature. The server-side caching techniques require a high capacity of main memory to cache all the data read or written by the client programs. Hence, most of the DFSs follow client-side caching as it is more efficient and cost effective than server-side caching. In the literature, client-side caching techniques like collaborative caching or cooperative caching are discussed to improve the file read performance in the DFS. The problem with these approaches is that they follow synchronous approach in which the client programs will have to wait for the remote server file system to send the data by reading from the disks attached with the server system. Even though these caching techniques improve the file read performance, still there is a scope for reducing the file read access time if asynchronous approach is followed.

In [8], authors presented a decentralized hint-based algorithm for cooperative caching. As per this hint-based algorithm, clients were allowed to take local decisions based on hints with less overhead. This hint-based algorithm reduces the manager load, block lookup traffic, and replacement traffic.

In the literature, a speculation-based technique has been proposed for improving the performance of read access in the DFS [6]. This technique implements the asynchronous approach in which the *read* system call issued by the client program reads the requested data (*d*) from the local cache (if it is available) maintained in the system where that client program is getting executed. Simultaneously, the server

system is requested to send the time stamp (T_1) of the latest modified (d). Let us consider that time stamp of (d) available in client-side cache is T_2 , and if T_2 is equivalent to T_1 which indicates that the *read* system call issued by the client program has read the recently modified data, the *read* system call can store the (d) in the local buffer (address is specified in the *read* system call) of the client system. Otherwise, the *read* system call execution is rolled back and it will start reading the (d) from the server file system. This type of asynchronous reading process will reduce the waiting time of the *read* system call issued by the client program.

In this paper, we have proposed a novel read algorithm by considering modern DFS. The DFS that we have considered has got one name node where the global directory system (metadata) is maintained and one or more data nodes. Note that the data nodes consist of local file system and a part of the global DFS. Client programs (applications) can be executed in the data node and the data generated by those programs will be stored in the part of the DFS maintained in that data node or in some other data nodes which will be decided by the name node. The DFS follows replication strategy to implement the reliability feature of the DFS. So the same copy of a file may be available in more than one data nodes. For example, in Hadoop DFS, three copies of a file are maintained in the data nodes by default [9]. In this paper, we have proposed a novel read algorithm for improving the performance of file read operations in the DFS based on anticipated parallel processing. The proposed algorithms carry out two anticipated parallel executions for a *read* system call issued by the client program. One execution (O_1) reads the data from the local cache maintained in the client data node (CDN) (the data node where the client program is getting executed) and the other execution (O_2) reads the data from the cache maintained in one of the data nodes which is very similar to the client data node. Note that we have defined a similarity measure for identifying similar data node (SDN) for a CDN. Meanwhile, the *read* system call will simultaneously request the name node to provide the time stamp (T_1) of the file in which the requested data is stored. Let us assume that the time stamp of the data stored in the cache of the CDN is T_2 and the time stamp of the data stored in the cache of the SDN is T_3 . If T_1 is equivalent to T_2 , then O_1 is allowed to store the data in the local buffer (address is specified in the *read* system call) of the CDN and O_2 will be terminated. If T_1 is equivalent to T_3 , then O_2 is allowed to store the data in the local buffer (address is specified in the *read* system call) of the CDN and O_1 will be terminated. If T_1 is not equivalent to T_2 and T_3 , then O_1 and O_2 will be terminated and the name node will be consulted to get the address of the data node (DN1) where the data is available. Next, the data will be read from the disk that DN1 where the DFS part is stored. We have carried out the performance analysis of the earlier speculation-based algorithm with our proposed algorithms based on anticipated parallel processing. The results of the analysis indicate that our proposed algorithm can perform better than the earlier speculation-based algorithm proposed in the literature.

This paper is organized as follows. In the next section, we describe the techniques discussed in the literature for improving the performance of the DFS. In Sect. 3, we discuss our proposed approach in detail. In Sect. 4, we have done the detailed performance evaluation of the algorithms. Section 5 concludes the paper.

2 Related Work

In this section, first we discuss and describe the techniques discussed in the literature for improving the performance of the DFS.

Many client-side caching techniques have been used to improve the performance of distributed file systems.

In [2], cooperative caching technique is discussed. It is a method in which caches maintained in the client computer systems connected in the local area network (LAN) co-operate to serve the request given by the client applications. This technique works as follows. First, client application verifies in the local cache whether the requested data is available in it or not. If the data is not available in the local cache, then it will be searched in the remote client caches where the data is available. The addresses of the client machines where the data is available will be collected from the server. This technique improves the performance of the DFS. However, this technique relies on the information available in the server for locating remote client for accessing data which leads to single point of failure.

To avoid single point of failure, a decentralized cooperative caching algorithm was proposed [8]. This approach uses hints and in this method, each client system keeps track of the probable location of the each file block and uses this information to check for file blocks in the cooperative cache. The main drawback of this approach is that applications have to be modified to support the caching of the file blocks which are accessed by them from the server.

In [4], a new technique called collective caching is discussed. This technique considers all the tasks that run the same application program as a single client to the server. This allows the data available in the cache to be managed collectively within the client processes and makes every caching operation to be known to all the tasks. The disadvantage of this approach is that the access pattern of individual task may be lost after the client file access requests are aggregated. So this type of operation has a negative impact on the performance of the caching algorithm.

The authors of [7] presented an aggressive proactive mechanism in which authors have taken efforts to shape the file system resource management which is suitable to the needs of I/O intensive applications. According to this mechanism, the I/O intensive applications will inform their future accesses information to the file system to achieve I/O parallelism. The informed future access information is also used to pre-fetch the informed blocks and caching the informed blocks for the purpose of reuse.

In paper [3], authors proposed an approach for cache utilization based on locality of file access patterns. In this approach, file block reuse distance was used to decide the strength of the locality of accessed block, and clients keep strong locality blocks in its cache. In order to balance the utilization of caches across the clients, a block is forwarded from one client to another client, and in case the forwarding, client has high cache utilization than the receiving client.

A speculation-based method has been proposed for improving the performance of read access in the DFS [6]. Speculative reading of the data from the client cache

can improve the performance of distributed file systems by reducing number of disk I/O operations. Rather than making the client program to wait for the server file system, the *read* system call does speculative reading from the client cache and then checking for the validity of the cache copy maintained in the cache by verifying with server. If the copy read from the client cache is the one available in the server file system, then the *read* system call issued by the client program can continue with its current execution. Otherwise, the client program has to execute the same *read* system call again, and this time the server file system will read the data from the disk and then send the same to the client program. This type of asynchronous processing will improve the performance by reducing the waiting time and by avoiding disk access.

The Hadoop DFS (HDFS) is the storage system developed by Hadoop which is an open-source project [9]. HDFS is designed to install on commodity machines with reliability feature implemented. For data-intensive applications, HDFS is able to provide good performance.

3 Proposed Algorithm Based on Anticipated Parallel Processing

In this section, we discuss, regarding anticipated parallel execution, the disadvantages of speculation-based algorithm and then the proposed algorithm.

3.1 Anticipated Parallel Execution

The main idea behind anticipated parallel execution is to do some task before it is known whether that task will be required at all. By doing so, it is possible to avoid a delay that would have happened by doing that task after knowing whether it is required or not. If the task done is not needed, then the effects of the task are undone and the results produced by that task are ignored.

Anticipated parallel execution has been used in pipelined microprocessors while executing conditional branch instructions [1]. In this approach, the branch instructions are allowed to continue in the pipeline. If the condition of the branch instruction is evaluated to be false which is identified in the later stage, then the instruction execution can continue, and if the condition is evaluated to be true, then the pipeline will be drained and then the instruction will be fetched from the target location specified in the instruction. For compiler optimization also, the anticipated parallel processing is used [5].

3.2 *Disadvantages of Speculation-Based Algorithm*

In the literature, a speculation-based method has been proposed for improving the performance of read operations [6]. In this paper, the authors have proposed a speculation-based method in which the read request given by the client application program reads the required data from the cache where the client application program get executed and proceed with the execution in a speculative manner. Meanwhile, the server system is contacted to know whether the read data from the cache is the recent one or old one by comparing the time stamp of the cache data and the time stamp of the data in the server. If both time stamps are matching, then the speculative execution of the client program is allowed to complete, or else the speculative execution is stopped and then the client application program has to contact the server file system for reading the data from the server's disk. This type of processing can improve the performance of the read operations by reducing the waiting time of the client application program and by avoiding disk access. The disadvantage of this algorithm is that it carries out only one speculative execution, and if the data is not available in the local cache, then it has to be read from the server file system. However, the same data may be cached in multiple nodes of the distributed system and reading data from one such cache will reduce the read access time. So it is required to propose an effective algorithm for improving the performance of the read operations by considering the data cached in multiple nodes of the distributed system.

3.3 *Proposed Algorithm*

In this subsection, we describe regarding the assumptions of the caching system maintained in the DFS, three parts of the proposed algorithm, and then the proposed algorithm.

Assumptions:

We consider a DFS which consists of one name node and multiple data nodes and all these nodes connected using switches in a LAN. The name node stores the metadata (global directory) and the data nodes store the data in the files. Also, note that data nodes are used to execute the application programs. Data nodes also execute DFS client programs. The DFS client program is used to communicate with the name node to read the metadata of a file (F1). Once the address of the data nodes (DNs) where F1 is stored is known, then the DFS client program contacts the nearby data node among DN's to read the contents of F1. We consider that three copies of the same file are kept in three different data nodes. Also, we consider that data nodes maintain caches in the main memory and caching will be done whenever a file is read from the disk storage system of a data node and entire file content is stored in the cache. Note that no cache coherence technique is followed in the DFS, in order to avoid the communication and other overheads. Also, we consider that

whenever a data node caches a file, it requests the name node to provide the address of one more data node in which copy of the same file is cached. Each data node maintains a cache directory (CD) in which the metadata of the cached files are stored. Also, for each cached file, address of a physically close data node where a copy of the file is cached is also stored in the CD if the copy of the file is available in any one of the data nodes of the DFS.

Three parts of the algorithm:

Our algorithm consists of three parts. The first part describes the steps to be followed for the main thread of execution of the *read* procedure of the DFS, the second part describes the steps to be followed by the anticipated execution (AE1), and the third part describes the steps to be followed by the anticipated execution (AE2).

/ A client program C running in D1 has issued read procedure to read the contents of the file F2 */*

(I) *Algorithm for main thread of execution:*

Step 1. If E1 and E2 are not created or E1 and E2 are terminated, then the following steps are executed or else go to Step 5.

Step 2. DFS client running in D1 contacts the name node to get the addresses of the data nodes (DNs) where F2 is stored.

Step 3. The DFS client program contacts one of the nearest data nodes among DN's to read the contents of F2 from the disk storage system of that data node.

Step 4. The content available in F2 is transferred to C and also cached in C.

Step 5. Stop.

(II) *Algorithm for anticipated execution (E1):*

Step 1. If F2 is available in the local cache, C can read it (Anticipated parallel execution (E1)).

Let us consider that time stamp value of this cached copy of F2 is T1.

Step 2. E1 will continue until the time stamp value of F2 is received from the name node (T2) or its completion.

Step 3. If $T1 = T2$, then E1 will be allowed to complete its execution; or else E1 will be terminated

Step 4. Stop.

(III) *Algorithm for anticipated execution (E2):*

Step 1. C verifies the CD to get the address of another data node (D2) where F2 is cached.

Step 2. If address of the data node is available in CD, then C can read F2 from the remote cache of D2 and C can start one more anticipated parallel execution (E2) or else goto Step 4. Let us consider that time stamp value of this cached copy of F2 is T3.

Step 3. E2 will continue until the time stamp value of F2 is received from the name node (T2) or its completion.

Step 3. If $T3 = T2$, then E2 will be allowed to complete its execution; or else E2 will be terminated.

Step 4. Stop.

(Note that request message is sent to name node to send the time stamp value of F2 and then both the algorithm steps II and III are executed in parallel).

4 Performance Evaluation

In this section, we discuss regarding the performance evaluation of algorithms through mathematical analysis and simulation experiments.

4.1 Assumptions

Let us consider that a file consists of only ten blocks (block size is 4 KB) for the purpose of performance evaluation. Here, we assume that all the data nodes and the name node are connected in a LAN. We consider that the average communication delay (ACD) required for transferring 4 KB of data is 4 ms and for transferring time stamp and metadata information as 0.125 ms based on the recent analysis by considering the switched LAN. Also, the average time required to access a data block (4 KB size) from the disk storage system is 12 ms by considering the latest disk storage devices. Also, we consider that the average time required to access 4 KB of data block from the main memory is 0.005 ms by considering latest dynamic random access memory technologies. Let us consider that local cache hit ratio as c and remote cache hit ratio as $c1$.

4.2 Mathematical Analysis

Average read access time for a 4 KB data block of DFS (without anticipated parallel processing) = time required to access name node to collect meta data + reading 4 KB data block from a specific data node from the disk + reading 4 KB data block from the main memory of source data node + transferring the 4 KB data block to the destination data node + time required for transferring the data block to client's address space). By following above formula, the average access time is computed as 16.135 ms (0.125 ms + 12 ms + 0.005 ms + 4 ms + 0.005 = 16.135 ms).

Average read access time for a 4 KB data block of DFS (with speculation – earlier approach) = c * (time required to access the local memory + time required to access name node to collect time stamp) + $(1 - c)$ * (time required to access the local memory + time required to access name node to collect time stamp + reading

4 KB data block from a specific data node from the disk + reading 4 KB data block from the main memory of source data node + transferring the 4 KB data block to the destination data node + time required for transferring the data block to client program’s address space).

By following the above formula, the average access time is computed as $(16.26 - 16.13c)$ ms (F1). Note that we have ignored the time required to start speculative execution in these calculations.

For the proposed approach, we have to calculate the time required to access 4 KB of data block from remote memory which is equivalent to time required to access memory + time required to transfer the data from the remote node to the local node + time required for transferring the data block to client’s address space. This time is computed as 4.01 ms $(0.005 \text{ ms} + 4 \text{ ms} + 0.005 \text{ ms})$.

Average read access time for a 4 KB data block of DFS (our anticipated parallel processing approach = $c * (\text{time required to send the request message to the name node to get the time stamp of F2} + \text{time required to access the local memory}) + c1 * (\text{time required to send the request message to the name node} + \text{time required for transferring data block to C from the cache of remote data node} + \text{time required for transferring the data block to client program’s address space}) + (1-c-c1) * (\text{time required to send the request message to the name node to get the time stamp of F2} + \text{time required for checking local cache} + \text{time required for checking near by cache} + \text{time required to access name node to collect meta data} + \text{reading 4 KB data block from a specific data node from the disk} + \text{reading 4 KB data block from the main memory of source data node} + \text{transferring the 4 KB data block to the destination data node} + \text{time required for transferring the data block to client’s address space})$.

By following the above formula, the average access time is computed as $(20.26 - 20.01c - 16.13c1)$ ms (F2).

Based on the formulas (F1 and F2) discussed above, we have evaluated the performance of speculation-based algorithm and our proposed anticipated parallel processing-based algorithm by fixing the local cache hit ratio (c) and by varying the remote cache hit ratio ($c1$). In Fig. 1, we have fixed the c value as 0.3 and observed

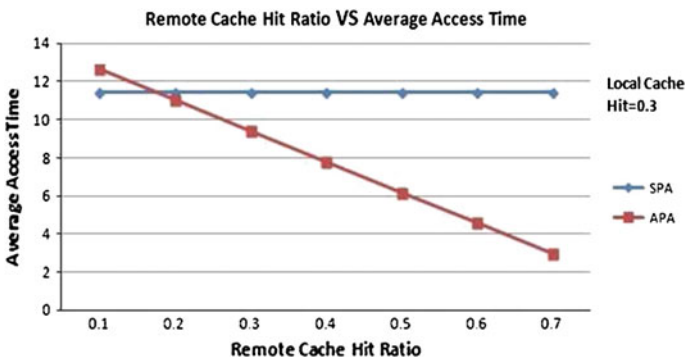


Fig. 1 Average read access time versus remote cache hit ratio

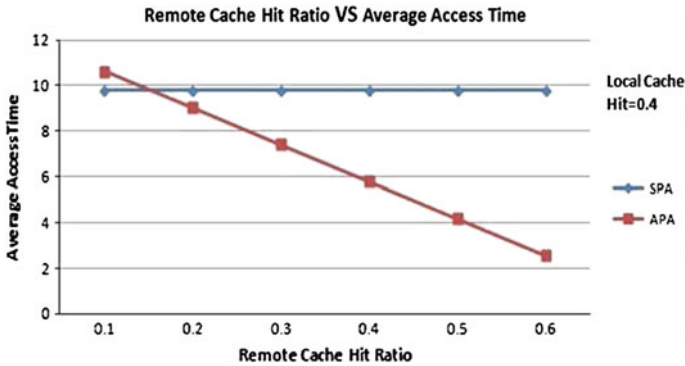


Fig. 2 Average read access time versus remote cache hit ratio

the performance of the algorithms. For the c_1 values 0.2 and above, the proposed anticipated parallel processing-based algorithm (APA) performs better than the speculation-based algorithm (SPA). Note that the performance of speculation-based algorithm is dependent only on c , whereas the performance of the proposed algorithm is dependent on both c and c_1 . We can also observe the similar trend in Fig. 2, where we have fixed the value of c as 0.4 and varied the values of c_1 from 0.1 to 0.6.

4.3 Simulation Experiments

We have developed a simulator for simulating speculation-based and anticipatory parallel processing-based algorithms and conducted simulation experiments by fixing the number of files available in the DFS, the number of cache blocks maintained in the local cache, nearby cache, and by varying the number of blocks available in the file.

Figure 3 shows the performance of the proposed algorithm (APA) and the speculation-based algorithm proposed in the literature (SP). Here, we have fixed the number of files present in the DFS as 50 and capacity of LC and NC as 100 blocks. We have varied number of blocks present in the files from 25 to 100 and observed the performance. We can observe that APA requires less average read access time than SP for all the cases.

Figure 4 shows the performance of APA and SP algorithms. Here, we have fixed the number of files present in the DFS as 50 and capacity of LC and NC as 200 blocks. We have varied number of blocks present in the files from 25 to 100 and observed the performance. We can observe that APA requires less average read access time than SP for all the cases.

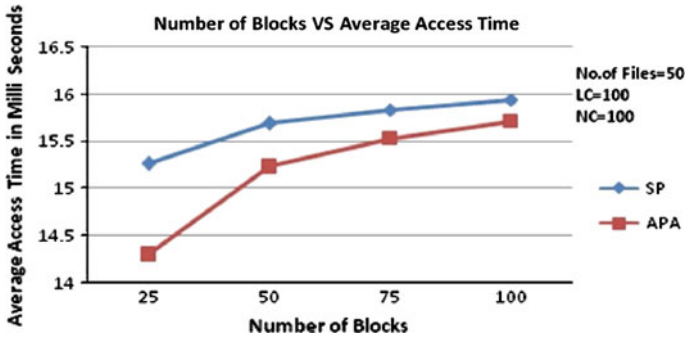


Fig. 3 Average read access time versus remote cache hit ratio

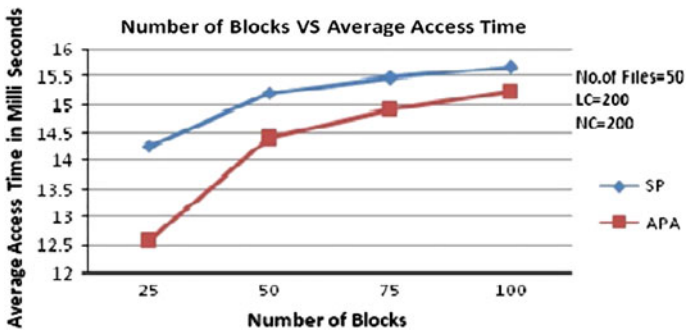


Fig. 4 Average read access time versus remote cache hit ratio

Overall, based on the results of our performance analysis, we conclude that the speculation-based algorithm performs better than the one without speculation, and the proposed anticipation-based algorithm can perform better than both speculation-based algorithm proposed in the literature and algorithm without speculation by considering the metric “Average Read Access Time.”

4.4 Discussion

See Tables 1, 2, 3, 4.

Table 1 Average access time when local cache hit = 0.3

S. no.	Remote cache hit ratio	Average access time (ms)	
		Speculation algorithm	Anti speculation algorithm
1	0.1	11.42	12.64
2	0.2	11.42	11.03
3	0.3	11.42	09.41
4	0.4	11.42	07.80
5	0.5	11.42	06.19
6	0.6	11.42	04.57
7	0.7	11.42	02.96

Table 2 Average access time when local cache hit = 0.4

S. no.	Remote cache hit ratio	Average access time (ms)	
		Speculation algorithm	Anti speculation algorithm
1	0.1	09.80	10.64
2	0.2	09.80	09.03
3	0.3	09.80	07.41
4	0.4	09.80	05.80
5	0.5	09.80	04.19
6	0.6	09.80	02.57

Table 3 Average access time when LC = 100 NC = 100

S. no	Number of blocks	Average access time (ms)	
		Speculation algorithm	Anti speculation algorithm
1	25	15.26	14.3
2	50	15.69	15.23
3	75	15.84	15.53
4	100	15.94	15.71

Table 4 Average access time when LC = 200 NC = 200

S. no.	Number of blocks	Average access time (ms)	
		Speculation algorithm	Anti speculation algorithm
1	25	14.26	12.58
2	50	15.22	14.42
3	75	15.50	14.92
4	100	15.67	15.24

5 Conclusion

Many IT organizations are deploying their applications in the cloud computing systems. Distributed file system is used as the main storage system in such cloud computing systems. Most of the cloud-based applications require a good performance from the distributed file system as for as file read operations are concerned. This is because most of the web applications deployed in the cloud computing systems perform file read operations more frequently, and less frequently these applications carry out file write operations. So improving the performance of file read operations in distributed file system is an important research issue.

In this paper, we have proposed a novel read algorithm for improving the performance of the DFS based on anticipated parallel processing. Also, we have carried out performance analysis for our algorithm and the speculation-based read algorithm proposed in the literature. The results of our analysis indicate that our proposed algorithm requires less read access time than the speculation-based read algorithm proposed in the literature.

References

1. Bernstein, D., Rodeh, M., Sagiv M.: Proving safety of speculative load instructions at compile-time. In: Krieg-Brackner B. (ed.) ESOP'92, volume 582 of Lecture Notes in Computer Science, pp. 56–72. Springer, Berlin (1992)
2. Dahlin, M.D., Wang, R.Y., Anderson, T.E., Patterson, D.A.: Cooperative caching: using remote client memory to improve file system performance. In: Proceedings of the 1st USENIX Conference on Operating Systems Design and Implementation, OSDI'94, USENIX Association, Berkeley (1994)
3. Jiang, S., Petrini, F., Ding, X., Zhang, X.: A locality-aware cooperative cache management protocol to improve network file system performance. In: 26th IEEE International Conference on Distributed Computing Systems, 2006. ICDCS 2006, pp. 42–42 (2006)
4. Liao, W.-K., Coloma, K., Choudhary, A., Ward, L., Russell, E., Tideman, S.: Collective caching: application-aware client-side file caching. In: Proceedings of the 14th IEEE International Symposium on High Performance Distributed Computing, 2005. HPDC-14, pp. 81–90. IEEE (2005)
5. Lilja, D., Bird, P.: The Interaction of Compilation Technology and Computer Architecture. Springer, US (1994)
6. Nightingale, E.B., Chen, P.M., Flinn J.: Speculative execution in a distributed file system. In: Proceedings of the Twentieth ACM Symposium on Operating Systems Principles, SOSP'05, pp. 191–205. ACM, New York (2005)
7. Patterson, R.H., Gibson, G.A., Ginting, E., Stodolsky, D., Zelenka, J.: Informed prefetching and caching. In: Proceedings of the Fifteenth ACM Symposium on Operating Systems Principles, SOSP'95, pp. 79–95. ACM, New York (1995)
8. Sarkar, P., Hartman, J.: Efficient cooperative caching using hints. In: Proceedings of the Second USENIX Symposium on Operating Systems Design and Implementation, OSDI'96, pp. 35–46. ACM, New York (1996)
9. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The hadoop distributed file system. In: 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), pp. 1–10. IEEE (2010)

Steerable Texture Descriptor for an Effective Content-Based Medical Image Retrieval System Using PCA

B. Jyothi, Y. MadhaveeLatha, P.G. Krishna Mohan and V.S.K. Reddy

Abstract Digital images have increased in quantity especially in the medical field used for diagnostics. Content-Based Medical Image Retrieval System will retrieve similar medical images from large database based on their visual features like texture, color, and shape. This paper focuses a novel method to increase the performance using Boundary detection, Steerable filter, and Principal Component Analysis. The content of the image was extracted with the help of region-based texture descriptor using steerable decomposition followed by extracting Principle Component Analysis which has better feature representation capabilities. The similar medical images are retrieved by comparing the extracted feature vector of the given query image with the corresponding database feature vectors using Euclidian distance as a similarity measure. The effectiveness of the proposed method is evaluated and exhibited via various types of medical images. With the experimental results, it is obvious that the region-based feature extraction method outperforms the direct feature extraction-based image retrieval system.

Keywords Content-based image retrieval · Boundary detection · Texture · Shape features · Principle component analysis · Similarity measure

B. Jyothi (✉) · Y. MadhaveeLatha · P.G. Krishna Mohan · V.S.K. Reddy
Department of Electronics and Communication Engineering,
Malla Reddy College of Engineering and Technology, JNTUH, Hyderabad, India
e-mail: b jyothi815@gmail.com

Y. MadhaveeLatha
e-mail: madhaveelatha2009@gmail.com

P.G. Krishna Mohan
e-mail: pgkmohan@yahoo.com

V.S.K. Reddy
e-mail: mrcet2004@gmail.com

1 Introduction

In the medical field, everyday huge volumes of different types of medical images are produced for diagnostics therapy and also to provide great support to the doctors in computer-aided diagnosing centers and research. Nowadays, different types of medical images are produced in various hospitals as well as in various medical diagnostic centers [1]. Content-based image retrieval has many momentous applications especially in medical diagnosis, education, and research fields. These images are also important because the historical images of different patients in medical centers have valuable information for the upcoming diagnosis with a system which retrieves similar cases that make more accurate diagnosis and decide on appropriate treatment [2]. So there is a necessity for efficient and effective retrieval system.

Content-based image retrieval system (CBIRS) is defined as a process of searching a digital image from the large database on the basis of their visual features like shape, color, and texture. Now the main intention of this research work is to retrieve the most similar medical images matching the given query image from large medical databases using feature extraction and similarity measurement techniques. However, due to the nature of medical, they are represented in gray level rather than color.

Medical images are enriched with specific textural patterns. These patterns have specific information. Low resolution, blurring effect, and contrast results are difficult to identify the tissue patterns and structure of organs, and these are some causes of reduction in the retrieval output.

Feature extraction plays a key role in improving the performance of the medical image retrieval system. In this paper, we are concentrating on the development of a novel and effective approach for extracting content information about region. Medical images have poor boundaries. Hence, there are a few practical problems while acquiring the medical images during scanning such less illumination and noise. A boundary is a line that marks the limits or edges of an entity and separates it from the background. It helps us in object recognition and understanding the shape of an image. In the proposed method, we have extracted the boundary of an image followed by extracting texture feature [3]. Krit Somkath has introduced a technique to overcome the problem of boundary detection for ill-defined edges using edge following algorithm defined in [4] based on intensity and texture gradient features which works better in detecting boundaries, and hence we have adapted it. We extend the framework to extract the content by means of steerable filter at various orientations followed by extracting texture features using principle component analysis. Finally, similar medical images are retrieved by matching the region features with the database images.

The rest of this paper is structured as follows. Section 2 describes the CBMIR system. Section 3 explains boundary detection procedure. Section 4 focuses on features extraction algorithm. Section 5 explains the experimental results. Section 6 concludes the paper.

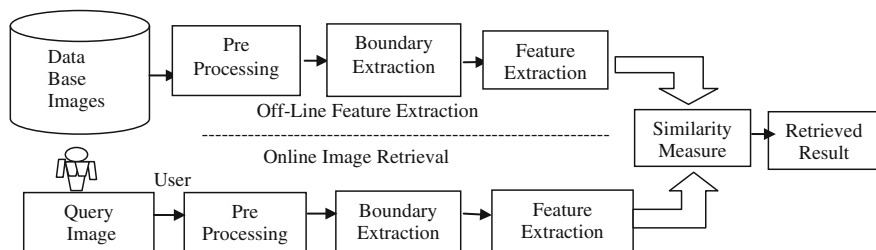


Fig. 1 Content-based medical image retrieval system

2 Proposed CBMIR System Architecture

The block diagram of a typical conceptual content-based retrieval system is illustrated in Fig. 1. It consists of two phases: offline feature extraction phase and online image retrieval phase. In offline phase, the images stored in the database are pre-processed by image de-noising. Next, the boundary of the images is detected followed by extracting the visual content of the image. The texture features of the database images are extracted and described with feature vectors and stored in the database. The database consists of various classes of medical images characterized by certain objects such as liver, body outline, and spine for CT or MRI images. The same process is repeated for query image in online phase.

In online image retrieval process, the user submits query images as an example to the retrieval system for searching similar images. The system retrieves related images by computing the similarity matching between the feature vectors of the query image and those of the database images. Finally, the system returns the results which are most relevant to the query image.

3 Boundary Detection Approach

Medical images consist of many objects which can be identified with the help of proper boundaries. In this approach, the boundary of the given query image and database images are detected using edge following algorithm which provides complete information of an image by considering both magnitudes and directions [5].

3.1 Gradient Vector Field Model

The magnitude and directions of given image $f(x, y)$ are calculated according to the following equations in x and y directions, respectively:

$$H(i,j) = \frac{1}{H_r} \sum_{(i,j) \in N} \sqrt{H_x^2(i,j) + H_y^2(i,j)} \tag{1}$$

$$A(i,j) = \frac{1}{H_r} \sum_{(i,j) \in N} \tan^{-1} \left(\frac{H_y(i,j)}{H_x(i,j)} \right) \tag{2}$$

$$H_x(i,j) = -G_y x f(x,j) \approx \frac{\partial f(x,y)}{\partial y} \tag{3}$$

$$H_y(i,j) = -G_x x f(x,j) \approx \frac{\partial f(x,y)}{\partial x} \tag{4}$$

where G_x and G_y are the Gaussian weight difference mask in the x and y directions and H_r is the total number of pixels in the 3×3 window neighborhood N . The idea of boundary extraction in unclear images is shown in Fig. 2.

3.2 Edge Following Technique

Applying from Law’s texture followed by double gradient detects, edges of objects in an image are derived from Law’s texture followed by canny edge detection. The texture feature of images is computed by convolving an input image $F(x, y)$ with

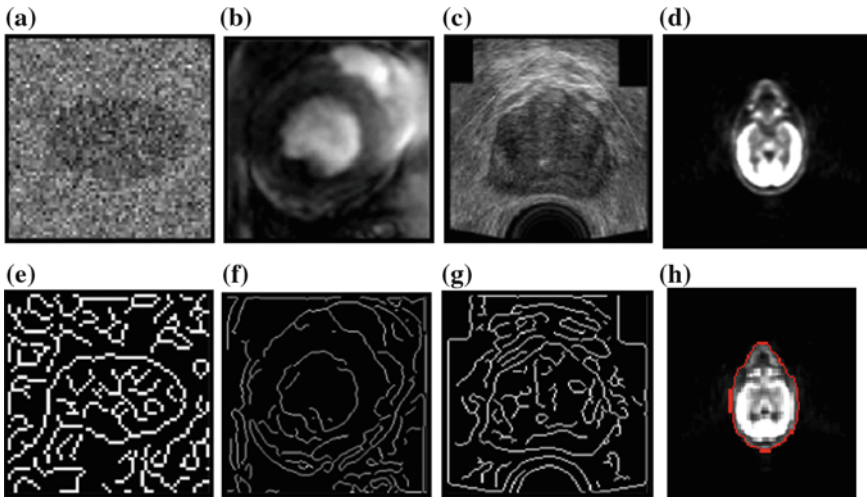


Fig. 2 **a** Synthetic noisy image. **b** Left ventricle MR image. **c** Ultrasound image. **d** Brain image. **e-f** Corresponding boundary detection with the help of edge maps derived from Law’s texture

each of the texture mask $T(x, y)$ which is defined as law's texture [6] and canny detection on the resultant image $R(x, y)$:

$$T(x, y) = L.L^T \quad (5)$$

$$L = (1, 4, 6, 4, 1)^T \quad (6)$$

$$R(x, y) = F(x, y) * T(x, y) \quad (7)$$

The average gradient vector field and edge map at the position (i, j) of an image are exploited to find the boundary of an image:

$$L_{ij}(r, c) = \alpha H_{ij}(r, c) + \beta A_{ij}(r, c) + \varepsilon R_{ij}(r, c) \quad (8)$$

$$0 \leq r \leq 2, 0 \leq c \leq 2$$

$$H_{ij}(r, c) = \frac{H(i+r-1, j+c-1)}{\text{Max}_{i,j} H_{i,j}};$$

$$A_{ij}(r, c) = 1 - \frac{|A(i, j) - A(i+r-1, j+c-1)|}{\pi};$$

$$R_{ij}(r, c) = R(i+r-1, j+c-1), 0 \leq r \leq 2, 0 \leq c \leq 2$$

where α , β , and ε are the weights; $H(i, j)$ and $A(i, j)$ are the magnitude and average of gradient edge vector fields; and $R(i, j)$ is the edge map from Law's texture.

4 Features Extraction

A feature is defined to express certain visual properties of an image, either locally for a small group of pixels or globally for the entire image. The contents of an image were described with the help of their features. In the proposed system, the local feature extraction is used in spite of extracting whole image. The input image is $F(x, y)$, where x and y are the pixel coordinates in the image. The image boundary is detected using edge following method, and corresponding region-based texture features are extracted in different orientations followed by extracting principal component analysis. In this section, we will discuss statistical region-based feature extraction in detail.

4.1 Texture Features Extraction

Medical images are mostly represented in gray level, and most medical image surfaces exhibit texture. Texture is a natural surface property, and it has repeated

pixel information regarding the structural arrangement and it also gives the relationship between the surface and external environment. In this paper, we implemented a rotation-invariant and scale-invariant texture representation for image retrieval based on steerable filter decomposition defined in [7].

Steerable filter is a special class of filter which is synthesized at different orientations as a linear combination of basis filters. It allows one to adaptively steer a filter to any orientation and to determine the filter output as a function of orientation.

The steering constraint is

$$F_{\theta}(m, n) = \sum_{k=1}^N b_k(\theta)A_k(m, n) \tag{9}$$

where $b_k(\theta)$ is the interpolation function based on the arbitrary orientation θ and $A_k(m, n)$ are the basis filter rotated version of impulse response at θ . An image at an arbitrary orientation was synthesized by taking linear combination of the image filtered convolved with the basis filters.

The steerable filter structure is illustrated in Fig. 3.

Texture information can be extracted by applying steerable filter in various oriented sub-bands. In this paper, we extract the texture features from six oriented sub-bands (0° , 60° , 120° , 180° , 240° , 300°), as shown in Fig. 4.

$$B_i(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1)P_i(x - x_1, y - y_1) \tag{10}$$

where $B_i(x, y)$ and P_i denote the horizontal and directional band pass filters at orientation ($i = 1, 2, 3, 4, 5, 6$).

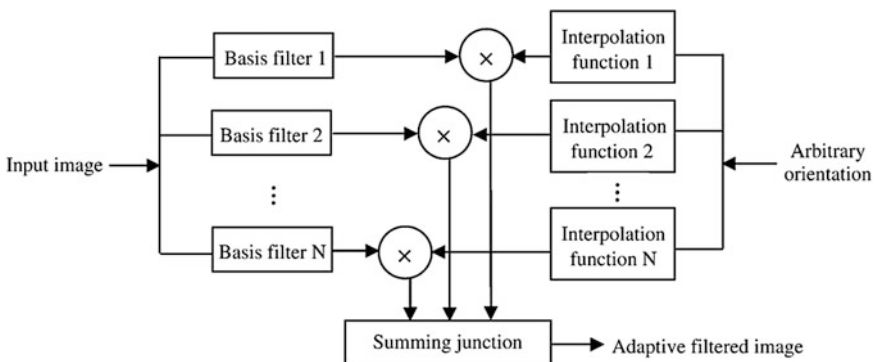


Fig. 3 Structure of steerable filter

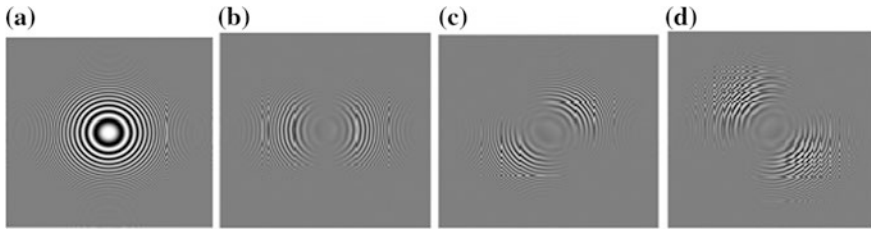


Fig. 4 The steerable filter image at different orientations. **a** The original image; **b** the image in horizontal orientation; **c** the image for rotation of 45°; **d** the image for rotation of -45°

4.2 Texture Representation Using PCA

The principal component analysis (PCA) is one of the most unbeaten techniques that have been used in image recognition applications. PCA is a statistical linear transformation method, used for the purpose of large dimensionality reduction of the data space.

PCA is linear transformation technique which is a variable reduction method and generally useful when data have high redundancy. This will result into reduction of information into smaller size which is called principal components. PCA is applied on filtered output at different orientations to reduce the dimensionality of a huge dataset [8]

The following steps give the detailed description of the mathematical calculation of the principal component analysis:

- (1) A 2-D image pixels can be represented as 1-D vector by re-arranging each column or row into a single vector and p_j 's represent the pixel values:

$$X_i = [A_1 \dots \dots \dots A_n], i = 1 \dots M. \tag{11}$$

- (2) The mean image was subtracted from each image vector to center the images. Let C represents the mean image:

$$C = \frac{1}{M} \sum_{k=1}^N X_i \tag{12}$$

The mean centered image is defined as

$$W_i = X_i - C \tag{13}$$

- (3) Covariance matrix of the dimensions $m \times m$ for mean centered images is calculated:

$$D = WW^T \quad (14)$$

(4) The eigen values were computed:

$$\text{Det}(D - \lambda I) = 0 \quad (15)$$

The nonzero eigenvalues was sorted from high to low where most image data can be represented.

Algorithm for Content-Based Image Retrieval Scheme

- Step 1 Input query medical image.
- Step 2 Convert image from RGB mode to Gray.
- Step 3 Preprocess the image with the help of median filter that suppresses undesired distortions and enhances image features which are very important for further processing
- Step 4 Extract the boundary of an image with the help of gradient vector image model and edge map techniques.
- Step 5 Apply steerable filter at six orientations and principal component analysis for extracting texture features.
- Step 6 For retrieving the similar images, these features were compared with the corresponding features in database.
- Step 7 Rank the images according to the sorted virtue distances and display the top 50 images.

5 Experimental Results and Discussion

The proposed method with the input dataset consists of the various images of different modalities. These images were acquired from Frederick National Laboratory for Cancer Research, which is open to all to provide an essential medical imaging research resource for supporting collection of data for private and internal projects and publishing results by project investigators. Our database consists of 1000 medical images of various classes of X-Ray, CT, and MRI scanned images such as chest, lungs, brain, abdomen, etc. We have tested the performance and effectiveness of the proposed approach. In the preprocessing stage, the noise in the images was corrected using median filter followed by segmenting the image for boundary extraction followed by extracting multiple features. The same process was repeated with the database images and query image. Similar medical images have been retrieved by computing the Euclidian distance [9] between the given query image feature vector and corresponding feature of the database images which is defined as

$$D(p_1, p_2) = \sqrt{\sum_{\forall i} (p_1(i) - p_2(i))^2} \tag{16}$$

The overall performance of a proposed CBMIR system is evaluated by measuring recall rate, mean average precision (MAP), and Error Rate, which have been employed in [10]. The performance of the proposed medical image retrieval system is compared with other systems which are based on various individual features and without extracting boundary of an image. From the experimental results as shown in the Table 1, it is clear that the proposed approach is more prominent for the content-based medical image retrieval as it compares with the existing system. This high performance is due to the combination of the shapes by extracting boundary of an image followed by extracting texture features.

$$\text{Recall Rate} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Relevant Images in Database}} \tag{17}$$

$$\text{Mean Average Precision} = \frac{\sum P(\text{NR})}{N} \tag{18}$$

where P (NR) is the precision after first NR images are retrieved.

$$\text{Error Rate} = \frac{\text{Number of Non Relevant Images Retrieved}}{\text{Total Number of Images Retrieved}} \tag{19}$$

The retrieval performance of the proposed CBMIR method was compared with existing feature extraction methods such as GLCM, LLBP, Tamura, and Gobor Features discussed in [11, 12]. Compared to all the above-mentioned techniques, the proposed feature extraction method gives better retrieval performance which works effectively for computer-aided diagnosis retrieval applications which are very effective for various image databases.

Table 1 Comparison of the recall rate, error rate, and mean average precision measures for various methods

Feature extraction technique	Performance measure using without boundary extracted features			Performance measure using with boundary extracted features		
	Recall rate (%)	Error rate (%)	MAP (%)	Recall rate (%)	Error rate (%)	MAP (%)
1. GLCM	42.12	58.88	56.11	43.58	57.42	58.56
2. Tamura	41.23	59.77	57.81	43.78	57.22	59.23
4. LLBP features	43.65	57.35	64.93	50.34	50.66	72.67
5. Gobor features	68.43	32.57	73.45	73.36	26.64	76.36
9. Proposed method Steerable + PCA	73.22	26.78	78.47	76.54	23.46	81.58

6 Discussion and Conclusion

Feature extraction is a significant task in efficient medical image retrieval. We have proposed a novel approach for extracting region features from the object of an image. It can easily be implemented and is computationally more effective than the traditional methods. In this paper, we have projected a new algorithm in which boundary of an image can be detected with the help of edge following algorithm followed by extracting region-based texture features such features with the help of steerable filter followed by extracting principal component analysis. The results have shown that the proposed approach improves the retrieval performance compared to the existing system. In future, the proposed scheme can further be improved by incorporating multiple features and relevance feedback.

References

1. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *Med. Inform.* **1**, 73 (2004)
2. Khoo, L.A., Taylor, P., Given-Wilson, R.M.: Computer-aided detection in the United Kingdom national breast screening programme: prospective study. *Radiology* **237**, 444–449 (2005)
3. Chun, Y.D., Kim, N.C., Jang, I.H.: Content-based image retrieval using multiresolution color and texture features. *IEEE Trans. Multimedia* **10**(6), 1073–1084 (2008)
4. Somkantha, K., Theera-Umpon, N.: Boundary detection in medical images using edge following algorithm based on intensity gradient and texture gradient features. *Proc. IEEE Trans. Biomed. Eng.* **58**(3), 567–573 (2011)
5. Veeralakshmi, S., Sivagami, S.V., Devi, V.V., Udhaya, R.: Boundary exposure using intensity and texture gradient features. *IOSR J. Comput. Eng. (IOSRJCE)* ISSN: 2278-0661, **8**(1), 28–33 (Nov–Dec 2012). ISBN: 2278-8727, www.iosrjournals.org
6. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny like criteria. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(8), 1007–1019 (2004)
7. Wang, X.-Y., Yu, Y.-J., Yang, H.-Y.: An effective image retrieval scheme using color, texture and shape features. *Comput. Stan. Interfaces CSI-02706* **33**(1), 59–68 (2011)
8. Navaz1, A.S.S., Dhevi sri, T., Mazumder, P.: Face recognition using principal component analysis and neural networks. *Int. J. Comput. Netw. Wirel. Mobile Commun.* **3**(1), 245–256 (Mar 2013). ISSN: 2250-1568
9. El-Naga, I., Yang, Y., Galatsanos, N.P., Nishikawa, R.M., Wernick, M.N.: A similarity learning approach to content-based image retrieval: application to digital mammography. *IEEE Trans. Medical Imaging* **23**(10), 1233–1244 (2004)
10. Arealillo-Herráez, M., Domingo, J., Ferri, F.J.: Combining similarity measures in content-based image retrieval. *Pattern Recognit. Lett.* **29**, 2174–2181 (2008)
11. Rasli, R.M, Muda, T.Z.T., Yusof, Y.: Comparative analysis of content based image retrieval techniques using color histogram: a case study of GLCM and K-means clustering. In: 3rd International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 283–286 (2012). ISBN: 978-1-4673-0886-1
12. Dubey, R.S., Choubey, R., Bhattacharjee, J.: Multi feature content based image retrieval. *Int. J. Comput. Sci. Eng.* **02**(06) (2010). ISSN : 0975-3397 2145 2145-2149

A Novel Multi-view Similarity for Clustering Spatio-Temporal Data

Vijaya Bhaskar Velpula and M.H.M. Krishna Prasad

Abstract With the enhanced usage of sensors and GPS devices, obtaining spatial and spatio-temporal data has become easy and analyses of these data in real-time applications are increasing day to day. Clustering is a data mining technique used for analyzing and obtaining unknown/hidden knowledge from the data/objects. Distance-based methods are helpful for analyzing and grouping the objects. In general, based on the type of data, Euclidean or Cosine distance-based techniques are used for grouping the data. Traditional techniques are point-based techniques and are based on single-view point, which may not produce efficient information and cannot be utilized for analyzing spatio-temporal objects. Hence, this paper presents a novel multi-view similarity technique for clustering spatio-temporal objects. Authors demonstrated the effectiveness of the proposed technique by adopting DBSCAN and implementing JDK1.2 on benchmarked datasets with respect to FMI indicator.

Keywords Clustering · Euclidean distance · Multi-view similarity · Spatio-temporal data

1 Introduction

Due to the enhanced usage of GPS and other sensory devices, availability of spatial and spatio-temporal data is growing day to day, e.g., repositories of remote-sensing images, sequential climatic changes, and collections of trajectories using handheld/motor attached GPS devices. Analysis of these spatio-temporal data is

V.B. Velpula (✉)

Department of Computer Science and Engineering, GEC, Guntur, AP, India

e-mail: vbr.velpula@gmail.com

M.H.M. Krishna Prasad

Department of Computer Science and Engineering, UCEK, JNTUK, Kakinada, AP, India

e-mail: krishnaprasad.mhm@jntucek.ac.in

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_30

increasing in various applications viz., land usage, environment and climate changes, and monitoring and predicting spread of disease.

Data mining/knowledge discovery from databases is a concept used for analyzing and obtaining hidden and unknown patterns from large datasets. In recent years [1, 2], research on spatio-temporal data is increasing day to day. For example, analysis of spatio-temporal data especially moving objects helps the analyst to design, analyze, evaluate, and optimize navigation spaces. Analysis of moving data/trajectories has gained more emphasis in different fields such as analyses of virtual environments, traffic management, location services, hurricane predictions, etc. Along with this, by performing spatio-temporal object analysis, a number of interesting patterns are developed which are useful in real-world life, e.g., using sensory data (trajectories) to identify migration patterns of selected groups of animals, analyzing virtual environments, and storing surveillance monitoring system data (mining customer movements will help in the arrangement merchandise and improve return on investment).

In literature, various authors presented the usage of clustering techniques to identify novel patterns. Any clustering technique requires a similarity/dissimilarity metric for measuring similarity between trajectories. The purpose of the similarity measure is to compare two trajectories (i.e., vectors of points) and to compute a single number which evaluates their similarity. In other words, the objective is to determine what extent two variables co-vary, which is to say, the same values for the same cases have. Some authors proposed and some adopted partitioning methods to generate a hierarchical structure of the document corpus by iteratively partitioning a large cluster into smaller clusters.

The challenging problem is measuring the similarity between two spatio-temporal objects. A number of cluster-based methods [3, 4] are used for measuring similarity. For example, [5] demonstrated the usage of Euclidean distance as a measure between time series of equal length of similarity, and the proposal has been generalized in [6] for subsequence matching. Reference [7] proposed the usage of Hausdorff similarity to measure distance between trajectories. A partition-and-group algorithm was used to cluster similar trajectories by considering three measurements, i.e., perpendicular distance, parallel distance, and angular distance [8].

While the above models find the spatio-temporal object similarity based on geographic features, some recent approaches introduce the semantic tags to enhance the accuracy of the measurement. But all the existing techniques are single-view point techniques, which cannot support the analyst to obtain effective patterns.

Hence, in this paper, authors propose a novel approach to measure the similarity between spatio-temporal data using multi-view point similarity.

2 Related Work

The main goal of clustering is to support analyst by grouping data/objects into clusters such that the intra-cluster similarity as well as the inter-cluster dissimilarity is maximized. Reference [3] proposed a spectral clustering framework that achieves

the multi-view goal by co-regularizing the clustering hypotheses, and proposed two co-regularization schemes to accomplish the task.

In literature, several authors proposed many state-of-the-art clustering techniques that do not employ any specific form of measurement, for example, non-negative matrix factorization [9] and probabilistic model-based method [10]. Even though, in general, to achieve better results, some form(s) of measurement is needed to determine similarity or dissimilarity. In this paper, due to wide usage, authors adopted Euclidean distance measure.

Some researchers proposed graph partitioning methods with variety of strategies and criterion functions viz., Average Weight [11] and Normalized Cut [12], and all of these procedures are efficiently applied for document clustering using cosine similarity measure [13]. In addition to the cosine measure, the extended Jaccard coefficient can also be adopted to represent similarity between nearest objects.

In [14], an empirical study was conducted to compare a variety of criterion functions for clustering. [15] demonstrated CLUTO a popular graph-based clustering technique. Initially, this technique models the objects with a nearest-neighbor graph and later using min-cut algorithm splits the graph into clusters.

In [16], in the survey paper, authors analyzed the existing four similarity measures, viz., Euclidean, Pearson correlation, extended Jaccard, and cosine, and stated that cosine and extended Jaccard perform well on web documents, when applied with popular clustering techniques (SOM, random, etc.). In nearest-neighbor graph clustering methods, such as the CLUTOs, the concept of similarity is somewhat different from the previously discussed methods. Initially, authors [17] evaluated the multi-view similarity measure for clustering a document corpus.

Recently, [18] demonstrated a method to compute dissimilarity between two categorical values of the sample based on their relationship with all other attributes. In continuation, [19] proposed a context-based distance learning technique for categorical data.

Euclidean measure is the most popular similarity/dissimilarity measure used to compute distance between a pair of continuous attributes and due to its easy computation and interpretation. By adopting the Euclidean measure, in the following section, authors describe the way to find the similarity between two spatio-temporal objects, and subsequently clustered using density-based clustering technique, dbSCAN.

3 Multi-view Similarity Measure

A spatio-temporal object S follows through a media as a function of time. Mathematically, which is defined as a sequence of movements, $S = [(t1, s1), \dots, (tn, sn)]$, this shows the consecutive positions of the object over a period of time. Here, n is defined as the length of S . Suppose, if two objects are similar, they should be close enough to each other in the problem space, and further both should have similar direction movement. But the challenge is how to measure closeness?

(i.e., similarity). Based on this, trajectories can only be considered as two (x - y plane)- or three (x - y - z plane)-dimensional time series data.

Considerable research has been done on one-dimensional data to measure similarity measure. Unfortunately, the similarity measure functions and indexing methods proposed for one dimension cannot be applied directly to the moving objects due to their characteristics [4]:

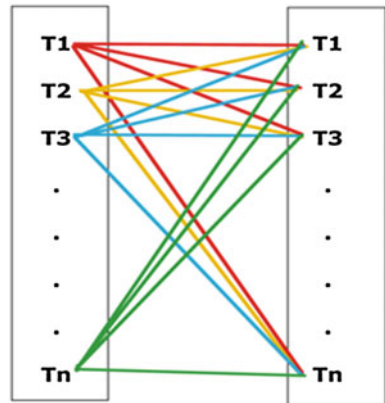
- In general, spatio-temporal objects are two- or three-dimensional data sequences with different lengths. Traditionally proposed, time-series-based similarity measures are focused on one-dimensional time series data only.
- Due to some failures, errors, and disturbances on data capture, many outliers may appear inside the data. This gap may lead to inaccuracy similarity measurement.

The purpose of a measure of similarity is to compare two objects and compute a quantitative value which represents their similarity. Euclidean distance is one of the regular metrics for geometrical problems. Most of the algorithms use Euclidean distance as a common distance measure between two points and can be easily measured in two-dimensional space and simply can be extended for three-dimensional space. The distance between two vectors X and Y is defined as follows:

$$d(x, y) = \sqrt{\sum_i^n (xi - yi)^2}. \tag{1}$$

Particularly, similarity of two spatio-temporal objects Si and Sj is represented as point vectors di and dj . $SpatDist(Si, Sj)$ is defined as the mean sum of the Euclidean distances between each point vector di and dj in the objects (Ref. Fig. 1), which can be adopted in single-view clustering techniques:

Fig. 1 Traditional measure (used in single-view clusterings)



$$SpatDist(S_i, S_j) = \frac{1}{\min(|S_i|, |S_j|)} \sum_{d_i \in T_i, d_j \in T_j} d(d_i, d_j). \tag{2}$$

Here, (in Eq. 2) S_i, S_j represents vector of points of the spatio-temporal object, d_i represents the vector point of the object i , and d_j represent the vector point of the object j . $|S_i|$ represents magnitude of the object S_i .

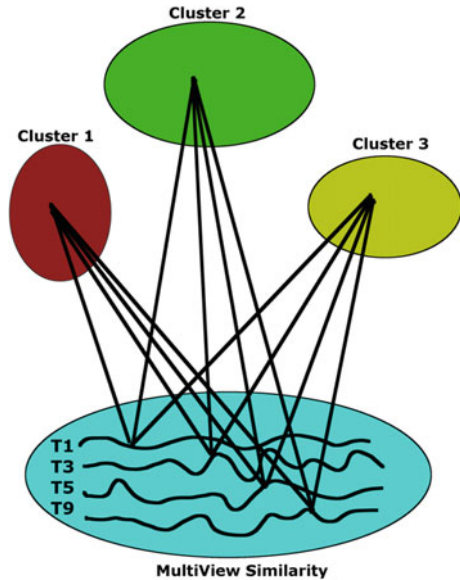
To construct a new concept of similarity, it is proposed to use more than one object of reference, i.e., from a third object S_h , the directions and distances to S_i and S_j are indicated, respectively, by the Euclidean measure.

Authors define similarity between the two spatio-temporal objects as below:

$$Sim_{i,tj \in S_r}(S_i, S_j) = \frac{1}{n - nr} \sum_{th \in S \setminus S_r} \sum_{ph \in S_h} Sim(pi - ph, pj - ph). \tag{3}$$

As described by Eq. (3), similarity of two objects S_i and S_j (if they are in the same cluster) is defined as the average similarity measured relatively from the views of all other spatio-temporal objects outside that cluster (Ref. Fig. 2). By this, one may have a more accurate judgment of how close or distant pair of trajectories are, by looking at them from many different viewpoints.

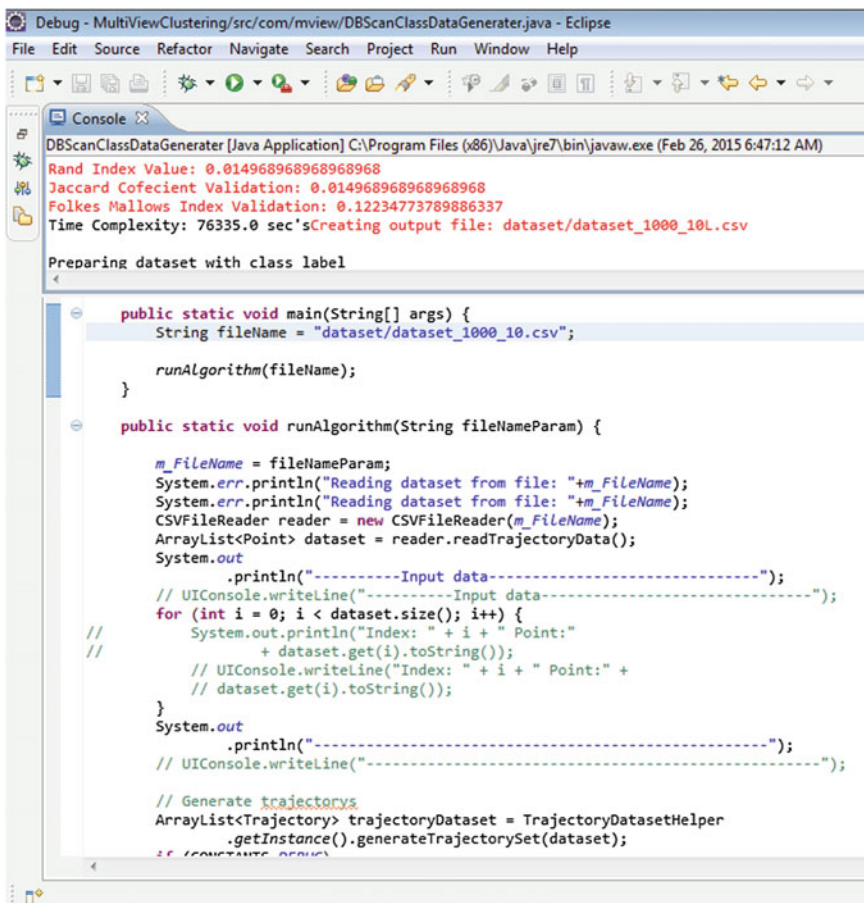
Fig. 2 Conceptual view of multi-view similarity measure



4 Experimental Observations

To demonstrate the effectiveness of the proposed technique, authors adopted DBSCAN with this proposed measure using JDK1.6 and experimented on synthetic and real-time datasets namely *Microsoft T-Drive* [20]. *Microsoft T-Drive* dataset contains the GPS trajectories with taxi id, time, and its position (in the form of longitude and latitude), obtained for every 5 s of time gap. During preprocessing, we extracted trajectories for every 5 min gap.

As the multi-view similarity can be implemented on labeled datasets, initially we clustered using standard DBSCAN (with Euclidean as dissimilarity measure) and labeled the trajectories to corresponding classes, i.e., cluster ids. Figure 3 shows the screenshot showing the obtained output and corresponding indicators (viz., Rand,



```

Debug - MultiViewClustering/src/com/mview/DBScanClassDataGenerator.java - Eclipse
File Edit Source Refactor Navigate Search Project Run Window Help

Console
DBScanClassDataGenerator [Java Application] C:\Program Files (x86)\Java\jre7\bin\javaw.exe (Feb 26, 2015 6:47:12 AM)
Rand Index Value: 0.014968968968968968
Jaccard Coeficient Validation: 0.014968968968968968
Folkes Mallows Index Validation: 0.12234773789886337
Time Complexity: 76335.0 sec'sCreating output file: dataset/dataset_1000_10L.csv

Preparing dataset with class label
4

public static void main(String[] args) {
    String fileName = "dataset/dataset_1000_10.csv";

    runAlgorithm(fileName);
}

public static void runAlgorithm(String fileNameParam) {
    m_FileName = fileNameParam;
    System.err.println("Reading dataset from file: "+m_FileName);
    System.err.println("Reading dataset from file: "+m_FileName);
    CSVFileReader reader = new CSVFileReader(m_FileName);
    ArrayList<Point> dataset = reader.readTrajectoryData();
    System.out
        .println("-----Input data-----");
    // UIConsole.writeLine("-----Input data-----");
    for (int i = 0; i < dataset.size(); i++) {
        System.out.println("Index: " + i + " Point:"
            + dataset.get(i).toString());
        // UIConsole.writeLine("Index: " + i + " Point:" +
        // dataset.get(i).toString());
    }
    System.out
        .println("-----");
    // UIConsole.writeLine("-----");

    // Generate trajectories
    ArrayList<Trajectory> trajectoryDataset = TrajectoryDatasetHelper
        .getInstance().generateTrajectorySet(dataset);
    // (CONSTANTS DEFINE)
}

```

Fig. 3 Screenshot showing the label generator for the trajectories using simple DBSCAN

Jaccard and Fowlkes-Mallows Index (FMI) [21]). FMI is a validation technique that is not dependent on the number of clusters.

Figure 3, captured screenshot, shows the obtained output after clustering 1000 trajectories, where one can observe that the rand index value is 0.0149, Jaccard Coefficient is 0.0149, and FMI is 0.12234.

Now, labeled trajectories are clustered using DBSCAN with proposed multi-view similarity measure and compared the obtained indices with indices of standard DBSCAN. Figure 4 shows the obtained screenshot after clustering the above labeled trajectories with this proposed measure, where one can observe that the rand index value is 0.968, Jaccard Coefficient is 0.968, and FMI is 0.98399; as FMI is giving more index (in Figs. 3 and 4), we considered FMI for other test cases and observed values are reported in Table 1.

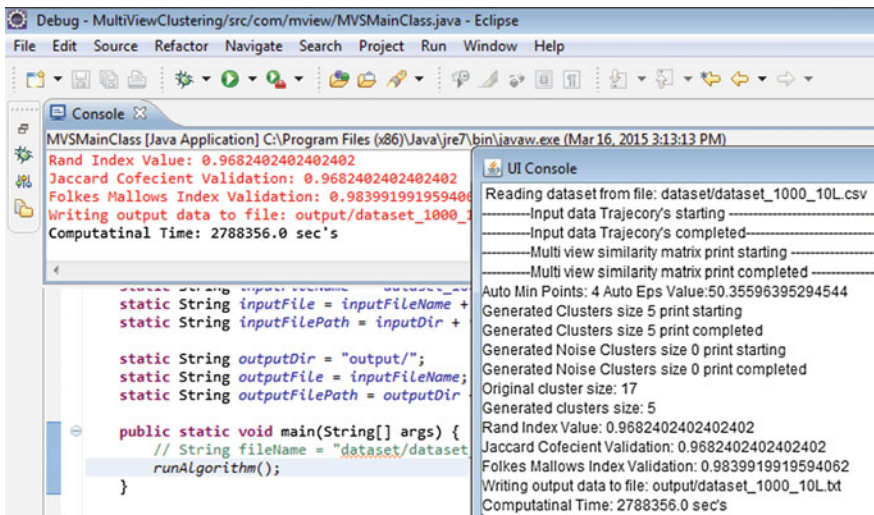


Fig. 4 Screenshot showing the output of the DBSCAN (using proposed measure)

Table 1 Validation results

Dataset size	FMI	Time (ms)	FMI	Time (s)
100	0.386384	272	0.979695	397
200	0.272436	773	0.998765	758
300	0.223033	1339	0.968745	1254
400	0.193924	2653	0.987484	22,116
500	0.173124	4696	0.975976	76,825
600	0.157841	7373	0.978315	118,432
700	0.146224	11,287	0.984274	134,192
800	0.137028	13,742	0.745154	193,258
900	0.129066	18,957	0.845154	214,192

From Table 1, one can easily note that the proposed technique is consistently performing well but it consumes more computation power. Moreover, particularly for the dataset size of 800 and 900, the FMI value is considerably dropped and, as per our observation, it is due to the presence of outliers.

5 Conclusion and Future Scope

This paper presents a novel multi-view similarity technique for measuring similarity between spatio-temporal objects, i.e., trajectories. We adopted the simple DBSCAN and clustered the trajectories with the proposed measure on Synthetic and *Microsoft T-Drive* data. From our experimental observations, it is clear that the proposed method produces better clusters, and same is observed in the form of FMI values and consumes more computational power.

In future, authors want to experiment on complex and real-time data. And also, as the computational time requires more in the proposed technique, we want to utilize multi-core programming and planning to develop a visualization tool justify the proposal.

References

1. Lei, C., Tamer, M.: Robust and fast similarity search for moving object trajectories. In: ACM SIGMOD Conference, pp. 491–502. ACM, New York (2005)
2. Nesrine, A., Lorna, S.: Graph Clustering: Complexity, Sequential and Parallel Algorithms. University of Alberta, Edmonton (1995)
3. Abhishek, K., Piyush, R., Hal, D.: Co-regularized multi-view spectral clustering. In: Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems, pp. 1413–1421. Spain (2011)
4. Jae-Gil, L., Han, J., Whang, K.: Trajectory clustering: a partition-and-group framework. In: ACM SIGMOD, pp. 593–604 (2007)
5. Agrawal, R., Faloutsos, C., Swami, A.: Efficient similarity search in sequence databases. In: Fourth International Conference on Foundations of Data Organization and Algorithms, pp. 69–84. London (1993)
6. Faloutsos, C., Ranganathan, M., Manolopoulos, Y.: Fast Subsequence matching in time-series databases. In: ACM SIGMOD International Conference on Management of Data, pp. 419–429. Minnesota (1994)
7. Hazarath, M., Lucio, I., Luca, C.: CAST: a novel trajectory clustering and visualization tool for spatio-temporal data. In: First International Conference on Intelligent Human Computer Interaction, pp.169–175. India (2009)
8. Lee, J., Han, J., Whang, K.: Trajectory clustering: a partition-and-group framework. In: ACM SIGMOD International Conference on Management of Data, pp. 593–604. Beijing (2007)
9. Wei, X., Xin, L., Yihong, G.: Document clustering based on non-negative matrix factorization. Paper presented at the meeting of the SIGIR (2003)
10. Arindam, B., Inderjit, S.D., Joydeep, G., Sra, S.: Clustering on the unit hypersphere using von mises-fisher distributions. *J. Mach. Learn. Res.* **6**, 1345–1382 (2005)

11. Hongyuan, Z., Xiaofeng, H., Chris. H.Q.D., Ming, G.: Spectral relaxation for K-means clustering. In: *Advances in Neural Information Processing Systems* **14**, pp. 1057–1064 (2001)
12. Jianbo, S., Jitendra, M.: Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000)
13. Dhillon, I.S.: Co-clustering documents and words using bipartite spectral graph partitioning *KDD*, pp. 269–274. *ACM* (2001)
14. Zhao, Y., Karypis, G.: Empirical and theoretical comparisons of selected criterion functions for document clustering. *Mach. Learn.* **55**(3), 311–331 (2004)
15. Karypis, G.: CLUTO a Clustering Toolkit. Technical report, Department of Computer Science, University of Minnesota. <http://glaros.dtc.umn.edu/~gkhome/views/cluto> (2003)
16. Strehl, A., Ghosh, J., Mooney, R.: Impact of similarity measures on web-page clustering. In: *17th National Conference Artificial Intelligence: Workshop of Artificial Intelligence for Web Search (AAAI)*, pp. 58–64. July (2000)
17. Prasad, V.V.D., Krishna Prasad, M.H.M., Velpula, V.B.: Evaluation of multi-view point similarity based document clustering. In: *Third International Conference on Recent Trends in Engineering and Technology*, pp. 132–135. India (2014)
18. Ahmad, A., Dey, L.: A method to compute distance between two categorical values of same attribute in unsupervised learning for categorical data set. *Pattern Recogn. Lett.* **28**(1), 110–118 (2007)
19. Ienco, D., Pensa, R.G., Meo, R.: Context-based distance learning for categorical data clustering. In: *Eighth International Symposium. Intelligent Data Analysis (IDA)*. pp. 83–94 (2009)
20. T-Drive trajectory’s dataset downloaded on 8 Sep 2013. <http://research.microsoft.com/apps/pubs/?id=152883>
21. Fowlkes, E.B., Mallows, C.L.: A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* **78**, 383 (1983)

WRSP-Miner Algorithm for Mining Weighted Sequential Patterns from Spatio-temporal Databases

Gurram Sunitha and A. Rama Mohan Reddy

Abstract Not allowing priorities in the mining process does not support user-directed or focus-driven mining. The work proposed in this paper provides support to include user prioritizations in the form of weights into the mining process. An algorithm *WRSP-Miner* is proposed for the purpose of mining *Weighted Regional Sequential Patterns* (WRSPs) from spatio-temporal event databases. *WRSP-Miner* uses two interestingness measures *sequence weight* and *significance index* for efficient mining of WRSPs. Experimentation has been performed on synthetic datasets and results proved that the proposed WRSP-Miner algorithm has achieved the purpose of its design.

Keywords Data mining · Spatiotemporal database · Event · Sequential pattern · Weighted patterns

1 Introduction

Ultimately, the goal of data mining process is to discover knowledge from the databases. The knowledge thus gathered must be of interest to the user and help him to perform better decision making. Sequential patterns are mined from the spatio-temporal databases to discover temporal relationships between data. Let the spatio-temporal database be an event database consisting of event records that have physically occurred in the real world and of various event types. Users may not have the same concern towards all the event types and his interest may vary with time and with geographical regions as well. Not allowing priorities in the mining

G. Sunitha (✉) · A. Rama Mohan Reddy
Department of CSE, S. V. University College of Engineering,
S. V. University, Tirupati, Andhra Pradesh, India
e-mail: gurramsunitha@gmail.com

A. Rama Mohan Reddy
e-mail: ramamohansvu@yahoo.com

process does not allow user-directed or focus-driven mining. Also generally, the data in the database may not be evenly distributed with respect to the frequency, geographic regions, etc. This leads to problems in fixing the minimum threshold values, where if minimum threshold value is set too high, some important patterns with low support will be lost, and setting it too low produces many spurious patterns difficult to handle. Such problems can be handled by the use of prioritization. The concept of prioritization can be implemented by associating weights to the data and performing mining on weighted data to discover weighted patterns. Mining weighted sequential patterns is one of the ways of reducing or pruning the resultant sequential pattern set. The idea has been proven efficient in many real-life applications. Till date, no research has been made to perform weighted sequential pattern mining from spatiotemporal databases. In this view, an algorithm called as WRSP-Miner has been proposed in this paper and is discussed in detail. WRSP-Miner uses two interestingness measures *sequence weight* and *significance index* for efficient mining of WRSPs.

2 Related Work

The work on discovering temporal relationships (sequential patterns) from transactional databases has been initiated by Agrawal and Srikant [1] which has been extended to spatio-temporal databases by Tsoukatos and Gunopulos [2]. Tsoukatos et al. [2] have proposed DFS_MINE algorithm for extracting sequential patterns from environmental data. [3–7] have proposed different techniques to discover sequential patterns from spatio-temporal databases. [8–10] stand as the milestone researches in the problem of introducing weights into the mining process for discovering weighted association rules. Yun et al. have studied in-depth the problem of mining weighted association rules and have introduced many methods [11–13]. Yun has also proposed two algorithms WSpan [14] and WIS [15] to extend the weighted mining concept to the process of sequential patterns. Another method has been proposed by Chang [16] recently.

3 Problem Statement

Let $STDB$ be the spatio-temporal database which consists of a set of events $\{e_1, e_2, \dots, e_n\}$. Let the event types be $\{E_1, E_2, \dots, E_m\}$ where m represents the number of event types. Each of the event type E_i comprises of a set of events. Let the geographical space under observation R be divided into sub-regions $R = \{R_1, R_2, \dots, R_r\}$. Let the time be divided into time intervals $T = \{T_1, T_2, \dots, T_t\}$. The spatio-temporal space covered by a geographical sub-region R_{ij} and a time interval T_k forms a spatiotemporal cell denoted as G_{ijk} . Let $W_E = \{w_1, w_2, \dots, w_m\}$ be the weights assigned to the event types and $W_R = \{w_1, w_2, \dots, w_w\}$ be the weights

assigned to the spatiotemporal cells. The weights assigned to the event types as well as to the spatio-temporal cells are non-negative real numbers that reflect their importance in the spatio-temporal event database. The sum of weights of all the event types should be equal to 1. Also, the sum of weights of all the spatio-temporal cells must be equal to 1. The weights are pre-assigned by the domain expert according to the relative importance of the event types, geographical regions, and time intervals.

A WRSP is a sequence of event types where the event types in the pattern have significant weight and the patterns are mined considering events that have occurred in the spatio-temporal cells with significant weight. The discovery process considers the significance of the event types, geographical regions, and time intervals. Such discovered patterns satisfy the significance and weight constraints. Hence, *weighted regional sequential pattern mining* can be defined as the problem of finding WRSPs that satisfy the significance constraint and weight constraint from the given spatio-temporal event database.

The work discussed in this paper is an extension to the work proposed in [17]. The RSP-Miner algorithm proposed in [17] assumes that all the event types, geographical regions, and time intervals are of equal importance, i.e., it assumes that the weight of all event types, geographical regions, and time intervals is 1. The concept of discovering weighted sequential patterns is based on assigning varying weights to different parameters to vary their importance and to efficiently simulate their priorities in the real world. Pushing weight constraints deep into the mining process is a non-trivial task. In aim of achieving such task, WRSP-Miner algorithm is proposed as an extension for the RSP-Miner algorithm proposed in [17]. The data model used for the purpose of mining WRSPs is the same as the data model presented in [17]. In this work, event types, geographical regions, and time intervals are considered to be the weighted attributes. The priorities of event types, geographical regions, and time intervals are domain-dependent and are user-defined.

4 Definitions

This section presents the necessary definitions for the purpose of mining weighted sequential patterns. The definitions defined in [17] are applicable for WRSP-Miner algorithm. The remaining definitions are as follows:

Definition 1 (*Weight of an event type*) It is the weight given to the event type to define its relative importance within the given event types. For a given event type E , it is denoted as $\text{weight}(E)$.

Definition 2 (*Weight of an event*) Consider an event e . Let the event type of e be E . Then the weight of the event e is the weight of its event type, i.e., $\text{weight}(e) = \text{weight}(E)$ where $e \in E$.

Definition 3 (*Weight of a spatiotemporal cell*) It is the weight given to a spatio-temporal cell (to a geographical region with respect to a time interval) to define its

relative importance within the given spatiotemporal cells. For a given spatiotemporal cell G_{ijk} , it is denoted as $\text{weight}(G_{ijk})$.

Definition 4 (Region Weight) Consider a spatiotemporal cell G_{ijk} . Its region weight $\text{Rweight}(G_{ijk})$ is defined as the product of the average weight of the events in G_{ijk} and the weight of G_{ijk} .

$$\text{R weight}(G_{ijk}) = \text{average}_{e \in G_{ijk}} (\text{weight}(e)) \times \text{weight}(G_{ijk}).$$

Definition 5 (Weight of a k-sequence) Consider a sequence $S = E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k$ of length k . Its weight is defined as the average weight of the event types in the sequence.

$$\text{Weight}(S) = \text{average}_{E_i \in S} (\text{weight}(E_i)).$$

Definition 6 (Sequence Weight) Consider a sequence $S = E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k$ of length k . Let the set of precursor nodes of the sequence S be P_N . Let the set of spatiotemporal cells existing be G . The sequence weight of S is defined as the ratio of product of weight of S and the sum of weights of precursor nodes to the sum of region weights of precursor nodes.

$$\text{Sweight}(S) = \frac{\text{weight}(S) \times \sum_{G_{ijk} \in P_N} \text{weight}(G_{ijk})}{\sum_{G_{ijk} \in G} \text{Rweight}(G_{ijk})}.$$

Definition 7 (Weighted Regional Sequential Pattern) A k -sequence of form $E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k$ is said to be a *Weighted Regional Sequential Pattern* if and only if it satisfies both weight constraint and significance constraint, i.e., if and only if

- (a) $\text{Sweight}(E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k) \geq \text{Min_SW}$ and
- (b) $\text{SI}(E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k) \geq \text{Min_SI}$,

where Min_SW is the minimum sequence weight threshold and Min_SI is the minimum significance threshold.

Definition 8 (k -Weighted Regional Sequential Pattern) A k -weighted regional sequential pattern is a weighted Regional Sequential Pattern with k event types in the pattern.

Definition 9 (Weighted Sequential Pattern Mining) Given a discretized spatio-temporal event database STDB, weight vector of event types W_E , weight vector of spatio-temporal cells W_R , maximum length of WRSPs to be generated maxLen , minimum sequence weight threshold Min_SW , and minimum significance index threshold Min_SI , the weighted regional sequential pattern mining is defined as the process of extracting all WRSPs of length up to maxLen .

5 WRSP-Miner Algorithm for Mining WRSPs

The working procedure is same as that of RSP-Miner [17] except that instead of prevalence constraint, WRSP-Miner works using weight constraint. The WRSP-Miner algorithm is shown in Fig. 1. Let $\rho \rightarrow f$ be a new sequence formed by appending an event type to the end of the already proven significant pattern. Then for the sequence, the sequence weight newSW and significance index newSI are calculated. The sequence $\rho \rightarrow f$ is considered as significant sequential pattern if

Algorithm WRSP-Miner

Input:

ρ – Sequential pattern to be expanded

L – Length of ρ

P – Set of Follower Nodes of ρ (Precursor Nodes of newSeq)

SI – Significance Index of ρ

Output :

R_{SEQ} – Set of Weighted Regional Sequential Patterns

Variables :

$maxLen$ – Maximum length of sequences to be generated

$newSeq$ – New sequence generated

N – Set of Follower Nodes of newSeq

FDR – Follower Density Ratio of newSeq

$newSI$ – Significance Index of newSeq

$newSW$ – Sequence Weight of newSeq

Min_SW – Minimum Sequence Weight Threshold

Min_SI – Minimum Significance Index Threshold

```

1:  for each event type  $f$  in STDB do
2:       $L \leftarrow L+1$ 
3:       $newSeq \leftarrow (\rho \rightarrow f)$ 
4:       $FDR \leftarrow findFDR (newSeq, P)$ 
5:      if  $L = 2$  then
6:           $newSI \leftarrow FDR$ 
7:      else
8:           $newSI \leftarrow MIN(SI, FDR)$ 
9:      end if
10:      $newSW \leftarrow findSW (newSeq, P)$ 
11:     if  $newSI \geq Min\_SI$  then
12:         if  $newSW \geq MinSW$  then
13:              $R_{SEQ}$ .add ( $newSeq$ )
14:         end if
15:         if  $L < maxLen$  then
16:              $N \leftarrow findFollowerNodes(newSeq, P)$ 
17:             call WRSP-Miner( $newSeq, L, N, newSI$ )
18:         end if
19:     end if
20: end for

```

Fig. 1 WRSP-Miner algorithm for mining WRSPs

$newSI \geq Min_SI$ and is considered to be a potential candidate for further extension to higher lengths. But it is added to the resultant pattern set only if $newSW \geq Min_SW$. Min_SW and Min_SI are defined by the user. Min_SW value should be between 0 and 1. A sequence weight value 1 of a sequence S signifies the fact that it contains most significant event types and the sequence is repeating itself in most significant spatio-temporal cells. Considering both sequence weight and significance index measures at a time makes sure that patterns having higher importance and worth considering are only presented to the user. Also, pruning the final pattern set reduces the set to a more manageable size.

6 Experimental Evaluation

For evaluating the performance of the proposed WRSP-Miner algorithm, a comparison is made with the RSP-Miner algorithm [17]. For the experimental evaluation of the WRSP-Miner algorithm, synthetic dataset *syn1* is used. The complete description of the same is provided in [17]. For the purpose of discretizing the space and time, an uniform rectangular grid of size (5, 6.5, 6.5) is considered. The Min_PI and α values for the WRSP-Miner are set to 0, and minimal spatio-temporal cell pruning is used. The Min_SI is considered to be 1. The weight vector of event types is adjusted so that half of the elements are of high importance and half of them are of least importance. It is assumed that all the geographical regions and time intervals have equal importance and hence their weight vector is a unit vector.

When both the algorithms are extensively studied using *syn1* dataset by varying the input parameters, it is observed that both the algorithms have the same time complexity, but the number of patterns discovered varied. The number of patterns discovered by the algorithms with respect to the varying *maxLen* parameter is shown in Fig. 2. Figure 2a shows the number of patterns discovered when Min_SW parameter value is taken as 0.5. Figure 2b shows the number of patterns discovered

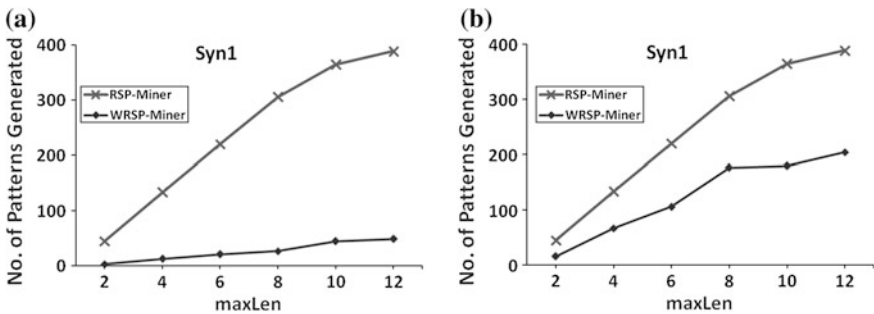


Fig. 2 Number of patterns discovered from *syn1* dataset

when Min_SW parameter value is taken as 0.8. It can be seen that the number of patterns discovered by WRSP-Miner algorithm was less when compared to the number of patterns discovered by RSP-Miner algorithm. This is because of inducing the weight constraint into the mining process. Hence, final pattern set presented to the user contained only most significant and highly weighted patterns.

It is observed that even though WRSP-Miner algorithm does not reduce the time complexity of the mining process, it efficiently manages to discover the interesting patterns and hence prune the final pattern set thus achieving the purpose of its design. This is particularly observed when Min_SI value is low. Adjusting weight vectors and integrating the results will result in identifying the consistently repeating patterns or top-k% patterns. In RSP-Miner approach, Min_SI value is the only factor which can be adjusted to adjust the final pattern set size. It is a proven fact that adjusting minimum threshold values is a too complex process and has been a major research area by itself. In this view, WRSP-Miner is designed to give lineage to the process of fixing the Min_SI threshold value. By fine adjustment of weight vectors, the number of patterns discovered by WRSP-Miner is remarkably less when compared to RSP-Miner to several orders or magnitudes. This can be achieved by the help of invaluable domain expertise. The achievability of the WRSP-Miner algorithm becomes bigger when Min_SI value is set too low. Hence, to prove the efficiency of the proposed algorithm, the Min_SI value is taken as 1, which is the possible minimum value in the experiments.

Table 1 shows the number of WRSPs extracted from *syn1* dataset with various Min_SI and Min_SW values. It is noticed that WRSP-Miner can generate fewer WRSPs using different minimum threshold values. The number of patterns discovered, dramatically decreases, with the increasing minimum threshold values.

For example, consider the scenario when Min_SI value is equal to 1. The number of patterns discovered by WRSP-Miner when Min_SW value is 0, 0.5, and 0.8 are 365, 180, and 45, respectively. The reduction in pattern set size is further achieved by increasing the Min_SI value. This shows the effectiveness of the weight constraint used in WRSP-Miner. In this way, the proper number of weighted sequential patterns can be found by adjusting the minimum sequence weight threshold and minimum significance index threshold. It is observed that the sequential patterns consisting of high number of insignificant event types are automatically eliminated by the use of weight constraint. This ensures that only quality patterns are included into the final sequential pattern set.

Table 1 Effectiveness of minimum weight in WRSP-Miner

Min_SI	Number of patterns discovered		
	Min_SW = 0	Min_SW = 0.5	Min_SW = 0.8
1 (minimum value)	365	180	45
10 % (of maximum Min_SI)	264	127	32
15 % (of maximum Min_SI)	169	77	24
20 % (of maximum Min_SI)	97	19	11

7 Conclusion

The contribution in this paper is focused on efficient pruning of the resultant pattern set, and hence supporting user-directed mining. An extension of RSP-Miner algorithm [17] is conceived for the purpose. The idea was to allow users induce their priorities into the mining process through setting the weights of the elements of the dataset. In this work, the proposed method supports users to set varying weights to event types, geographical regions, and time intervals to define their relative importance in the real-world. An algorithm called as WRSP-Miner is developed that uses *sequence weight* and *significance index* measures to refine the mining process. The experimental studies have proved that the proposed method is efficient in pruning the discovered patterns and has given lineage in fixing the minimum significance index threshold value.

References

1. Agrawal, R., Srikant, R.: Mining sequential patterns. In: Proceedings of 1995 International Conference on Data Engineering (1995)
2. Tsoukatos, I., Gunopulos, D.: Efficient mining of spatiotemporal patterns, pp. 425–442. Springer, Berlin, Heidelberg, (2001)
3. Wang, J., Hsu, W., Lee, M.L.: Flow miner: finding flow patterns in spatio-temporal databases. In: 16th IEEE International Conference on Tools with Artificial Intelligence, pp. 14–21 (2004)
4. Wang, J., Hsu, W., Lee, M.L.: Mining generalized spatio-temporal patterns. In: Database Systems for Advanced Applications, pp. 649–661. Springer, Berlin, Heidelberg (2005)
5. Huang, Y., Zhang, L., Zhang, P.: A framework for mining sequential patterns from spatio-temporal event data sets. IEEE Trans. Knowl. Data Eng. **20**(4), 433–448 (2008)
6. Salas, H.A., Bringay, S., Flouvat, F., Selmaoui-Folcher, N., Teisseire, M.: The pattern next door: towards spatio-sequential pattern discovery. In: Advances in Knowledge Discovery and Data Mining, pp. 157–168. Springer, Berlin, Heidelberg (2012)
7. Fabrègue, M., Braud, A., Bringay, S., Le Ber, F., Teisseire, M.: Including spatial relations and scales within sequential pattern extraction, pp. 209–223. Discovery Science, Springer, Berlin, Heidelberg (2012)
8. Cai, C.H., Chee Fu, A.W., Cheng, C.H., Kwong, W.W.: Mining association rules with weighted items. In: Proceedings of the Sixth International Conference on Intelligent Data Engineering and Automated Learning (1998)
9. Wang, W., Yang, J., Yu, P.S.: Efficient mining of weighted association rules (WAR). In: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 270–274 (2000)
10. Tao, F.: Weighted association rule mining using weighted support and significant framework. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 661–666 (2003)
11. Yun, U., Leggett, J.J.: WFIM: Weighted frequent itemset mining with a weight range and a minimum weight. In: Proceedings of the Fourth SIAM International Conference on Data Mining, pp. 636–640 (2005)
12. Yun, U., Leggett, J.J.: WLPMiner: weighted frequent pattern mining with length decreasing support Constraints. In: Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 555–567 (2005)

13. Yun, U., Leggett, J.J.: WIP: mining weighted interesting patterns with a strong weight and/or support affinity. *SDM* **6**, 3477–3499 (2006)
14. Yun, U., Legget, J.J.: WSpan: weighted sequential pattern mining in large sequence databases. In: 3rd International IEEE Conference on Intelligent Systems, pp. 512–517, IEEE (2006)
15. Yun, U.: WIS: Weighted interesting sequential pattern mining with a similar level of support and/or weight. *ETRI J.* **29**(3), 336–352 (2007)
16. Chang, J.H.: Mining weighted sequential patterns in a sequence database with a time-interval weight. *Knowl. Based Syst.* **24**(1), 1–9 (2011)
17. Sunitha, G., Rama Mohan Reddy, A.: A region-based framework for mining sequential patterns from spatio-temporal event databases. *Int. J. Appl. Eng. Res.* **9**(24), 28161–28175 (2014)

Performance Analysis in Free Space Optical Communication System Using Aperture Averaging

Saloni Bhatia, Monika Singh and Hemani Kaushal

Abstract The advancement of the terrestrial free space optical communication system (FSO) has shown a drastic increase. But despite the various advantages provided by FSO communication, there are various issues related to the design of optical links. To overcome the problems like scattering, diffraction, turbulence, etc., different techniques were developed, and one of the most widely employed techniques is aperture averaging. The aperture averaging is the main focus of this paper and using on-off (OOK) modulation technique, the bit error rate (BER) performance curves have been analysed for varying data rates and aperture sizes at the receiver using optisystem software.

Keywords FSO · BER · SNR · Turbulence · Aperture averaging

1 Introduction

The demand of bandwidth is increasing day by day and due to licensed and limited radio frequency (RF) spectrum which provides with the data rate in the range of 10 Mbps to several 100 Mbps, the wireless communication companies are now switching on to the alternate solution which can provide with much higher data rate and increased bandwidth. The answer to this alternative lies in the FSO communication where the optical links between source and destination are established using free space as the medium.

The FSO communication system possesses various advantages over the RF and millimetre waves. It provides with the data rate of 10 Gbps and guarantees increased information capacity. The optical carriers with the frequency lying between 10^{12} and 10^{16} Hz can provide maximum data bandwidth up to 2000 THz.

S. Bhatia (✉) · M. Singh · H. Kaushal

Department of EECE, Institute of Technology and Management, Gurgaon,
Haryana 122001, India
e-mail: salonibhatia13@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_32

It also enables greater data security and provides with an extremely narrow beam width with the diffraction limit of 0.01–0.1 mrad. The unlicensed band had proven to be a boon for both the customers and the service providers. It ensures less or no interference between the optical frequencies.

Beside various advantages provided by the FSO communication system, there are various problems which lead to the degradation of the optical link and hinder its usage. Due to the presence of turbulence and the various suspended particles in the free space, the beam gets diverted and erroneous signal could be received at the receiver side. Thus, implementation of long-range FSO link is still a great challenge.

The main cause of the turbulence in the atmospheric channel is due to the temperature inhomogeneities present between the atmospheric layers, which results in the formation of packets of air called eddies, having different refractive indices. To mitigate the turbulence effect, various techniques are adopted like diversity techniques, orthogonal frequency division multiplexing (OFDM), higher data rate coding techniques and aperture averaging, etc. The aperture averaging technique is discussed in this paper and the scenario was designed using optisystem software. BER performance analysis has been carried out using OOK modulation scheme to understand the effect of different aperture sizes over the received signal.

The paper is divided into five sections. Section 2 gives the description of the atmospheric turbulence effects. Section 3 describes the simulation set up using aperture averaging technique. Section 4 presents the simulation results and the plots of BER versus SNR followed by the conclusions in Sect. 5.

2 Turbulence Effects

The particles suspended in the atmospheric channel and various natural phenomena cause the degradation in the quality of the received signal. The sun ray, when strikes the Earth's surface, causes the rise in the temperature and the air near the surface becomes warm as compared to the air at higher altitudes. This warm air rises up and gets mixed up with cool air due to which the characteristic property of channel changes and causes turbulence effect.

The temperature inhomogeneities result in the formation of the air packets also called as eddies which have different refractive indices. When the optical beam passes through these air packets having varying refractive indices, it will get refracted with different angles and get scattered thus causing beam divergence and beam path deviation.

The various other problems caused due to turbulence are beam steering, beam scintillations, beam spreading, polarisation fluctuations and spatial coherence degradation [1].

The atmospheric turbulence depends upon the altitude, wind speed and the inhomogeneities present in the atmosphere, pressure and temperature. This results

in the random fluctuations in the refractive index of atmospheric channel which is given by [2]

$$n = 1 + \frac{77.6(1 + 7.52 * 10^{-3} * \lambda^{-2})P}{T} * 10^{-6}, \tag{1}$$

where

- n refractive index
- λ wavelength of light
- P atmospheric pressure
- T temperature in Kelvin.

The smaller size eddies are given by small inner scale length l which lies within few millimetres, whereas larger size eddies are given by outer scale length L . Due to the high turbulence flow, the larger size eddies having size equal to flow dimension start moving randomly, thus forming smaller scale eddies and as the turbulence flow becomes equal to the inertial flow, the further disintegration of larger size eddies stops [1, 3]. Thus, the energy from larger size eddies having outer scale length L begins to cascade and redistribution of energy takes place from larger size eddies to smaller size eddies having inner scale length l as shown in Fig. 1.

The FSO channel undergoes slow fading since the coherence length [1, 2] is much larger than the symbol duration. These eddies are generally fixed but sometimes moves with the wind speed at right angle to the transverse wave.

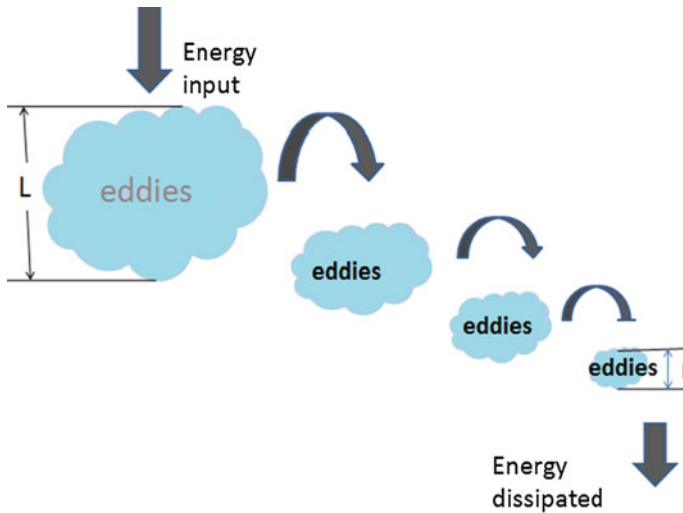


Fig. 1 Redistribution and dissipation of the energy from larger eddies to the smaller eddies due to inertial flow

The turbulent flow is a result of the inertial flow which occurs due to the varying velocities of light in various eddies. The turbulent flow is characterised by the Reynolds number Re , given as [1]

$$Re = \frac{UL}{\nu}, \quad (2)$$

where

U characteristic flow velocity

L characteristic dimension of viscous flow

ν kinematic viscosity of fluid.

As the characteristic flow velocity increases, the value of Re increases due to which the turbulent flow also increases. For turbulent flow to occur, the value of Re should be greater than 2500–5000. When optical beam propagates through such a turbulent atmosphere, it will cause fluctuations in both amplitude and phase of the received beam. This will make the optical links unreliable for the long-range communication [4, 5].

3 Aperture Averaging

The aperture averaging is one of the techniques by which the amount of radiations reaching the receiver can be increased by increasing the receiver aperture size. This will reduce the loss of radiations due to increased beam divergence. The aperture size of the receiver is increased so that most the radiance is collected by the lens [6, 1].

3.1 Simulation Set Up

The performance of aperture averaging technique was evaluated for an On–Off Keying (OOK) which is also known as intensity modulation scheme. The OOK modulation scheme is the most usable scheme in the terrestrial FSO communication because of its ease of employment and simplicity. It can be employed with both NRZ (non return to zero) and RZ (return to zero) line codes. The BER performance was analysed for various receiver aperture diameters at the wavelength of 1550 nm. The simulation setup is illustrated in Fig. 2 and was designed in Optisystem Software. The scenario consists of the following components:

1. *Pseudo-Random Bit Sequence Generator (PRBS)*: It was used to generate the random sequenced data.

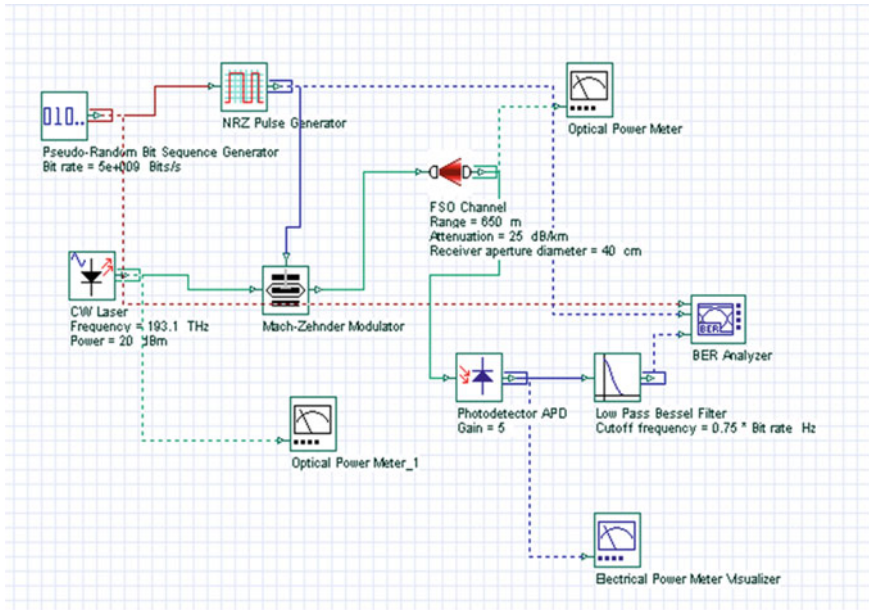


Fig. 2 The scenario designed for the OOK modulation scheme in the optisystem software depicting the various components used

2. *NRZ Pulse Generator*: The output of PRBS is given as the input to the pulse generator which generates the rectangular pulses according to random data provided by PRBS. Its output is given to Mach-Zehnder (MZ) modulator.
3. *Mach-Zehnder Modulator (MZ modulator)*: It is used to intensity modulate the logical data which is provided by NRZ pulse generator with the light waves coming from the CW laser.
4. *Optical Power Meter*: It is used to measure the amount of power transmitted and received.
5. *FSO Channel*: It acts as an unguided medium for the light propagation. The attenuation of the channels was set at 25 dB/km and the link distance was varied to obtain the readings.
6. *Photodetector*: At receiver side, Avalanche Photodiode (APD) is used to convert the light energy into electrical energy. The gain of APD was set to 5.
7. *Low-Pass Bessel Filter*: It is used to filter the signals at the receiver side to obtain a transmitted signal.
8. *Bit Error Rate Analyser*: It is used to evaluate the BER, Quality factor and eye diagram of the signals received.

4 Simulation Results

The simulation was carried out at two data rates, 5 and 10 Gbps, considering three different link distances, 650 m, 800 m and 1 km. The noise power was taken to be 1 mW (0 dBm) and graphical results were obtained. The operating wavelength was chosen to be 1550 nm.

Case 1. Aperture diameter of 20 cm

The graphical results obtained for bit rates of 5 and 10 Gbps are as shown in Fig. 3a, b.

Following conclusions can be drawn for graphical results obtained in Fig. 3.

1. As link distance increases, the required SNR to achieve same BER performance also increases irrespective of data rate.
2. The increase in the required SNR with increase in data rate is very marginal for smaller link distances, and required SNR increase as link distance increases to achieve same BER performance as shown in Table 1.

Case 2. Aperture diameter of 40 cm

The values obtained at an aperture diameter of 40 cm are depicted in Table 2 at two different data rates of 5 and 10 Gbps and graphical results are shown in Fig. 4a, b.

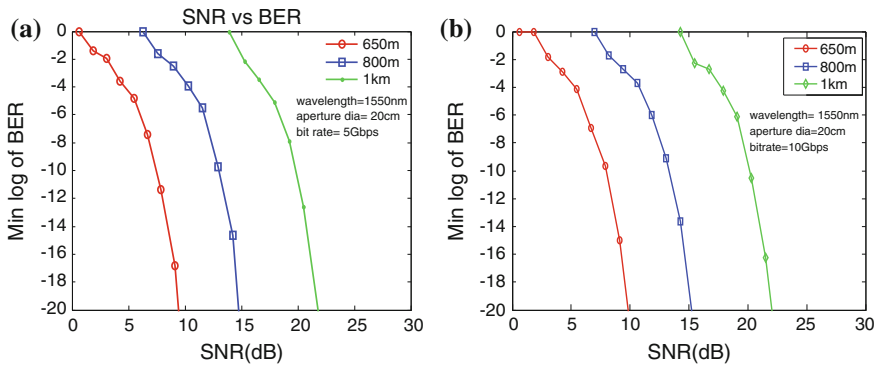


Fig. 3 BER versus SNR (dB) plot for link distances of 650 m, 800 m and 1 km for aperture diameter of 20 cm at bit rates of **a** 5 Gbps and **b** 10 Gbps

Table 1 Values of required SNR (dB) at data rates 5 and 10 Gbps for link distances of 650 m, 800 m and 1 km with aperture diameter of 20 cm at BER 10^{-9}

Link distance	SNR (dB) BER 10^{-9}	
	Bit rate (5 Gbps)	Bit rate (10 Gbps)
650 m	7.1875	7.89
800 m	12.72	13.6
1 km	19.54	21.56

Table 2 Values of required SNR (dB) at data rates 5 and 10 Gbps for link distances of 650 m, 800 m and 1 km with aperture diameter of 40 cm at BER 10^{-9}

Link distance	SNR (dB) BER 10^{-9}	
	Bit rate (5 Gbps)	Bit rate (10 Gbps)
650 m	0.9284	1.405
800 m	6.26	8.994
1 km	13.57	13.83

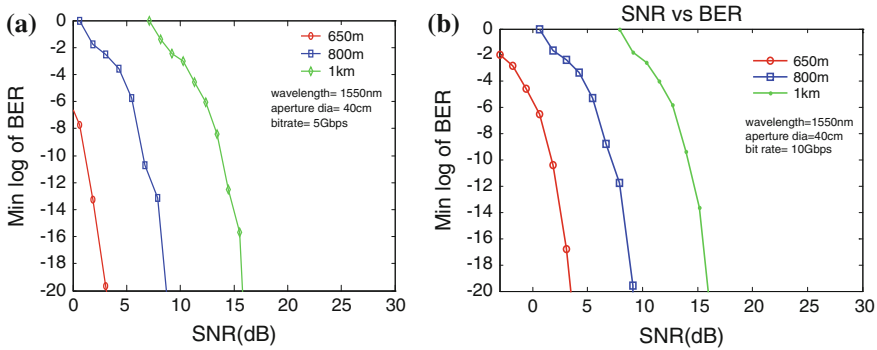


Fig. 4 BER versus SNR (dB) plot for link distances of 650 m, 800 m and 1 km for aperture diameter of 40 cm at bit rates of **a** 5 Gbps and **b** 10 Gbps

Comparing Tables 1 and 2, it is observed that for the same bit rate and link distance, the values of required SNR decreases with the increase in the aperture diameter. Hence, for a given BER performance, increase in receiver aperture diameter reduces the transmitter power requirement to transmit the data at same data rate and link distance.

Case 3. Aperture diameter of 60 cm

The required SNR values obtained at a BER 10^{-9} are depicted in Table 3 and the graphical plots for the same are shown in Fig. 5a, b at the bit rates of 5 and 10 Gbps, respectively.

Table 3 Values of required SNR (dB) at data rates 5 and 10 Gbps for link distances of 650 m, 800 m and 1 km for aperture diameter of 60 cm at BER 10^{-9}

Link distance	SNR (dB) BER 10^{-9}	
	Bit rate (5 Gbps)	Bit rate (10 Gbps)
650 m	-2.507	-1.998
800 m	3.227	3.408
1 km	9.941	10.47

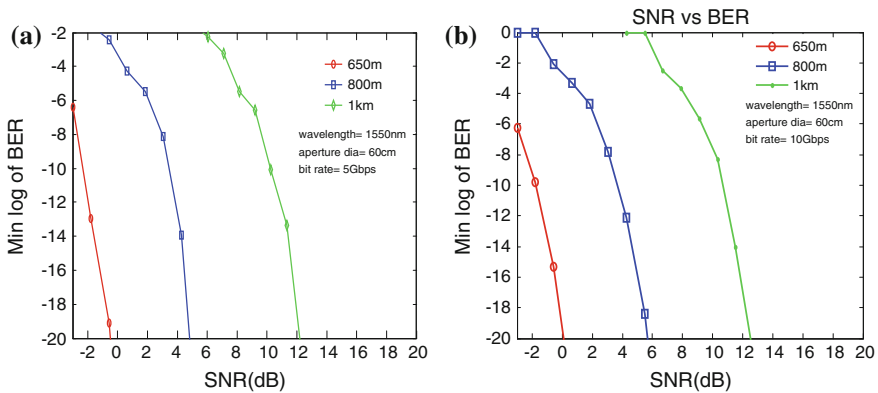


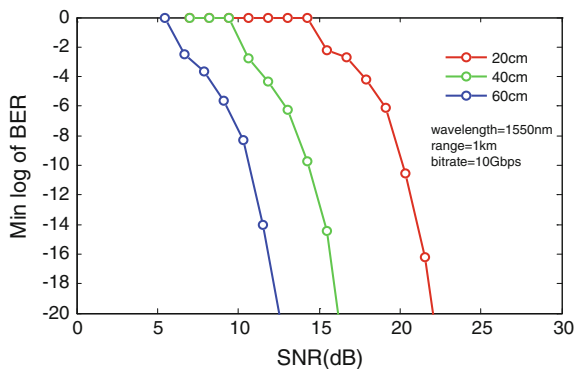
Fig. 5 BER versus SNR (dB) plot for link distances of 650 m, 800 m and 1 km for aperture diameter of 60 cm at bit rates of **a** 5 Gbps and **b** 10 Gbps

The same conclusions can be drawn for Case 3 as for Case 1 and Case 2 but the required SNR value for this case decreases even further. This depicts that less transmitter power is required to transmit the same data with increase in aperture diameter. However, required aperture area cannot be increased indefinitely as it will lead to increase in background noise as well.

5 Conclusion

1. For same aperture diameter, the required SNR increases with increase in data rate for given BER performance.
2. If same data rate is considered, then with increase in aperture diameter, the required SNR decreases. For example, at data rate of 5 Gbps and link distance of 1 km, the SNR (dB) values for aperture diameters of 20, 40 and 60 cm are 21.56, 13.83 and 10.47 dB, respectively. This can be concluded from Fig. 6.

Fig. 6 Comparison of BER versus SNR (dB) at a data rate 10 Gbps for aperture diameters of 20 cm, 40 cm and 60 cm



3. The relative increase in required SNR is less as aperture size increases for a given link distance, and thus the required transmitter power is less. For example, considering 650 m link distance for a data rate of 5 Gbps, using Tables 1, 2 and 3 for the relative increase in aperture diameter from 20 to 40 cm, the SNR is 6.259 dB, and for the relative increase in aperture diameter from 40 to 60 cm the SNR is 3.434 dB.

Thus, for the same BER performance, it is affordable to work at higher data rates with increase in aperture diameter as the value of required SNR decreases. However, the main disadvantage which lies in the use of increased aperture size is the increase in the noise power. Thus, a compromise has to be made between the required SNR and aperture diameter to get the desired results.

References

1. Wasiecko, L.M.: Techniques to mitigate the effects of atmospheric turbulence on free space optical communication link. University of Maryland, UK, Ph.D. thesis (2004)
2. Ghassemlooy, Z., Popoola, W.O.: Terrestrial free-space optical communications. Optical Communications Research Group, NCR lab, Northumbria University, Newcastle upon Tyne, vol. 5–7, pp. 195–212, (2014)
3. Churnside, J.H.: Aperture averaging of optical scintillations in the turbulent atmosphere. *Appl. Opt* **30**(15) 1982–1994 (1991)
4. Andrews, L.C.: Aperture-averaging factor for optical scintillations of plane and spherical waves in the atmosphere. *J. Opt. Soc. Am* **9**(4), 597–600 (1992)
5. Churnside, J.H.: Aperture-averaging factor for optical propagation through the turbulent atmosphere. NOAA Technical Memorandum ERL WPL-188 (1990)
6. Gilbreath, C.G., Rabinovich: Research in free space optical data transfer at the U.S Naval Research laboratory. In: Proceedings of SPIE, vol. 5160, Free Space Laser Communication and Active Laser Illumination III (2003)

Performance Evaluation of Free Space Optical Link Under Various Weather Conditions

Monika Singh, Saloni Bhatia and Hemani Kaushal

Abstract Free space optics (FSO) is emerging as a viable complementary technology to address the need for larger bandwidth and high data rate at affordable cost. FSO communication systems face severe link availability and reliability challenges under different weather conditions, and this is a limitation for the wide-scale acceptability of the FSO technology. The main objective of this paper is to analyze the impact of fog, snow, and rain on FSO link, and hence evaluate the performance of the FSO system for various weather conditions. It is analyzed that bit error rate (BER) and link margin of FSO system are very poor for heavy fog, rain, and snow. However, decreasing the data rate for a particular weather condition can improve both these performance parameters.

Keywords Weather effects · Simulation · Link length · Bit error rate · Data rate · Link margin · Visibility · Attenuation

1 Introduction

Ever-increasing demands of the communication world for high data rates and bandwidth can easily be met by free space optics. By the virtue of inherent high optical carrier frequency (in 20–375 THz range), FSO communication system can provide highest data speeds making it suitable for first/last mile access connectivity. Due to ease of deployment and light weight systems used in FSO technology, it finds application in disaster recovery to provide ad hoc connectivity using temporary links. Moreover, being a cost-effective technology, FSO can be used to provide communication over difficult terrains where installation of conventional communication systems is a costly affair. With the advent of more mature

M. Singh (✉) · S. Bhatia · H. Kaushal
Department of Electrical, Electronics and Communication Engineering,
ITM University, Sec 23A, Gurgaon 122017, Haryana, India
e-mail: monikasingh9@gmail.com

optoelectronic devices, FSO communication has seen a steep rise in its growth and acceptance. A major challenge being faced by FSO communication is its sensitivity to the weather conditions and the turbulence in the atmosphere. This leads to FSO link availability and reliability issues. An insight in the performance behavior of the FSO links under various weather conditions is one of the requirements to decide FSO system design parameters prior to its deployment. Fog, rain, and snow which are most deterrent weather phenomenon lead to attenuation in the optical signal, and these attenuation can be parameterized/measured using visibility range of the optical signal in various weather conditions.

This paper has been organized as follows: Section 2 describes the characterization of snow, fog, and rain attenuation using empirical models. Section 3 enumerates the detailed aspects and schematic for simulation. In Sect. 4, performance of the FSO system under different weather conditions is analyzed, and finally conclusions of the analysis are discussed in Sect. 5.

2 Characterization of Attenuation Due to Weather Phenomenon

2.1 Fog

Fog particles are considered to be major photon scatterers since their size is comparable with the wavelength of interest in FSO communication (0.5–2 μm). It causes significant attenuation of the signal for considerable amount of time and this is deterrent for achieving high availability in FSO transmissions. The optical attenuation in case of fog droplets can be accurately predicted by applying Mie scattering theory. An approach is based on visibility range information, in which the fog attenuation is predicted using common empirical models. Visibility range is defined as the distance that a parallel luminous beam travels in the atmosphere until its intensity drops 2 % of its original value. The wavelength of 550 nm is usually taken as the visibility range reference wavelength. Equation 1 defines the attenuation coefficient given by common empirical models for Mie scattering [1]:

$$\beta_{\text{fog}} = \frac{3.91}{V} \left(\frac{\lambda}{550} \right)^{-p} \quad (1)$$

where V (km) stands for visibility range, λ (nm) is the operating wavelength, and p is the size distribution coefficient of scattering.

According to Kim model [2] p is given as

$$p = \begin{cases} 1.6 & V > 50 \\ 1.3 & 6 < V < 50 \\ 0.16V + 0.34 & 1 < V < 6 \\ V - 0.5 & 0.5 < V < 1 \\ 0 & V < 0.5 \end{cases} \quad (2)$$

According to Kruse model [3], p is given as

$$p = \begin{cases} 1.6 & V > 50 \\ 1.3 & 6 < V < 50 \\ 0.585V^{\frac{1}{3}} & V < 6 \end{cases} \quad (3)$$

Different weather conditions can be specified based on their visibility range values. Table 1 summarizes the visibility range for different weather conditions [4].

2.2 Effect of Snow

Snow attenuation can vary depending upon the snowflake size and snowfall rate. Since snowflakes are larger in size than rain drops, hence, they produce deeper fades in the signal as compared to the rain drops. Snowflake size can be as large as 20 mm which can completely block the path of the optical signal depending upon the beam width of the signal. For snow, FSO attenuation can be classified into dry and wet snow attenuation. The dry snow affects at the low snow rate, whereas wet snow affects at high snow rate. Specific attenuation is a fundamental factor used to determine attenuation at a defined distance and it is measured in dB/km. The specific attenuation for snow rate S in mm/h is given by [5]

$$\beta_{snow} = aS^b \quad (4)$$

For an operating wavelength λ , values of parameters a and b in dry and wet snow are $a = 5.42 * 10^{-5} + 5.4958776$; $b = 1.38$ and $a = 1.023 * 10^{-5} + 3.7855466$;

Table 1 Visibility range values corresponding to weather conditions

Weather conditions	Visibility range (km)
Heavy fog	0.2
Moderate fog	0.5
Light fog	0.770–1
Thin fog	1.9–2
Haze	2.8–40
Clear	18–20
Very clear	23–50

$b = 0.72$, respectively. The snow attenuation based on visibility range can be approximated by the empirical model [5] given as

$$\alpha_{\text{snow}} = \frac{58}{V} \quad (5)$$

2.3 Effect of Rain

The optical signal is randomly attenuated by rain and fog droplets. Although the main factor for attenuation is fog, the sizeable rain droplets can also cause wavelength-independent scattering. The attenuation produced by rainfall increases linearly with rainfall rate. The specific attenuation [6] against rain rate of R (mm/h) for a FSO link is given as

$$\beta_{\text{rain}} = 1.067R^{0.67} \quad (6)$$

The rain attenuation for FSO link can be reasonably well approximated by empirical formula [7] given as

$$\alpha_{\text{rain}} = \frac{2.8}{V} \quad (7)$$

where V is the visibility range in km and its values based on rainfall rates [1] are summarized in Table 2.

2.4 Visibility Calculation for Cloud

The cloud attenuation can be evaluated using the visibility condition, which is given by

$$V = a(LWC)^b \quad (8)$$

In the above expression, LWC stands for liquid water content. For a daylight-level illumination, values of the empirical constants ‘ a ’ and ‘ b ’ are given by $65 \leq a \leq 178$

Table 2 Rainfall rates and their visibility ranges

Rainfall type	Rainfall rate (mm/h)	Visibility range (km)
Heavy rain	25	1.9–2
Medium rain	12.5	2.8–40
Light rain/drizzle	0.25	18–20

Table 3 Visibility range values for different cloud types

Cloud type	LWC (g/m^{-3})	Visibility range (m)
Cumulus	1.0	65–178
Stratus	0.29	29.8–54.2
Stratocumulus	0.15	19.7–28.8

and $0.63 \leq b \leq 0.96$. Table 3 shows LWC which is taken from modified gamma distribution [7].

Once the visibility range is estimated, the cloud attenuation can be predicted using the visibility range-dependent Kruse [3] empirical model given by Eq. (3). Even though there are more cloud types present in the atmosphere, only three types of clouds (Stratus, Cumulus, and Stratocumulus) have been considered in this paper since these clouds are present at a lower height which is below 1 km, whereas the height of the buildings is only in meters.

3 Simulation

In this paper, numerical simulations using MATLAB have been carried out for various weather conditions based on which the performance of the FSO communication system has been analyzed. Furthermore, a FSO communication system with (On–Off keying Non–return to zero) OOK NRZ modulation scheme has been simulated using simulation software OptiSystem 12.0 to verify the results obtained in the numerical simulation. The simulation block diagram of the FSO communication system is shown in Fig. 1.

The transmitter block consists of laser diode as optical source, a modulating data source, and an external modulator (Mach–Zehnder modulator). OOK is the most

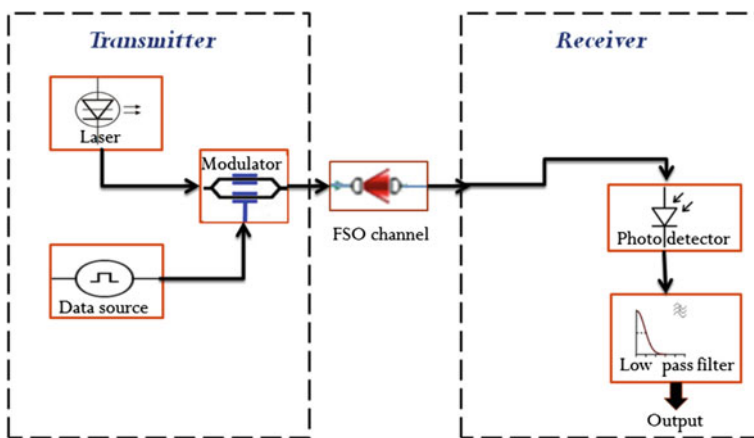


Fig. 1 Block diagram of FSO communication system with OOK modulation

commonly used modulation scheme in terrestrial FSO communication systems, primarily due to its simple implementation and resilience to laser nonlinearity.

Atmospheric channel has been used as FSO communication channel in the analysis. To analyze the effects of various weather conditions on FSO link, the attenuation of the channel has been kept variable. By varying the link range of the channel, a good approximation of the system performance under different visibility ranges can be measured. Figure 2 shows the simulation schematic of the FSO system with OOK modulation. All the simulations have been carried out at 1550 nm wavelength for the obvious reasons: its compatibility with third window wavelength division multiplexing networks, eye safety, and reduced scattering in light haze/fog. The simulation parameters are given in Table 4.

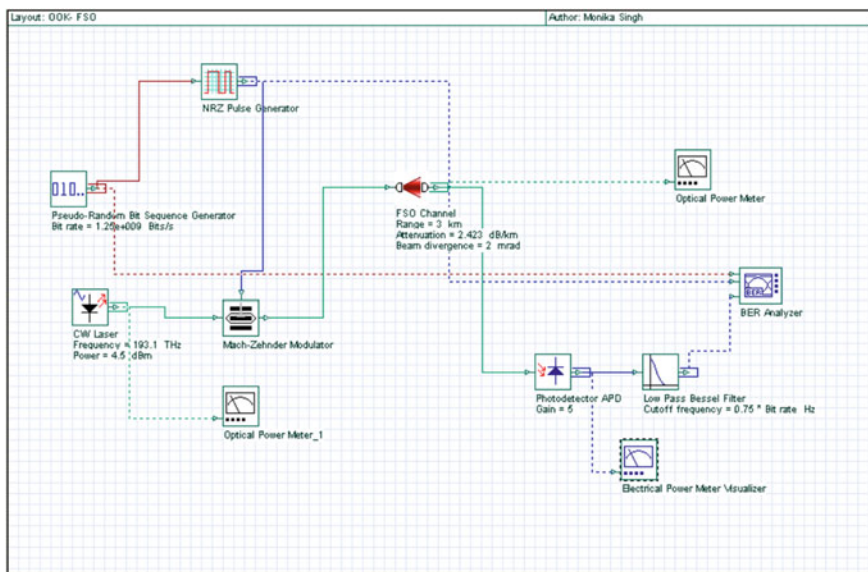


Fig. 2 OptiSystem simulation schematic of the FSO communication system

Table 4 System parameters used in FSO link analysis

Parameter	Symbol	Value
Transmitter optical power	P_t	5 mW
Divergence angle	θ	2 mrad
Receiver aperture diameter	D_R	15 cm
Transmitter efficiency	ψ_t	0.5
Receiver efficiency	τ_r	0.5
Laser wavelength	λ	1550 nm
Receiver sensitivity	N_b	-20 dBm

4 Results and Analysis

4.1 Attenuation Versus Visibility

Variation of attenuation with visibility for heavy and light fog is shown in Figs. 3 and 4, respectively. For a low-visibility weather condition, i.e., heavy fog, operating wavelength has a negligible effect on the specific attenuation, whereas for light fog when the visibility range is high (≥ 6 km), attenuation is quite less for an operating wavelength of 1550 nm as compared to 850 and 950 nm. Dependency of the attenuation on wavelength decreases when visibility increases beyond 20 km (clear weather).

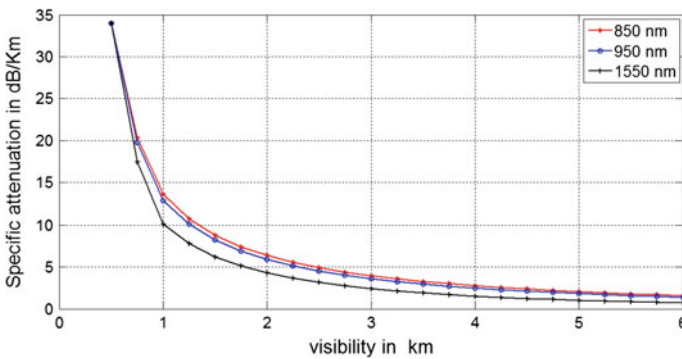


Fig. 3 Attenuation versus visibility plot for heavy fog

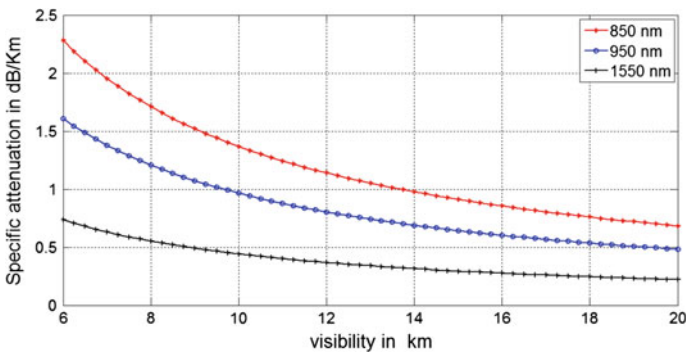


Fig. 4 Attenuation versus visibility plot for light fog

4.2 Attenuation Due to Weather at 1550 nm

The attenuation for different weather conditions, i.e., heavy fog and cloud, light fog, and haze at 1550 nm, are shown in Figs. 5 and 6, respectively, and are tabulated in Table 5. These values of attenuation at 1550 nm are used to characterize different weather conditions in further simulations.

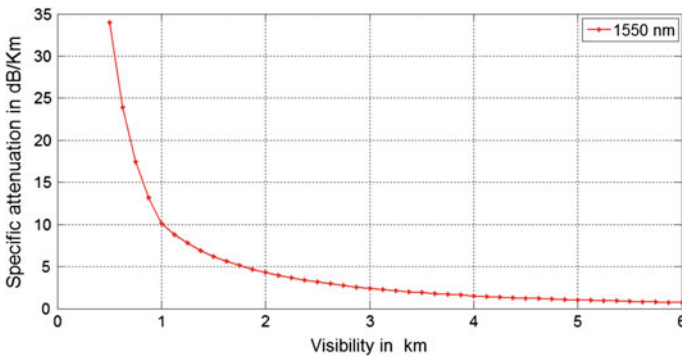


Fig. 5 Attenuation for heavy fog and cloud at 1550 nm

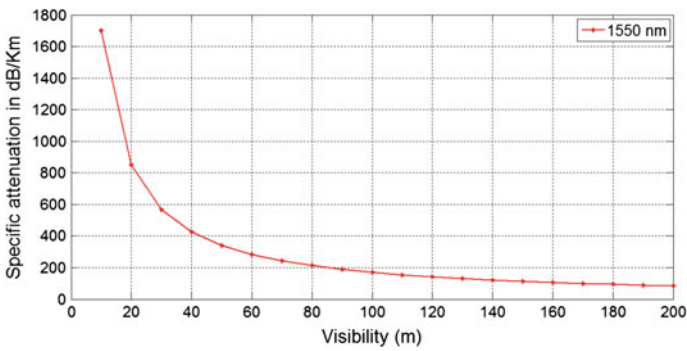


Fig. 6 Attenuation for light fog and haze at 1550 nm

Table 5 Attenuation at 1550 nm

Weather conditions	Visibility range (m)	Attenuation (dB/km)
Haze	3000	2.423
Light fog	875	13.17
Heavy fog	100	170
Cloud	30	566.7

4.3 Comparison of Attenuation for Fog, Snow, and Rain

Scattering in different fog conditions gives significantly high attenuation because the size of fog particles is comparable to the optical transmission wavelengths. It can be observed from Fig. 7 that attenuation is maximum for snow and less for rain as compared to fog. In case of fog, there is a sudden increase in attenuation for visibility range less than 150 m which exactly matches with the data provided in Table 1 (i.e., visibility range <200 m corresponds to heavy fog) for an operating wavelength of 1550 nm.

4.4 Received Power for FSO System

Consider a laser source L_s transmitting power P_T at a wavelength of 1550 nm. The received signal power P_R [3] can be obtained as

$$P_R = P_T \frac{D_R^2}{\theta^2 L^2} 10^{-\gamma L/10} \tau_t \tau_r \tag{9}$$

where D_R is the receiver aperture diameter, θ is the beam divergence, L is the link length, γ is the atmospheric attenuation factor, τ_r and τ_t are receiver and transmitter optical efficiencies, respectively. For a receiver sensitivity of -20 dBm, achievable transmission distances for cloudy, heavy fog, light fog, and haze are 45, 105, 435, and 691 m, respectively. These observations are made based on the graph shown in Fig. 8. Similar observations have been made using OptiSystem and it has been shown in Fig. 9.

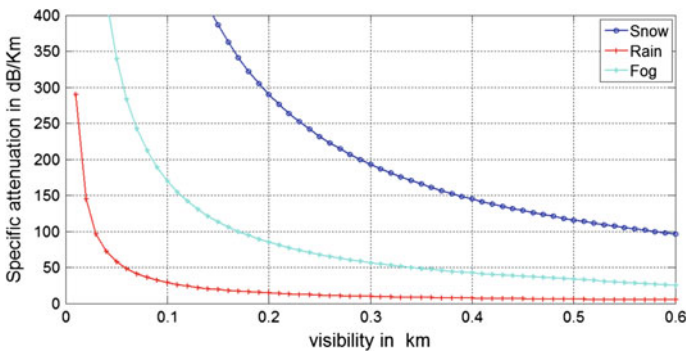


Fig. 7 Attenuation for fog, snow, and rain

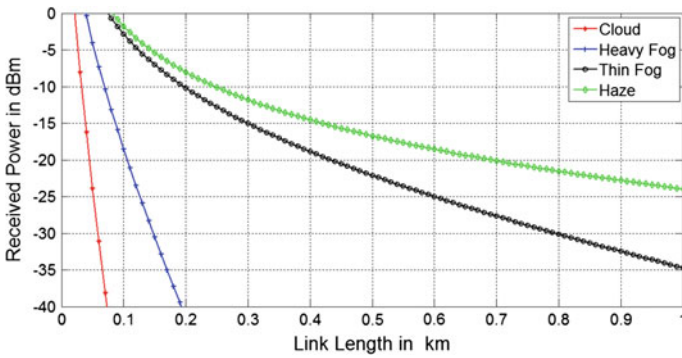
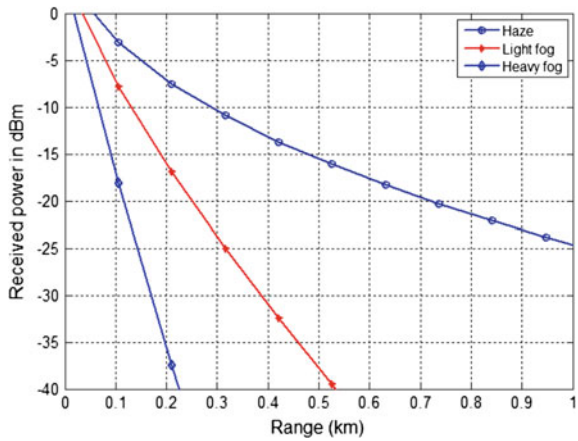


Fig. 8 Output power versus link length

Fig. 9 Output power versus link length (software simulated)



4.5 Data Rate for FSO System

For a given receiver sensitivity (in this case, it is -20 dBm equivalent to 327,000 photons/bit), the achievable data rate R [8] is given by

$$R = \frac{P_T D_R^2 10^{-\gamma L/10} \tau_t \tau_r}{\pi \left(\frac{\theta}{2}\right)^2 L^2 E N_b} \tag{10}$$

where $E = hc/\lambda$ is the photon energy at wavelength λ and N_b is the receiver sensitivity in photons/bit, h is the Planck's constant, and c is the velocity of light. Data rate of 1 Gbps is obtained at a link distance of 20 m for cloudy weather, 80 m for heavy fog, and 111 m for haze as shown in Fig. 10.

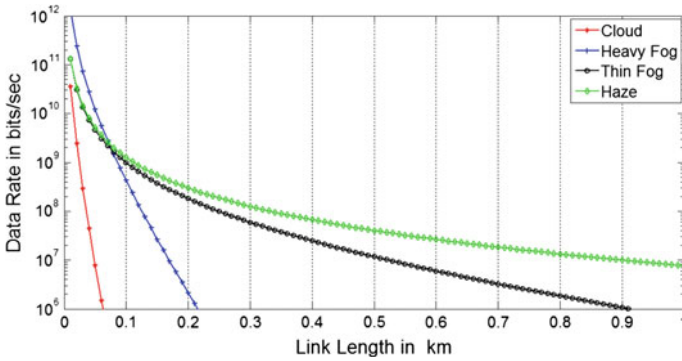


Fig. 10 Data rate versus link length

4.6 Link Margin for FSO System

The link margin is included in the link budget equation in order to compensate for scattering, absorption, and scintillation losses at a given range. Link margin [8] is defined as the ratio of the available power at the receiver to the power required for achieving a specified BER at a given data rate and it can be expressed as

$$L_M = \left[\frac{P_T \lambda}{N_b R h c} \right] * \left[\frac{a l D^2}{\theta^2 L^2} \right] * 10^{-\gamma L / 10} \tau_r \tau_t \tag{11}$$

Link margin has been analyzed for different weather conditions as well as for different data rates in case of haze.

For a fixed link range, link margin decreases with increase in bit rate. For a link range of 100 m in haze, link margin at data rates 500 Mbps, 1.25 Gbps, 2.5 Gbps, and 5 Gbps are 15, 11, 10, and 8 dB, respectively. This is shown in Fig. 11. It is

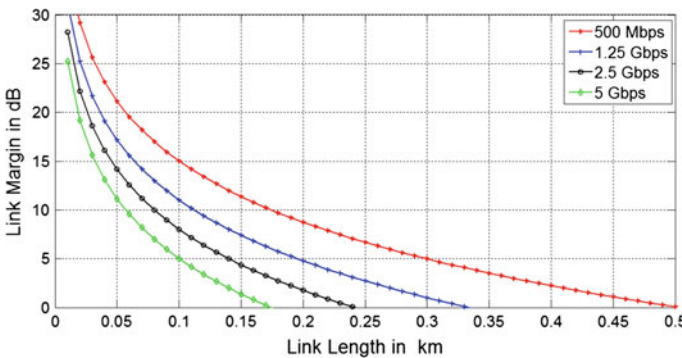


Fig. 11 Link margin for different data rates

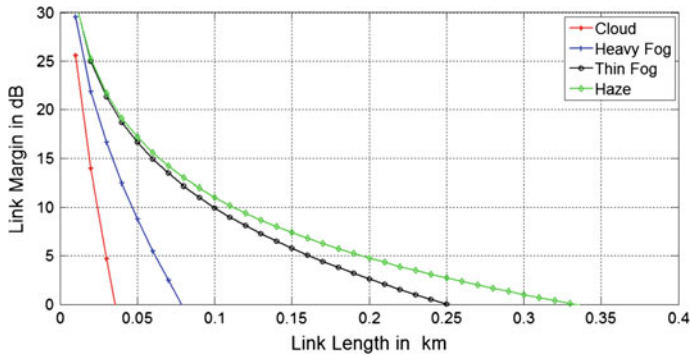


Fig. 12 Link margin for different weather conditions

apparent from Fig. 12 that link margin of 10 dB is obtained for link lengths of 24, 46, and 112 m in case of cloudy, heavy fog, and haze weather conditions, respectively.

4.7 Bit Error Rate

Bit error rate is the probability of an incorrect identification of a bit by the decision circuit of a receiver. For communication systems, criteria used are that the $BER \leq 10^{-9}$. The BER results for different weather conditions show that BER is very poor for heavy fog and snow as shown in Fig. 13. Using the simulation results in Fig. 14, we can infer that BER performance of the FSO system degrades with

Fig. 13 BER versus link range for different weather conditions

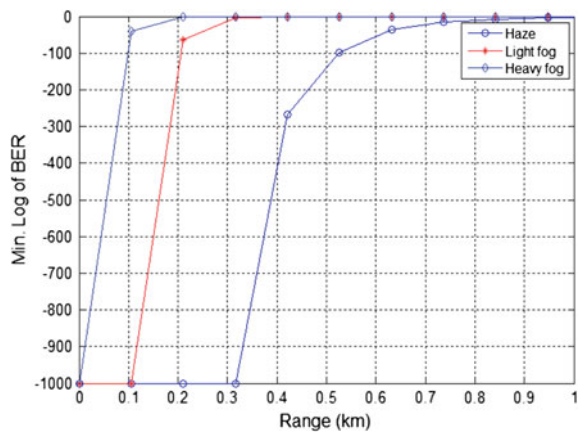
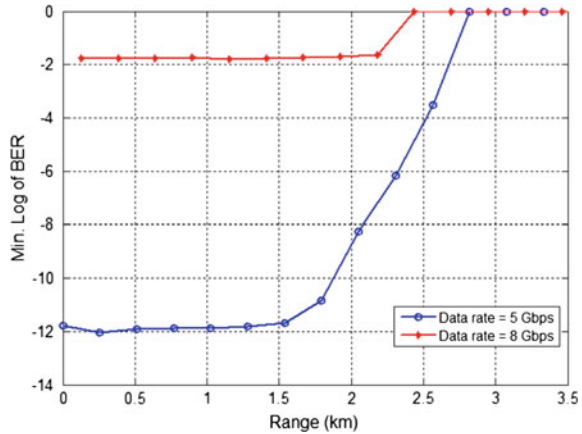


Fig. 14 BER versus link range for different data rates



increase in data rate. For a link range of 1.5 km, the FSO system degrades with increase in data rate. For a link range of 1.5 km, the FSO system with lower data rate say 5 Gbps gives BER of 10^{-12} .

5 Conclusions

Despite its major advantages, FSO access technology suffers from severe availability and reliability challenges mainly due to different weather effects like fog, rain, snow, and clouds in the earth atmosphere. Following are the conclusions made on the basis of the above simulation and analysis:

1. For a OOK-modulated FSO system, the simulation results show that attenuation due to absorption and scattering is minimum at 1550 nm, and thus using 1550 nm optical wavelength for transmission in an FSO system is a better choice.
2. Due to sudden increase in specific attenuation in case of heavy fog, there is a substantial degradation in the performance FSO communication link.
3. It was also observed that a link margin of 10 dB is obtained for longer link lengths in haze as compared to cloud and heavy fog, which shows that an FSO system’s performance would be less affected by haze as compared to cloudy and heavy fog weather.
4. There is a tradeoff between the data rate and link margin. For a particular weather condition, higher link margins can be obtained by decreasing the data rate and vice versa.
5. BER decreases with increasing data rates. Also, BER performance is very poor in heavy fog and snowy weather.

6. On increasing the link length, the achievable data rate drops down significantly for heavy fog and cloudy weather which can severely affect the communication by leading to complete link outage.

Future work can focus on performance analysis of the FSO systems with modulation schemes like PPM, DPPM, and DPIM under various weather conditions.

References

1. Ghassemlooy, Z., Popoola, W.: Terrestrial free-space optical communications. In: Fares, S.A., Adachi, F. (eds.) Intech, ISBN 978– 953-307-042 (2010)
2. Kim, I.I., McArthur, B., McArthur, A.E.K.B., Korevaar, E.: Comparison of laser beam propagation at 785 and 1550 nm in fog and haze for optical wireless communications (2001)
3. Kruse, P.W., McGlauchlin, L.D., McQuistan, R.B.: Elements of infrared technology: generation, transmission and detection. Wiley, New York (1962)
4. Farukh, N., Kvicera, V., Awan, M., Leitgeb, E., Muhammad, S., Kandus, G.: Weather effects on hybrid fso/rf communication link (2009)
5. O'SBrien, H.W.: Visibility and light attenuation in falling snow, vol. 9, pp. 671–683 (1970)
6. Atlas, D.: Shorter contribution optical extinction by rainfall, vol. 10, pp. 486–488 (1953)
7. Ramprasath, K., Prince, S.: Analyzing the cloud attenuation on the performance of free space optical communication, pp. 3–5, April 2013
8. Guo-yong, H., Chang-ying, C., Zhen-qiang, C.: Free-space optical communication using visible light (2006)

Evolutionary Improved Swarm-Based Hybrid K-Means Algorithm for Cluster Analysis

Janmenjoy Nayak, D.P. Kanungo, Bighnaraj Naik and H.S. Behera

Abstract Improvement in the quality of cluster centers and minimization of intra-cluster distance are two most challenging areas of K-means clustering algorithm. Due to predetermined number of clusters, it is difficult to predict the exact value of k . Furthermore, in case of non-globular clusters, K-means fails to get optimal cluster center in a data set. In this paper, a hybrid improved particle swarm optimization-based evolutionary K-means clustering method has been proposed to obtain the optimal cluster center. The hybridization of improved PSO and genetic algorithm (GA) along with K-means algorithm improves the convergence speed as well as helps to find the global optimal solution. In the first stage, IPSO has been used to get a global solution in order to get optimal cluster centers. Then, the crossover steps of GA are used to improve the quality of particles and mutation is used for diversification of solution space in order to avoid premature convergence. The performance analysis of the proposed method is compared with other existing clustering techniques like K-means, GA-K-means, and PSO-K-means.

Keywords Clustering · K-means · Improved PSO · GA · Hybrid GA-IPSO

J. Nayak (✉) · D.P. Kanungo · B. Naik · H.S. Behera
Department of CSE and IT, Veer Surendra Sai University of Technology,
Burla, 768018 Odisha, India
e-mail: mailforjnyak@gmail.com

D.P. Kanungo
e-mail: dpk.vssut@gmail.com

B. Naik
e-mail: mailtobnaik@gmail.com

H.S. Behera
e-mail: mailtohsbehera@gmail.com

1 Introduction

K-means clustering algorithm [1, 2] is one of the simple, efficient center-based hard clustering techniques used to solve different real-world problems. In the past decade, it has been a key interest among various researchers in diversified application areas like pattern recognition, biomedical applications, image processing, fuzzy logic, cloud computing, and many more. Ensafi et al. [3] used the fuzzy K-means with PSO algorithm for network anomaly detection and they obtained better convergence. By hybridizing ant colony optimization, particle swarm optimization, and K-means, [4] developed APSO-K-means clustering algorithm for speaker recognition. Kader [5] presented a hybrid two-phase GA-IPSO with K-means data clustering algorithm which performs fast data clustering and can avoid premature convergence to local optima. An improved PSO-based K-means algorithm was developed by Xiangwei and Yuanjiang [6] to avoid the local optima problem in normal K-means clustering. Li and Wu [7] proposed an improved K-means algorithm by combining the largest minimum distance algorithm and traditional K-means algorithm. Naik et al. [8] proposed a hybrid PSO—K-means clustering algorithm to get optimal cluster centers for cluster analysis. Govindarajan et al. [9] developed a PSO-based clustering algorithm which improves the quality of student's learning using cloud computing. By decreasing the number of iterations in K-means algorithm, Liao et al. [10] introduced a parallel K-means algorithm based on map reduce. Bai et al. [11] used the local geometrical information to represent the objects in the cluster center of K-means algorithm. A sample-based hierarchical adaptive K-means (SHAKM) clustering algorithm was proposed by Liao et al. [12] to retrieve large-scale videos. An improved K-means with hybridized PSO algorithm for web document clustering has been introduced by Jaganathan and Jaiganesh [13]. After the combination of K-means method and mathematical morphology, Yao et al. [14] developed an improved K-means method for fish image optimization. Monedero et al. [15] presented a modification of the celebrated K-means method for quasi-unsupervised learning by controlling the size of the cluster partitions and adjusted it by means of the Levenberg–Marquardt algorithm. Shahbaba and Beheshti [16] introduced a novel minimum ACE K-means (MACE) clustering method which has the advantage for use in synthetic and real data. To deal with distributed data and overcome the limitations of K-means, Naldi and Campello [17] proposed an evolutionary K-means algorithm for clustering. Scitovski and Sabo [18] analyzed the K-means algorithm for the case of data points occurring at the boarder of the clusters. To detect multilingual text for video indexing, Pavithra and Aradhya [19] proposed the hybridization of three interesting techniques viz., wavelet transform, Gabor filter, and K-means clustering method.

K-means algorithm adopts unsupervised pattern recognition process with the iterative generation of clusters. The objects in the cluster are nonoverlapping in nature. Some properties like robustness, easy understandability and implementation, and fast approach toward the closer cluster center make this algorithm more popular than other existing clustering techniques. Although it is highly influential for solving various

real-life applications compared to other algorithms, it still has some major limitations like sensitivity to local optimal solutions—the area in which more work has to be done. K-means algorithm is able to produce disjoint and flat clusters. In a globular multidimensional space, it performs better than in normal space. Also, the random selection of cluster centers in K-means may not be able to produce feasible solutions.

Both GA and PSO are popular nature-inspired metaheuristic optimization techniques frequently used in almost all fields of science and engineering. The former is inspired by Darwin's concept of evolution, while the later is a swarm-based algorithm inspired by the movement and nature of birds. But GA has some limitations like complex operational parameters including mutation and crossover, generation of global solution, variable response time, and random in nature. On the other hand, PSO suffers from the problem of slow convergence. So, suitable parameter selection and proper adjustments in the parameters of both the algorithms may be helpful to produce efficient cluster centers. Inspired by this, attempt has been made in this paper to obtain suitable optimal cluster centers by hybridizing both the genetic algorithm and improved PSO with K-means algorithm.

The remainder of the paper is organized in the following manner—Sect. 2 describes the basic preliminary concepts like K-means, PSO, IPSO, and GA. In Sect. 3, the proposed method (GA-IPSO-K-means) has been presented. Section 4 presents the experimental setup along with the results obtained. Finally, Sect. 5 concludes our work.

2 Preliminaries

2.1 *K-means Algorithm*

The K-means algorithm receives k number of input parameters and performs the partition on a set of n objects in multidimensional space. The method of K-means starts with the random selection of k number of objects and is represented as cluster means (cluster centers). Depending on the distance metric between the object and the cluster mean, for each of the residual objects, a similar object is assigned which helps to compute a new cluster mean. This process is continued till the convergence of criterion function. Hence, K-means is able to find the best cluster center points in the space. The general steps can be realized in Fig. 1.

2.2 *Particle Swarm Optimization (PSO) Algorithm*

Inspired by bird flocking and fish schooling, Kennedy and Eberhart [20] first proposed the popular metaheuristic algorithm called PSO. It has a few parameter settings unlike GA, which helps to reduce the computational complexity of the algorithm. The working theory of this algorithm is based on two terms called Particle (solutions) and Swarm (population). The particles will move around in the solution space by

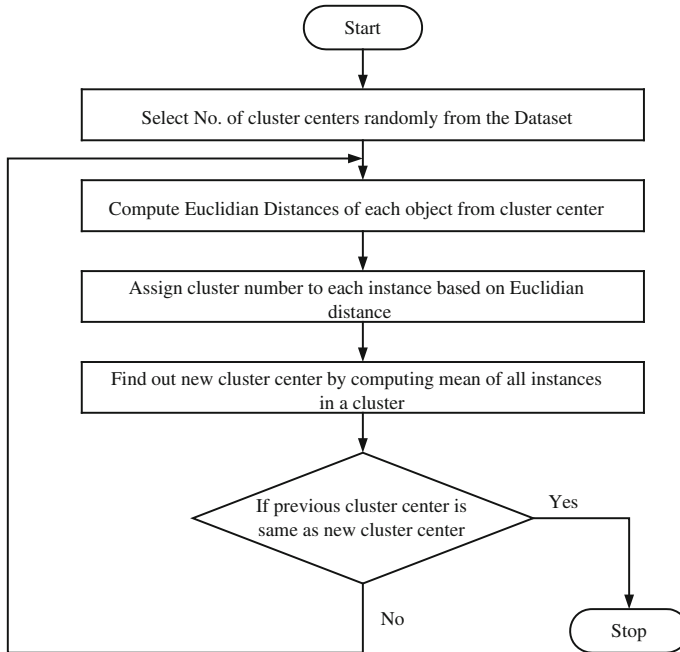


Fig. 1 Steps of k-means algorithm

adjusting their position and velocity. They also exchange the information about the current position with their neighbors in the search space as prior of their own earlier experience [21]. In a group, they travel for either food or shelter [22], without any collision among themselves. After communication of their group information, the particles modify both their positions and velocities as per the best position appeared in the current movement [23]. As a result, the particles would gradually get closer to the specified position and finally reach the optimal position with the help of interactive cooperation [24]. During the movement, each particle maintains their local best position ($lbest$) and the global best position ($gbest$) among themselves.

$$V_i^{(t+1)} = V_i^{(t)} + c_1 * \text{rand}(1) * (l_{best_i}^{(t)} - X_i^{(t)}) + c_2 * \text{rand}(1) * (g_{best}^{(t)} - X_i^{(t)}) \quad (1)$$

$$X_i^{(t+1)} = X_i^{(t)} + V_i^{(t+1)} \quad (2)$$

Equation 1 controls both cognition and social behavior of particles and the next position of the particles are updated using Eq. 2, $V_i(t)$ and $V_i(t + 1)$ are the velocity of i th particle at time t and $t + 1$ in the population, respectively, c_1 and c_2 are acceleration coefficient normally set between 0 and 2 (may be same), $X_i(t)$ is the position of i th particle and $lbest_i(t)$ and $gbest(t)$ denotes local best particle of i th particle and global best particle among local bests at time t , $\text{rand}(1)$ generates random value between 0 and 1.

2.3 Improved PSO Algorithm

In traditional PSO, three basic steps like calculation of velocity, position, and the fitness value [25] are iterated till the required criteria of convergence are met. The ending criteria may be the maximum change in the best fitness value. However, if the velocity of the swarm is fixed to zero or nearer to that and the best position is a fixed value, then the PSO may lead to be trapped at some local optima. This happens only due to the swarm's experience in the current and global positions. This experience is to be avoided and should be based on the mutual cooperation among all the swarms in a multidirectional manner [26]. So, in IPSO a new inertia weight factor λ is introduced to control both the local and global search behavior. The value of λ may be decreased quickly [27, 28] during the initial iterations and slowly during the optimal iterations. The new velocity and position updation can be realized through Eqs. (3) and (4).

$$V_i^{(t+1)} = \lambda * V_i^{(t)} + c_1 * \text{rand}(1) * (l_{\text{best}_i}^{(t)} - X_i^{(t)}) + c_2 * \text{rand}(1) * (g_{\text{best}}^{(t)} - X_i^{(t)}) \quad (3)$$

$$X_i^{(t+1)} = X_i^{(t)} + V_i^{(t+1)} \quad (4)$$

3 Proposed Algorithm (GA-IPSO-K-means)

The proposed GA-IPSO-K-means algorithm is a hybrid algorithm based on the combination of improved PSO, GA, and K-means algorithm for real-world data clustering. Due to the slow convergence speed of basic PSO and easier stuck at local optima in K-means algorithm, the hybridization of GA and improved PSO along with K-means algorithm will improve the convergence speed as well as help to find the global optimal solution. The initial clusters were produced by the Euclidian distance between the cluster object and center. The fitness value of each of the instance will represent the local best and the clusters having the best fitness will be the global best. Accordingly, the updated velocity of the object can be evaluated from the local best, global best, and initial velocity. After execution of the steps of IPSO, the crossover steps of GA are used to improve the quality of particles and mutation is used for diversification of solution space in order to avoid premature convergence. So, the advantages of both the algorithms have been incorporated to achieve an efficient result than the use of any individual population-based algorithms. After execution of pseudo-code of the proposed algorithm, select the solution vector (cluster centers) from the resultant population having maximum fitness value. Initialize the best cluster centers to K-means algorithm and iterate the steps until no significant changes in consecutive cluster centers.

3.1 Pseudo Code of GA-IPSO-K-means Algorithm

Pseudo code of GA-IPSO-K-means Algorithm

Initialize the position P and velocity V of particles randomly. Each particle is a potential solution for clustering problem.

A single particle represents the centroid of clusters. Hence the population of particle is initialized as follows: $P = \{X_1, X_2 \dots X_n\}$

Where X_i represents the centroid of clusters which is a single possible solution (particle) in search space and can be denoted as:

$$X_i = \{C_{i,1}, C_{i,2} \dots C_{i,m}\}$$

Where $C_{i,j}$ represents j^{th} cluster center among m^{th} clusters in the datasets.

Iter=1;

While (iter<=maxIter)

 Compute fitness of all particles X_i in population P by using following objective function in eq. (5):

$$F(X_i) = \frac{k}{\left(\sum_{j=1}^m \sum_{i_k \in C_{i,j}} (i_k - C_{i,j})^2 \right) + d} \quad (5)$$

If (iter==1)

 Assign Local best particle lbest=P.

Else

 Evaluate fitness of P and P'.

 Compare the fitness of particles based on their fitness in P and P'.

If fitness of i^{th} particle X_i in P is less than fitness of a particle in P'

 Then assign Lbest (i) = P'(i).

Else assign Lbest (i) = P(i).

End of if

End of if

 Select particles with best fitness value from Lbest as Gbest particle.

 Compute new velocity Vnew of the particle by using P, Lbest and gbest as follows:

$$Vnew_i^{(t+1)} = \lambda * V_i^{(t+1)} + c_1 * rand(1) * (l_{best_i}^{(t)} - X_i^{(t)}) + c_2 * rand(1) * (g_{best}^{(t)} - X_i^{(t)})$$

 Generate next positions of particles P' by using P and Vnew as follows:

$$X_i^{(t+1)} = X_i^{(t)} + Vnew_i^{(t+1)}$$

 Sort the particles in P' according to their fitness.

 Create a Mating pool of particles by replacing weak particles with global best Gbest particle.

 Perform two point crossovers on particles in P' to generate new feasible solutions P''.

If P' is same as P'',

 Then perform mutation on P'.

End if

 P'=P''.

 Iter = iter+1;

End of while

4 Experimental Setup and Result Analysis

The proposed GA-IPSO-K-means has been implemented in MATLAB and compared with other alternatives (K-means, GA-K-means, and PSO-K-means). All the clustering methods are tested with multidimensional real-world datasets (Table 1) from UCI repository [29] and have been compared in terms of fitness value of the cluster centers from Eq. 5. The comparison of clustering methods is listed in Table 2. The proposed method has been implemented using MATLAB 9.0 on a system with an Intel Core Duo CPU T5800, 2 GHz processor, 2 GB RAM, and Microsoft Windows-2007 OS.

In Eq. (5), k and d are the parameters used to calculate the fitness of clustering methods along with the proposed method. The simulation has been carried out by setting the values $k = 50$, $d = 0.1$ and the proposed clustering method is found to be

Table 1 Dataset information

Datasets	Number of pattern	Number of clusters	Number of attributes
Iris	150	3	4
Lenses	24	3	4
Haberman	306	2	3
Balance scale	625	3	4
Wisconsin breast cancer	699	2	10
Contraceptive method choice	1473	3	9
Hayesroth	132	3	5
Robot navigation	5456	4	2
Spect heart	80	2	22

Table 2 Performance comparison of GA-IPSO-K-means with other clustering methods

Datasets	Fitness values of clustering algorithms			
	K-means	GA-K-means	PSO-K-means	GA-IPSO-K-means
Iris	0.012395396	0.013826351	0.014528017	0.014582881
Lenses	0.339904827	0.351735427	0.360239542	0.360291735
Haberman	0.000317745	0.000328364	0.000348162	0.000363942
Balance scale	0.002573387	0.002628475	0.002810827	0.002922836
Wisconsin breast cancer	7.25935E-14	7.26287E-14	7.28928E-14	7.32617E-14
Contraceptive method choice	7.80139E-05	8.03819E-05	8.20198E-05	8.22049E-05
Hayes roth	4.59807E-05	4.70825E-05	4.73918E-05	4.74193E-05
Robot navigation	0.001583094	0.001828362	0.001898018	0.001938272
Spect heart	0.069341756	0.072648917	0.076041565	0.078285293

better than some existing methods mentioned in this work. The acceleration coefficients c_1 and c_2 are set to 1.4 for early convergence during IPSO iteration. The inertia weight is set between 1.8 and 2 for early convergence. For crossover operation in GA, two-point crossover has been used by using two random positions of the particles. If population of particle remains the same after crossover, then mutation is applied on the particles in the population for diversification of solution space. The proposed hybrid technique is able to produce good cluster center of an object. With the increase of number of iterations, the cluster center (initially chosen) will be attracted towards its corresponding similar clusters, which will lead to obtain the final cluster center with best fitness value. The change in local and global best solution will result in the updation of new position and velocity of the cluster.

5 Conclusion and Future Scope

Many researchers have successfully applied the K-means algorithm for clustering purpose. But the idea of selection of the initial clusters may vary and, hence, may lead to major/minor differences in the cluster distance. In order to overcome the drawbacks of the K-means algorithm, a hybrid evolutionary improved swarm-based K-means algorithm has been proposed in this paper to improve efficiency. In Table 2, the obtained results of the proposed GA-IPSO-K-means have been marked in boldface, which indicates the best results compared to other methods. The obtained fitness value of the clusters from the hybrid method of GA and PSO helped to get the nearer and optimal cluster centers. The proposed approach utilizes the advantage of both the optimization algorithms to get the global solution and fast convergence property of K-means method, which facilitates to avoid the limitations of both. The result of the proposed hybrid approach illustrates the better convergence with the fitness values than the other clustering techniques. But, as K-means algorithm needs the input as the number of clusters to be supplied by the user in advance and the algorithm is incapable of choosing the number of clusters, our future work may lead in that direction. Also, K-means is a spherical-based algorithm and it only performs well in a round-shaped space, for which the mean value will easily converge toward the cluster center. The future work may comprise to develop any soft clustering algorithm, which may lead to some better results in finding the optimal cluster centers as well as better fitness value.

Acknowledgments This work is supported by the Department of Science & Technology (DST), Ministry of Science & Technology, New Delhi, Govt. of India, under grants No. DST/INSPIRE Fellowship/2013/585.

References

1. Hartigan, J.A.: Clustering algorithms. 1975, Wiley, New York
2. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: A K-means clustering algorithm. *J. Royal Stat. Soc. C* **28**(1), 100–108. JSTOR 2346830 (1979)
3. Ensafi, R., Dehghanzadeh, S., Akbarzadeh, T.M.R.: Optimizing fuzzy K-means for network anomaly detection using PSO, pp. 686–693 (2008). doi: [10.1109/AICCSA.2008.4493603](https://doi.org/10.1109/AICCSA.2008.4493603)
4. Sha, M., Yang, H.: Speaker recognition based on APSO-K-means clustering algorithm. In: International Conference on Artificial Intelligence and Computational Intelligence, (2009). doi: [10.1109/AIICI.2009.17](https://doi.org/10.1109/AIICI.2009.17)
5. Kader, A.R.F.: Genetically improved PSO algorithm for efficient data clustering. In: Second International Conference on Machine Learning and Computing, 2010. doi:[10.1109/ICMLC.2010.19](https://doi.org/10.1109/ICMLC.2010.19) (2010)
6. Xiangwei, Z., Yuanjiang, J.: A study on educational data clustering approach based on improved particle swarm optimizer (2011). doi:[10.1109/ITIME.2011.6132144](https://doi.org/10.1109/ITIME.2011.6132144)
7. Youguo, L., Haiyan, W.: A clustering method based on K-means algorithm. *Phys. Procedia* **25**, 1104–1109 (2012)
8. Naik, B., Swetanisha, S., Behera, D. K., Mahapatra, S., Padhi, B.K.: Cooperative swarm based clustering algorithm based on PSO and k-means to find optimal cluster centroids. In: National Conference on Computing and Communication Systems (NCCCS), pp. 1–5 (2012). doi:[10.1109/NCCCS.2012.6413027](https://doi.org/10.1109/NCCCS.2012.6413027) (2012)
9. Govindarajan, K., Somasundaram, T.S. Vivekanandan, K.S.K.: Particle swarm optimization (PSO)-based clustering for improving the quality of learning using cloud computing. In: IEEE 13th International Conference on Advanced Learning Technologies, pp. 495–497 (2013). doi:[10.1109/ICALT.2013.160](https://doi.org/10.1109/ICALT.2013.160) (2013)
10. Liao, Q., Yang, F., Zhao, J.: An Improved parallel K-means clustering algorithm with map reduce, In: Proceedings of ICCT, pp. 764–768, (2013). doi:[10.1109/ICCT.2013.6820477](https://doi.org/10.1109/ICCT.2013.6820477)
11. Bai, L., Liang, J., Chao, S., Dang, C.: Fast global k-means clustering based on local geometrical information. *Inf. Sci.* **245**, 168–180 (2013)
12. Liao, K., Liu, G., Xiao, L., Chaoteng L.: A sample-based hierarchical adaptive K-means clustering method for large-scale video retrieval. *Knowl.-Based Syst.* **49**, 123–133 (2013)
13. Jaganathan, P., Jaiganesh, S.: An improved K-means algorithm combined with Particle Swarm Optimization approach for efficient web document clustering, pp. 772–776, (2013). doi:[10.1109/ICGCE.2013.6823538](https://doi.org/10.1109/ICGCE.2013.6823538)
14. Yao, H., Duan, Q., Li, D., Wang, L.: An improved K-means clustering algorithm for fish image segmentation. *Math. Comput. Model.* **58**, 790–798 (2013)
15. Monedero, D.R., Solé, M., Nin, J., Forné, J.: A modification of the k-means method for quasi-supervised learning. *Knowl.-Based Syst.* **37**, 176–185 (2013)
16. Shahbaba, M., Beheshti, S.: MACE-means clustering. *Sig. Process.* **105**, 216–225 (2014)
17. Naldi, M.C., Campello, R.J.G.B.: Evolutionary k-means for distributed datasets. *Neurocomputing* **127**, 30–42 (2014)
18. Scitovski, R., Sabo, K.: Analysis of the k-means algorithm in the case of data points occurring on the border of two or more clusters, *Knowl.-Based Syst.* **57**, 1–7 (2014)
19. Pavithra, M.S.: Aradhya Manjunath VN : A comprehensive of transforms, Gabor filter and k-means clustering for text detection in images and video. *Appl. Comput. Inf.* (2014). doi:[10.1016/j.aci.2014.08.001](https://doi.org/10.1016/j.aci.2014.08.001)
20. Kennedy, J., Eberhart, R.: Particle swarm optimization, In: Proceedings of the 1995 IEEE International Conference on Neural Networks, vol. 4,1942–1948 (1995)
21. Wei, J., Guangbin, L., Dong, L.: Elite particle swarm optimization with mutation. In: IEEE Asia Simulation Conference—7th International Conference on System Simulation and Scientific Computing, pp. 800–803 (2008)
22. Khare, A., Rangnekar, S.: A review of particle swarm optimization and its applications in solar photovoltaic system. *Appl. Soft Comput.* **13**, 2997–3006 (2013)

23. Babaei, M.: A general approach to approximate solutions of nonlinear differential equations using particle swarm optimization. *Appl. Soft Comput.* **13**, 3354–3365 (2013)
24. Neri, F., Mininno, E., Iacca, G.: Compact particle swarm optimization. *Inf. Sci.* **239**, 96–121 (2013)
25. Nayak, J., Naik, B., Kanungo, D.P., Behera, H.S.: An improved swarm based hybrid k-means clustering for optimal cluster centers. In: *Information Systems Design and Intelligent Applications*, pp. 545–553. Springer, (2015)
26. Yue-bo, M., Jian-hua, Z., Xu-sheng, G., Liang, Z.: Research on WNN aerodynamic modeling from flight data based on improved PSO algorithm. *Neurocomputing* **83**, 212–221 (2012)
27. Dehuri, S., Roy, R., Cho, S.B., Ghosh, A.: An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification. *J. Syst. Softw.* **85**, 1333–1345 (2012)
28. Kanungo, D.P., Naik, B., Nayak, J., Baboo, S., Behera, H.S.: An improved PSO based back propagation learning—MLP(IPSO-BP-MLP) for classification. *Smart Innovation, Systems and Technologies* 31, doi:[10.1007/978-81-322-2205-7_32](https://doi.org/10.1007/978-81-322-2205-7_32) (2015)
29. Bache, K., Lichman, M.: UCI machine learning repository [<http://archive.ics.uci.edu/ml>], Irvine, University of California, School of Information and Computer Science (2013)

Analysis of Energy Efficient, LEACH-Based Cooperative Wireless Sensor Network

V. Murali and Sindhu Hak Gupta

Abstract Energy efficient transmission is the foundation for energy efficient communication. Similarly, for (Wireless Sensor Network) WSN energy efficiency is of prime importance because of its constrained size that results in constrained battery. The transmission medium being wireless, fading plays a major role on the performance of WSN. It deteriorates the overall performance which leads to indirect consumption of battery of a sensor node. Cooperative communication combats fading and saves energy of a network. In this paper application of cooperative communication has been implemented and analyzed. Experimental results show that when cooperation is there the total energy consumed by the WSN is much less than the energy consumed by the WSN without cooperation. Cooperation makes WSN more energy efficient. It is a known fact that wireless transmission is energy consuming operation. LEACH is considered as most popular routing protocol which has better performance in saving the energy consumption. It is implemented in cooperative WSN and experimental results show that LEACH makes cooperative WSN more energy efficient.

Keywords LEACH · Cooperative communication · WSN · Energy efficiency

1 Introduction

Wireless sensor network (WSN) is a distributed network which comprises of a very large number of sensor nodes wirelessly positioned over a certain geographical area. WSN has been made possible by the collaboration of microelectromechanical

V. Murali (✉) · S.H. Gupta
Amity University, Noida Sec-125, 201301, U.P, India
e-mail: murali.prince08@gmail.com

S.H. Gupta
e-mail: shak@amity.edu

systems technology, digital and wireless communications, and digital electronics [1]. These sensor nodes comprise of RF trans-receivers, actuators, power supply, CMOS building blocks such as microprocessors, microcontrollers, interfacing circuitry, specialized and general purpose signal processing engines, etc. [2]. They are randomly deployed in a huge number, over a certain geographical area. They need to be tough enough to withstand various harsh conditions and particular environmental and geographical constraints. Limited physical size of these nodes makes them to have limited battery energy supply and hence limited net energy. Since they are randomly and abundantly deployed, it is not possible to recharge or replace their batteries easily [3]. There is a constant need to devise efficient ways which can take care of their energy issues. The main focus needs to be kept on the performance of WSN. Wireless sensor nodes, in order to save energy consumption, periodically self-organize themselves into clusters. Among group of sensors forming a cluster, one sensor is selected as a cluster head. The cluster head organizes the cluster, collects and aggregates data within the cluster, and finally transmits the aggregated data to the sink. Low energy adaptive clustering hierarchy (LEACH) [4] is such a protocol which achieves all this. It has been shown in [4] that LEACH is an energy efficient protocol and saves energy, which in turn saves the battery life of a WSN. The major portion of the total battery of a sensor network is consumed in the trans-receiving and processing section. For transmission and reception of data the medium in this case is wireless. One of the factors which adversely deteriorates the performance of wireless communication is fading. So, it can be summed up that fading has an indirect effect on the battery consumption in a sensor node.

On the other hand, cooperative communication appears as a potential candidate to combat the adverse effects of fading on wireless communication. At the same time it increases the transmit energy efficiency [5]. Apart from these advantages, cooperative communication provides other various major advantages such as spatial diversity and data throughput; although these advantages are provided by (Multiple-input multiple-output) MIMO systems also. In this scenario of WSN, it becomes difficult to implement MIMO, as it will be very difficult for a sizewise constrained sensor node to support multiple transmitters and receivers [6]. Cooperative communication helps in creation of Virtual MIMO in a WSN. It enables single antenna wireless sensor node into multiuser environment. This is made possible as sensor nodes share their antennas and achieve transmit diversity. The fundamental concept behind cooperative communication was in the wake of information theoretic properties of the relay channel [7, 8]. Amplify and forward, decode and forward, and coded cooperative communication are the few cooperative protocols [9, 10]. Since we get all the benefits of MIMO from cooperative communication, we can term it as Cooperative MIMO (C-MIMO) or Virtual MIMO (V-MIMO). For our representation in further research in this paper we will use the term (Cooperation in WSN) C-WSN. In this paper, we are implementing cooperative communication in WSN and are observing the results. We are also implementing LEACH in C-MIMO.

The main contribution of this paper is that:

1. It is shown that C-WSN makes it more energy efficient in comparison to the WSN without cooperation. It works on the principle of Single-input single-output (SISO). It is termed as SISO-WSN.
2. Total energy consumption of C-WSN is much less in comparison to SISO-WSN
3. C-WSN with LEACH is more energy efficient than C-WSN without LEACH

2 System Model and Energy Model

Figure 1 depicts a WSN. This network consists of N number of wireless sensor nodes. These sensor nodes are randomly deployed over a geographical area of $(M \times M)$ dimensions. All the N nodes of the network have a specific lifetime. During the specified lifetime the energy possessed by the nodes is utilized for gathering, processing, and transmitting information. All sensor nodes are assumed to be synchronized by the sink before the start of the operation. The sinks will cooperate among themselves to form a V-MIMO or C-WSN. Each round will consist of three stages: cluster formation stage, steady state stage, and cooperative communication transmission stage.

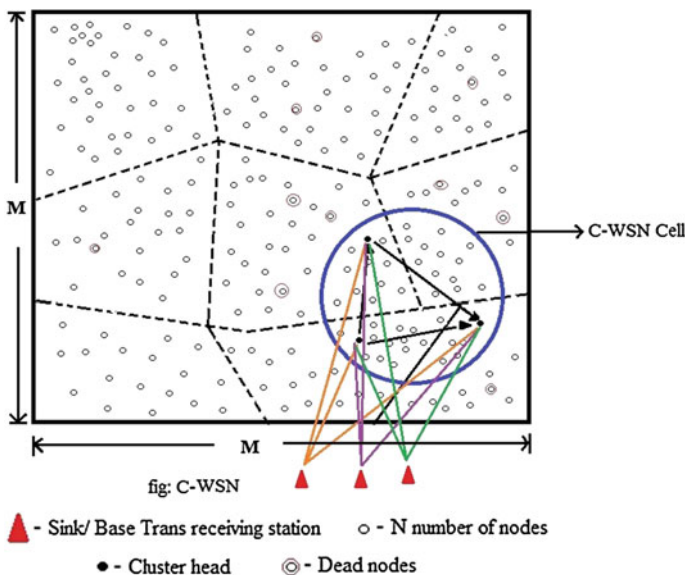


Fig. 1 Clustering in WSN

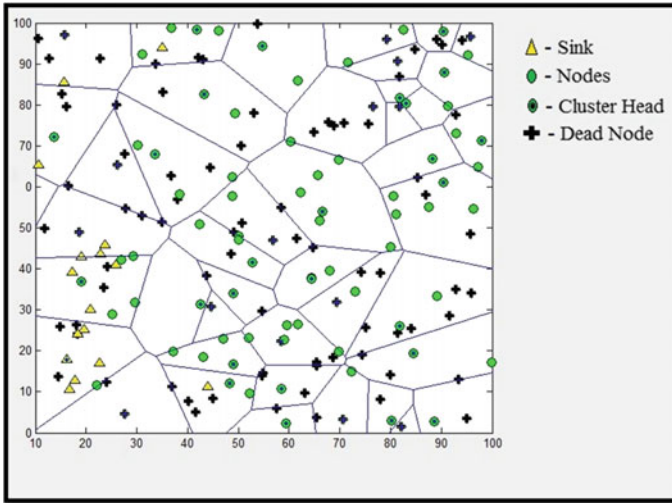


Fig. 2 Implementation of LEACH

Figure 2 depicts the capture of implementation of LEACH in WSN. It shows the random deployment of nodes, portioning, clustering, and formation of cluster heads. In this $P = 0.04$ has been considered. This indicates the probability of a node to become cluster head. Dead nodes, active nodes, sinks, etc. are clearly indicated in Fig. 2.

3 Cluster Formation Stage

There is no centralized control in LEACH. Each node using a distributed algorithm takes an autonomous decision and becomes part of a cluster. All the nodes of the cluster must become cluster heads same number of time (approximately) so that net lifetime of each node is approximately same. If any node is repeatedly chosen as cluster head, its net lifetime will be reduced drastically. It is assumed that all nodes will commence their operation with the equal quantity of energy [11]. Distribution of the cluster heads should be uniform throughout the network. Once the nodes have elected cluster head, the nominated cluster head will broadcast an advertisement message. This advertisement message is comprised of two parts. One part is the nodes ID and the other part is a header, whose task is to distinguish this broadcasted message as an announcement message. Each node will receive this broadcast message. Noncluster head nodes will determine to which cluster they belong based on signal strength of the advertisement message they received. Once each node has decided its cluster head, it must inform the respective head that it is

part of that cluster. Join request message is transmitted by each node to the nearest chosen cluster head. Once the cluster is formed, the cluster head node will act as a local control center and will control data transmission in the respective cluster [5].

4 Steady State Stage

The number of nodes per cluster is not uniform and evenly distributed in LEACH. The number of the nodes in the cluster determines the amount of the data which every node individually can remit to the cluster head. In cluster formation stage, the head sets up the TDMA schedule for cluster members, which via broadcast message is transmitted to each node. In steady state phase, data will be transmitted by each member to the head by multiple frames. In LEACH it is made sure that in each frame, data is transmitted by each cluster member in each frame, in its allocated transmission slot specified at cluster formation stage, in all other slots nodes will sleep to save the energy. The trans-receivers of the cluster head should be in the active mode throughout the operational time so that it is able to receive data from all the neighboring nodes. Data aggregation is done by the cluster heads. Data redundancy is removed by cluster heads via data aggregation.

5 Cooperative Transmission Stage

Once data aggregation is done by the cluster heads, they wait for the polling message which is broadcasted by the sink. This sink usually is a base trans-receiving station (BTS). When broadcast message transmitted by the sink is received by cluster heads, they reply back with an answer message. This answer message indicates the identity and location of each cluster head. Based on this retrieved information the sink will form groups of these cluster heads. Members of each group, which comprise of these cluster heads will cooperate among themselves to achieve and implement V-MIMO. Hence, a V-MIMO cell will be formed. As is indicated in Fig. 2. Sink will set up a TDMA schedule for Cooperative WSN (C-WSN) cell [12]. Now, cluster heads among a group which are cooperating among themselves will form a cooperative network. These cluster heads will cooperate among themselves to achieve V-MIMO. Which cluster head will cooperate with whom is executed with the help of relay channel algorithm. Two cluster heads will cooperate among themselves if they can decode each other's message properly. BER performance of each cluster head plays an important role. The aim while choosing a cooperating cluster head partner is that it should be optimal and should have a good BER. They will be implementing coded cooperation [13]. Sink/Base trans-receiving sets up the TDMA schedule. According to the schedule it is decided which C-WSN cell will transmit first, which will transmit second, and so on [12].

6 Energy Consumption Model of the Proposed Scheme

In this section the energy consumption model for the C-WSN scheme is derived before modeling the energy consumption of the proposed network. We will first model as described in [4] the energy consumption of one transmitting or receiving bit first.

Let P_T = Total power consumption of transmitting. P_T is comprised of two parts, first is the power consumed by all power amplifiers, i.e., P_{POA} and the second is the power consumed by all other circuit blocks. It is designated as P_{cb} .

$$\text{i.e., } P_T = P_{POA} + P_{cb} \quad (1)$$

The power consumption of all the power amplifiers (P_{POA}) can be calculated by the following equation [4]:

$$P_{POA} = (1 + \alpha) P_{out} \quad (2)$$

If de = Drain efficiency &
 ε = Peak average ratio [14]

$$\alpha = \left(\frac{\varepsilon}{de} - 1 \right) \quad (3)$$

R_b = Transmission bit rate
 ε_b = Average energy per bit [for a given BER specification]
 d = Transmission distance
 G_t = Antenna gain (Transmitter)
 G_r = Antenna gain (Receiver)
 λ = Wavelength of the carrier
 N_f = Noise figure of receiver
 M_1 = Link margin

Then [14]

$$P_{out} = (\varepsilon_b R_b) * \frac{(4\pi)^2 d^k}{G_T G_r \lambda^2} M_1 N_f \dots \quad (4)$$

$$N_f = \frac{N_r}{N_o} \quad (5)$$

N_r = Power spectral density of total effective noise at receiver i/p
 N_o = Power spectral density of total effective noise at room temperature

The total energy consumption per bit is

$$\varepsilon_{bt} = \frac{(P_{POA} + P_{CB})}{R_{ber}} \quad (6)$$

$$R_{ber} = \text{Actual bit error rate.} \quad (7)$$

The total energy consumption is estimated by multiplying ε_{bt} by the number of bits L to be transmitted. The value of R_{ber} in equation is replaced by parameters given in [15]. Now, we will be developing mathematical model for energy consumption in cooperative communication case [16].

$$\begin{aligned} E_{VM} = & \sum_{i=1}^{N_{tx}-1} L_i E_i^{tr} + E_{dip} \sum_{i=1}^{N_{tx}} L_i + E_{enec} \sum_{i=1}^{N_{tx}} \frac{L_i}{r} \gamma_i + (N_{tx} - 1) \sum_{i=1}^{N_{tx}} \frac{L_i}{r} \gamma_i \\ & + \sum_M^L \sum_{i=1}^{N_{tx}} \frac{L_i}{r} \gamma_i + \frac{1}{b_{mimo}} \sum_{i=1}^{N_{tx}} \sum_{j=1}^{N_{rx}} \frac{L_i}{r} \gamma_i b_l E_j^t \end{aligned} \quad (8)$$

In SISO approach, sensors transmit their data to the cluster head and cluster head transmit all their aggregated data to the destination node without any cooperation. Therefore,

$$E_{SISO} = \sum_{i=1}^{N_{tx}-1} L_i E_i^t + E_{da} \sum_{i=1}^{N_{tx}} L_i + E_{enco} \sum_{i=1}^{N_{tx}} \frac{L_i}{r} \gamma_i + E_s^l \sum_{i=1}^{N_t} \frac{L_i}{r} \gamma_i \quad (9)$$

E_s^l = SISO long have transmission, which can be calculated as a special case of MIMO with $N_{tx} = N_{RX} = 1$ [16].

7 Results & Discussions

Figure 3 shows the comparative total energy consumption w.r.t distance. It is observed that as the distance in meters increases the energy consumption is least in C-WSN (i.e., V-MIMO) in comparison to SISO-WSN. It is observed that at a distance of 40 m the energy consumption in case of SISO-WSN is 1.3 J, whereas in cooperative WSN it is 65 J. A detailed analysis of energy consumption is given in Table 1.

Figure 4 shows the comparative energy efficiency w.r.t distance. It is observed that as the distance in meters increases there is variation in energy efficiency. At a distance of 40 m, C-WSN is most energy efficient. Its energy efficiency is 20 % in comparison to SISO-WSN which is 0 %. Detailed analysis of energy efficiency w.r.t distance for the cases, SISO-WSN and C-WSN is given in Table 2.

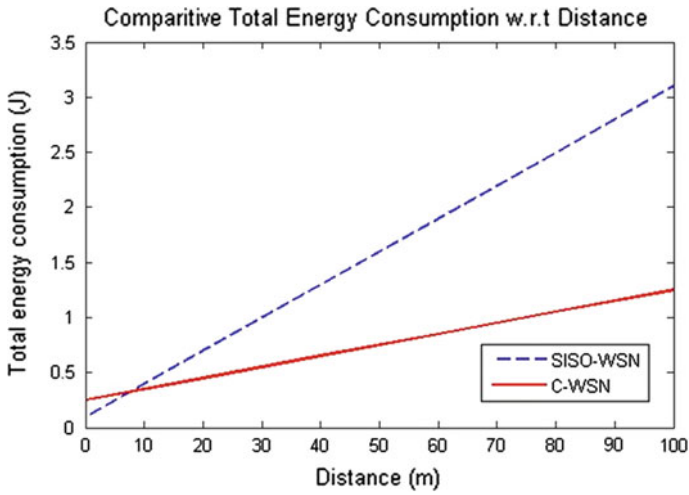


Fig. 3 Comparitive energy consumption w.r.t distance

Table 1 Comparative energy efficiency w.r.t distance

Distance (m)	Total energy consumption (J)	
	SISO-WSN	C-WSN
20	0.7	0.45
40	1.3	0.65
60	1.9	0.85
80	2.5	1.05
100	3.1	1.25

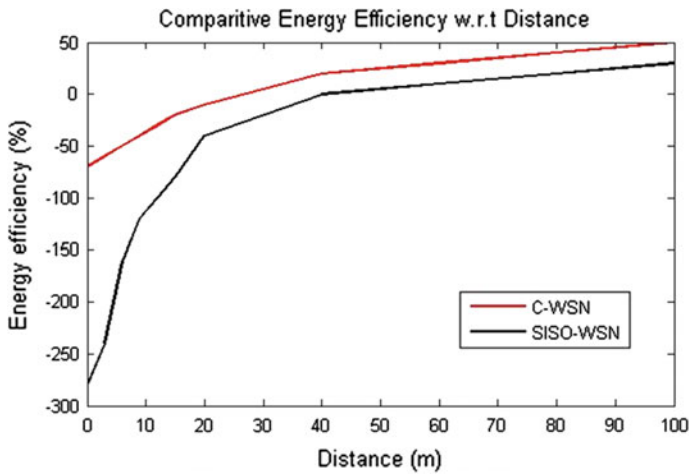


Fig. 4 Comparitive energy consumption w.r.t distance

Table 2 Comparative energy efficiency w.r.t distance

Distance (m)	Energy efficiency (%)	
	SISO-WSN	C-WSN
20	-40	-10
40	0	20
60	10	30
80	20	40
100	30	50

Fig. 5 Implementation of LEACH

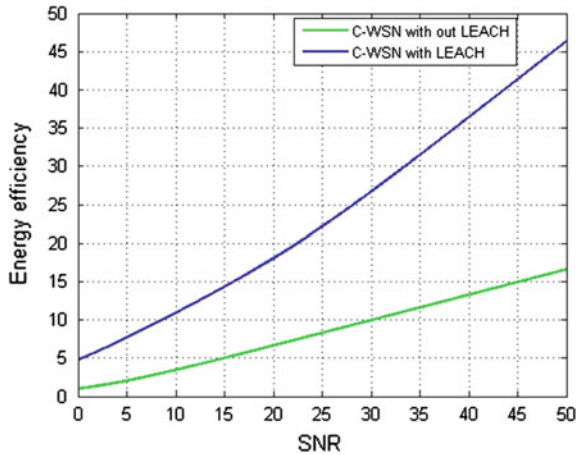


Table 3 Energy efficiency w.r.t SNR

SNR (db)	Energy efficiency (%)	
	C-WSN without LEACH	C-WSN with LEACH
10	3.45	23.59
20	6.65	41.86
30	9.96	61.53
40	13.29	81.64
50	16.61	101.4

Figure 5 shows the effect of implementation of LEACH in C-WSN. It is observed that LEACH makes C-WSN about 84 % more energy efficient. At an SNR of 40 db the energy efficiency of C-WSN without LEACH is 81.64 %. Detailed energy efficiency (%) of C-WSN without LEACH, and C-WSN with LEACH are given in Table 3.

8 Conclusion

One of the most important factors for WSN is energy. Fading degrades the performance of WSN to a huge extent. Cooperative communication combats fading and saves the energy of a network. In this paper, it is observed that when cooperative communication is implemented in WSN, it becomes C-WSN. C-WSN becomes more energy efficient as it is proven from the simulation results that total energy consumption for the cooperative WSN gets decreased by approximately 35 % in comparison to SISO-WSN. It is also analyzed that cooperative WSN is approximately 20 % more energy efficient than SISO-WSN. Further consider that LEACH is one of the most famous clustering mechanisms. It is implemented in cooperative WSN and it is observed that cooperative WSN with LEACH is approximately 83 % more energy efficient in comparison to C-WSN without LEACH.

References

1. Ian, F.A., Wellian, S., Yogesh, S., Edral, C.: Wireless sensor networks: a survey. *Comput. Netw.* **38**(4), 393–422 (2002)
2. Pottie, G.J.: Wireless Sensor Networks, Information Theory Workshop. In: IEEE 1998, pp. 139–140
3. Gungor, V., Hancke, G.P.: Industrial sensor networks challenges, design principles, and technical approaches. *IEEE Trans. Industr. Electron.* **56**(10), 4258–4265 (2009)
4. Cui, S., Goldsmith, A.J., Bahai, A.: Energy efficiency of MIMO and cooperative MIMO techniques in sensor networks. *IEEE J. Sel. Areas Commun.* **22**, 1089–1098 (2004)
5. Sadek, A.K., Yu, W., Ray Liu, K.J.: On the Energy Efficiency of Cooperative Communications in wireless sensor networks. *ACM Trans. Sens. Netw.* **6**(1), 5:1–5:21 (2009)
6. Gupta, S.H., Singh, R.K., Sharan, S.N.: Performance evaluation of efficient cooperative MIMO system wireless communication. *WSEAS Trans. Commun.* **13**, 9–14
7. Cover, S.N., Gamal, A.E.: Capacity theorems for the relay channel. *IEEE. Trans. Inform. Theor.* **25**(5), 572–84 (1979)
8. Sendonaris, A., Erkip, E., Aazhang, B.: User cooperation diversity Part I and Part II. *IEEE Trans. Commun.* **51**(11), 1927–1948 (2003)
9. Laneman, J.N., Wornell, G.W., Tse, D.N.C.: An Efficient protocol for realizing cooperative diversity in wireless Networks. In: Proceedings of the IEEE ISIT, Washington DC, pp. 294 June 2001
10. Hunter, T.E., Nosratinia, A.: Cooperative Diversity through coding. In Proceedings of the IEEE ISIT, Washington, DC, p. 220 (2001)
11. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **1**, pp. 660–670 (2002)
12. Ding, J., Liu, D., Wang, X., Wu, H.: An energy efficient virtual MIMO transmission scheme for cluster based wireless sensor networks. In: IEEE ICCT, pp. 809–813 (2010)
13. Nosratinia, A., Hunter, T.E.: Cooperative communication in wireless networks. pp 74–80 (2004)

14. Tan, T.K., Raghunathan, A., Lakshminarayan, G., Jha N.K.: High Level Software Energy Macro-Modeling. In Proceedings of Design Automation Conference. Las Vegas, NV, USA, pp. 605–610 June 2001
15. Rappaport, T.S.: Wireless Communications, Principles and Practices, 2nd edn. Prentice Hall: Upper Saddle River, NJ, USA 2002
16. Mohammad, R.I., Young, S.H.: Cooperative MIMO communication at wireless sensor network: an error correcting code approach. *Sensors* **11**(10), 9887–9903 (2011)

A Monotonic Digitally Controlled Delay Element-Based Programmable Trigger Pulse Generator

Amit Krishna Dwivedi, Manisha Guduri, Rishab Mehra
and Aminul Islam

Abstract This paper presents a resourceful utilization of a monotonic digitally controlled delay element (DCDE) to propose a programmable high frequency trigger pulse generator circuit (TPG). Performance evaluation of various analog and digital programmable delay elements (DEs) have been carried out to reach the conclusions presented. Further, this work exploits a monotonic DCDE along with an efficient XOR circuitry, to realize the proposed TPG. The proposed design generates a very high frequency ultra-thin pulses of pulse duration ranging from 56 to 170 ps for digital input vector ranging from '00000' to '11111', respectively. The proposed design has been extensively verified using SPICE @ 16-nm predictive technology model.

Keywords Programmable delay circuits · Trigger pulse generator · Digitally controlled delay element

1 Introduction

Low supply operated circuits have become a topic of much interest due to its potential applications in portable gadgets. Various circuits require a high threshold pulse to start their operation. This triggering of circuits requires large power

A.K. Dwivedi (✉) · M. Guduri · R. Mehra · A. Islam
Department of Electronics and Communication Engineering,
Birla Institute of Technology, Mesra, Ranchi 835215, Jharkhand, India
e-mail: amit10011.13@bitmesra.ac.in

M. Guduri
e-mail: manishaguduri@bitmesra.ac.in

R. Mehra
e-mail: rishabmehra03@gmail.com

A. Islam
e-mail: aminulislam@bitmesra.ac.in

consuming circuitry designed especially to initiate the switching operation. Devices, such as SCR, thyristor, etc., require a firing circuit implemented in the same IC to execute their operation. This requires a compact trigger pulse generator (TPG) circuit.

Conventional TPG employs RC circuits in their architecture, which are bulky to be incorporated in the nanoscale circuit designs. These designs are also affected by the time constant (numerical values of R and C). In other methods of generating such high frequency microwave pulses, step recovery diode (SRD) or nonlinear transmission line (NLTL) are utilized [1]. These pulse generators are inexpensive but they require several volts of input drive. Moreover, they exhibit moderately high level of random jitters and fluctuation in pulse width due to the PVT (process, voltage, and temperature) variations. In other 42Ts design mentioned in [2] utilizes a buffer delay circuit, which increases the complexity. Similar delay-based design approach mentioned in [3] requires CNFETs to implement the design. This design provides precise control over delay, however, the technology integration and large number of transistors makes the design complex. Moreover, CNFET-based designs suffer from CNT-specific imperfections. It is also difficult to grow single chirality tubes as only few techniques have been developed to short semiconducting tube.

All the abovementioned designs of TPG circuits are non-programmable; means only fixed duration of trigger pulse can be produced. As the duration of desired trigger pulse also varies for specific applications, it demands to implement a variable pulse duration TPG circuit to increase its versatility. For example, circuits, such as intensifiers, lasers, optical switches, modulators, optical components, and so on, require wider pulse (in range of nanoseconds) as compared to a clock or stimulus (in range of picoseconds) which can be injected into a device under test (DUT). Thus, a variable width TPG with compact size and low power consumption is desirable. In view of above, this paper makes following contributions:

- (1) Various programmable delay elements (DEs) are investigated.
- (2) Monotonic digitally controlled delay element (DCDE) is utilized to develop proposed high frequency variable pulse generator circuit.

The rest of the paper is organized as follows. Section 2 investigates various programmable DEs. Section 3 proposes a proficient use of monotonic DCDE as a programmable high frequency TPG. Simulation results and discussions are mentioned in Sect. 4. Finally, the concluding remarks are provided in Sect. 5.

2 Various Programmable Delay Elements

The delay introduced by programmable DEs can be manipulated by external control. The control can be of analog or digital type. Thus, programmable DEs can be investigated under two classes.

2.1 Analog Programmable Delay Elements (APDEs)

Cascading even number of inverter circuits is a simplest way to incorporate the delay between the two signals. However, the amount of delay introduced by such delay elements can be controlled using an external control [4]. The dependence of the introduced delay (t_d) by analog voltage controlled delay elements (VCDEs) can be defined with following relationship:

$$t_d = \frac{C_L}{I_{C/D}} V_{FSW}. \tag{1}$$

where, C_L is the load capacitance, $I_{C/D}$ indicates the charging/discharging current of the load capacitance, and V_{FSW} is the full swing voltage of the inverter circuit. From the equation, it can be observed that desired delay can be achieved by controlling $I_{C/D}$ through the load capacitance. Linear relation exists between the delay introduced and the V_{FSW} or C_L . Thus, variable C_L is required to linearly control the delay amount. However, variable capacitor is realized as inversely polarized PN junction and its value dominantly depends on the inverse voltage [5]. This nonlinear dependence of variable capacitance makes its realization challenging. Another method to provide variable delay is to control the drain current of the pull-up or pull-down or both the network using an external control mechanism. Such delay circuits come under family of voltage controlled DEs that can be classified as— n -voltage controlled, p -voltage controlled, and np -voltage controlled DEs [6]. The circuit-level models of these DEs are shown in Fig. 1 [4–6].

Figure 1a shows NMOSs (MN3/MN4) connected in series to the source terminals of the transistors MN2/MN1, respectively, to act as a pull-down network. Further, an external control V_n is provided to the gate terminal of the transistors

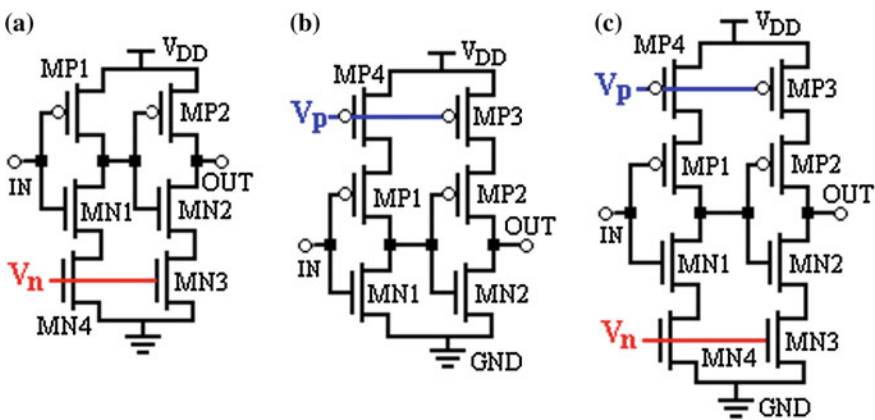


Fig. 1 Circuit diagram of **a** n -voltage controlled, **b** p -voltage controlled and **c** np -voltage controlled deadly elements

Table 1 Design metrics of analog voltage controlled delay ELEMENTS @ 700 mV supply for 16 nm technology node

S. no.	Delay element	No. of transistors	Minimum delay $\times 10^{-10}$	Power (W) $\times 10^{-8}$	Delay variability (a.u.) (σ/μ)
1	<i>n</i> -Voltage controlled	6Ts	1.7438	154.85	0.256
2	<i>p</i> -Voltage controlled	6Ts	2.296	182.57	0.582
3	np-VC	8Ts	2.103	209.16	0.656

MN4/MN3 to control the pull-down current. However, *n*-voltage controlled DE shows its effectiveness only when transition from high to low occurs. Figure 1b utilizes V_p as a control voltage for the pull-up drain current through the transistors MP3/MP4. Such type of control mechanism is restricted only for low to high transitions as it comes into action only for pull-up switch network. Thus, to provide control over both low to high or high to low transitions, both *n*-voltage control and *p*-voltage control are applied in a single architecture. Figure 1c shows the circuit-level model of the np-voltage control DE. Here, V_n controls the pull-down network; meanwhile inverted control signal V_p controls the pull-up network. Thus, np-VCDEs provide more precise control over the amount of delay introduced. Table 1 summarizes the analysis in terms of various design metrics of the DEs presented in Fig. 1. The *n*-voltage control and *p*-voltage control delay elements give considerable amount of delay but effectiveness of control mechanism is restricted to high to low and low to high transitions, respectively. This can be further eliminated in np-voltage control circuits. However, such circuits still lack the signal integrity, which is required while processing signal. Schmitt trigger is introduced in np-voltage controlled DEs to provide signals with larger hysteresis, but at the cost of more power consumption due the increase number of transistors [6].

In analog VCDEs, the charging/discharging current (I_{CD}) at the load capacitance (C_L) varies due to geometric imperfections. This results in the nonidentical rise time and fall time of the output signal. Apart from this, the analog VCDEs require ADC to make it compatible to the digital environment. As most of the circuits are progressing toward digital domain, demand for digital control DEs are increasing.

2.2 Digital Programmable Delay Elements (DPDEs)

As in analog control mechanism where either voltage or current is utilized to control the variable amount of delay, similarly, DCDEs utilize discrete voltage or capacitance to control the delay amount, digitally [7]. In DCDEs, an NMOS switch network is placed at the source terminal of NMOS (MN1) of the CMOS inverter as shown in Fig. 2a. This shows effectiveness in the falling edge of the output signal. Further, to control the rising edge of the output signal, PMOS switch network is

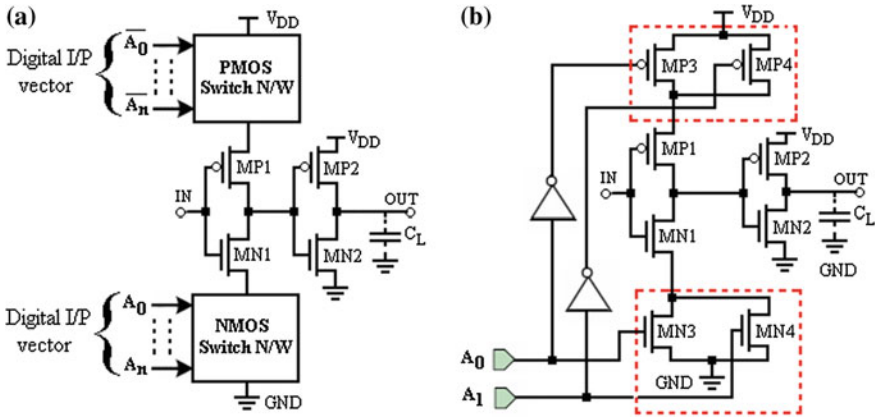


Fig. 2 a Basic structure of digital controlled delay elements and b 2-I/P vector programmable delay element

introduced at the source terminal of the PMOS MP1 of the inverter as shown in Fig. 2a.

However, the number of transistors comprising the NMOS/PMOS switch networks depends upon the desired resolution and delay to be introduced. A stack of $n \times m$ NMOS/PMOS transistors can be incorporated in the design to control the charging/discharging of the C_L and hence, the delay. The equivalent resistance (R_{eq}) offered by the NMOS/PMOS networks constitute the delay introduced which can be further varied based on the digital input vectors [7]. The basic architecture of DCDE for 2-I/P vector is demonstrated in Fig. 2b. This work aims to present an input vector control-based pulse generator circuit. To achieve this task, monotonic DCDE-based mechanism has been employed in design presented. A 5-binary input (abcde) control modified circuit-level model of monotonic DCDE is demonstrated in Fig. 3. Binary input vector (abcde) controls the switching characteristics of PMOSs (MP2-MP6). This further controls the current (I_D) through the MOS (MN6). As the drain current of MN6 is mirrored to the drain terminal of MN2 (because MN6-MN2 form a current mirror), same current (I_D) is obtained through the MN2. Now, MN2 provide a pull-down path to MN1. The delay introduced by the circuit shown in Fig. 3 depends upon I_D . Further, MP1 and MN2 control the rise time and fall time of the inverter composed of MN1/MP1, which finally controls the delay. The output of the inverter composed of MN1/MP1 is further passed through cascaded inverter stages to provide the delayed output signal at node OUT.

This architecture provides variability in the incorporated delay with the incremental delay. To achieve this, delay controlling PMOSs are sized in the binary fashion such that W/L of MP3 is twice that of W/L of MP2, and so on. Sizing of transistors (MP2-MP6) is tabulated in Table 2. The architecture can be utilized to implement the 2^N binary inputs, where N is the number of PMOS transistors. Such types of circuit are complex due to increase number of transistors and consumed

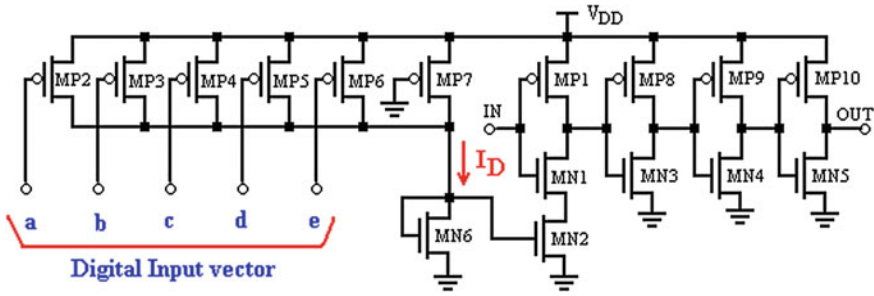


Fig. 3 Circuit-level model of monotonic digital control delay element (DCDE)

Table 2 Transistor sizing for the Presented DCDEs

Aspect ratio (nm)	MP2	MP3	MP4	MP5	MP6
$\frac{W}{L}$	$\frac{48}{16}$	$2 \times \frac{48}{16}$	$4 \times \frac{48}{16}$	$8 \times \frac{48}{16}$	$16 \times \frac{48}{16}$

Table 3 Performance metrics of the 5-I/P DCDEs

Delay element	Input vector	Minimum values of delay (s) $\times 10^{-12}$	Power consumption (Watts) $\times 10^{-6}$
5-I/P DCDE	00000	56	12.89
	11111	170	29.7

more power per unit transistor as compared to other DEs. However, the delay introduced by the DCDE depends upon the digital input vectors. This makes DCDEs a deserving candidate for precise delay introducing circuit to realize the proposed TPG. Table 3 summarizes the parametric performance of the DCDEs for five-binary I/P.

3 Proposed Programmable Trigger Pulse Generator Circuit

Programmable DEs often lacks uniform increment introduced delay based upon the incremental input vector. Hence, this paper utilizes the modified design of DCDE presented in [7] to propose an input vector controlled TPG circuit. Figure 4 illustrates the state-of-the-art of the TPG. Simple XORing of the two signals having certain amount of time delay existing between them can be utilized to obtain high frequency microwave pulses. The circuit-level model of the proposed TPG circuit realized using DCDE is presented in Fig. 5. It can be observed from the Fig. 4, that

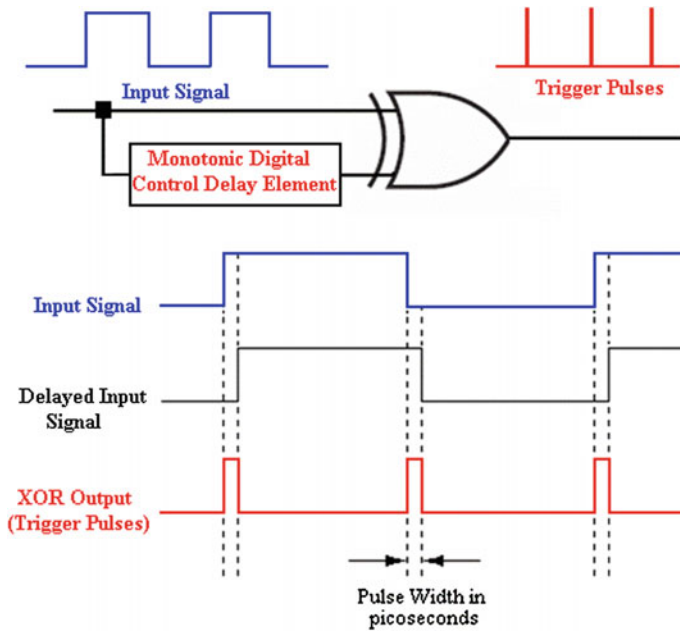


Fig. 4 Principle of generation of trigger pulses from the proposed buffered delay circuit

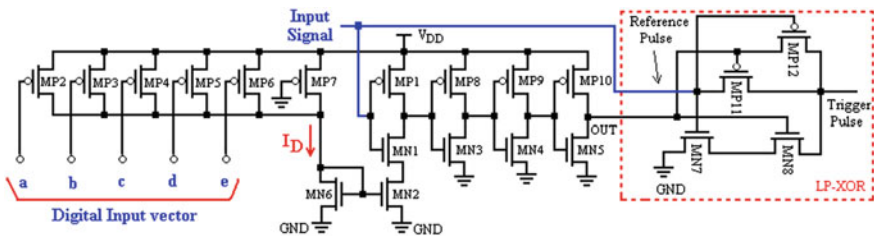


Fig. 5 Circuit-level model of the proposed monotonic digital controlled variable pulse width trigger pulse generator using LP-XOR

the proposed pulse generator requires a DE along with a XOR logic circuit. The monotonic DCDE is employed as a DE emerged from the analysis carried in the previous section. Apart from this, to perform the XOR operation of the reference signal and its delayed version, an efficient XOR circuit is required. This paper utilizes LP—XOR circuit presented in [8] to implement the proposed design, which consumes a small amount of power (1.096 nW) to perform the XOR operation. As the introduced delay is in the order of picoseconds (56–170 ps), the XORed output produces pulses with the pulse duration in picoseconds. Pulse generators capable of

generating pulses with widths approximately 100 picoseconds are often termed as microwave pulse generator. These pulse generators find application in various testing and fault finding circuits.

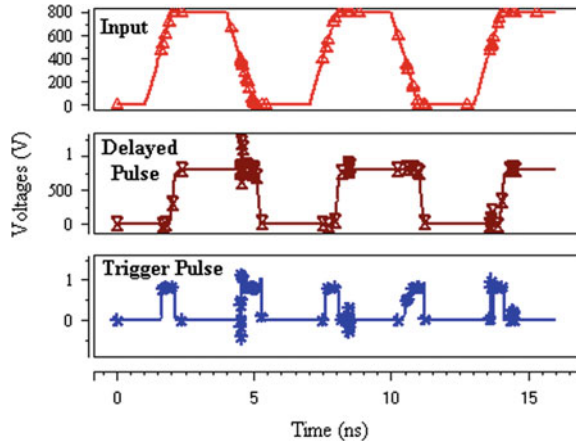
4 Simulation Results and Discussion

To verify the circuit-level model extensive simulation of delay circuits are performed on SPICE using 16-nm Predictive Technology Model (PTM) @ 0.7 V and $T = 27^\circ\text{C}$ [9]. Both APDEs and DPDEs have been studied in this paper. Design metrics such as minimum delay, delay variability, power consumption, and number of transistors utilized have been studied and summarized in Table 1. Further, for the five-I/P DCDE, the t_d varies from 56 to 170 ps and power consumption varies from 12.89 to 29.7 μW for input vector ‘00000’ – ‘11111’, respectively. The delay variability have been analyzed by varying device parameters such as channel length (L), gate width (W), channel doping concentration (N_{DEP}), oxide thickness (t_{OX}), threshold voltage (V_t), and carrier mobility (μ_0) up to $\pm 10\%$. For higher accuracy, 5000 samples of the delay circuits have been simulated using Monte Carlo simulations that take into account variability of device and environmental parameters. The simulation results obtained for the proposed TPG @ 16-nm technology node are demonstrated in the Fig. 6. Table 4 mentions a comparative study of the similar TPG circuits.

Table 4 Comparison with existing TPG Circuits @ 700 mV supply

Delay element	Input control	Minimum delay (s)	Remarks	Reference
5-I/P DCDE	00000–11111	$56\text{--}170 \times 10^{-12}$	Programmable	In this work
NLTL	NA	17×10^{-9}	1. Fixed duration 2. Jitters and fluctuation due to PVT variations.	[1]
RC-based TPG	NA	Depends upon time constant	1. Fixed duration 2. Bulky RC circuitry	[10]
CNFET-based TPG	NA	88.6×10^{-12}	1. Fixed duration 2. Suffers from CNT-specific imperfections	[3]
MOS-based TPG	NA	163.2×10^{-12}	1. Fixed duration 2. Complex circuitry (42Ts)	[2]

Fig. 6 Generation of trigger pulses from the proposed circuit-level model of pulse generator for input vector '11111'



5 Conclusion

As the resolution of the obtained pulse depends upon the preciseness of the introduced delay, various APDEs and DPDEs have been studied to extract an appropriate delay element for the proposed high frequency TPG. Series connected transistors whose gate voltages can be varied to regulate the current have been presented as analog VCDEs to incorporate delay. Further, DPDEs have been also presented whose delay can be controlled by digital input vector and modifies monotonically. The functionality of the monotonic DCDE is verified by utilizing it to develop a TPG circuit along with LP-XOR. The proposed TPG circuit proves its versatility by producing ultra-thin pulses of variable pulse duration. The pulse duration can be controlled monotonically by the various logic combination of digital input vector. The comparison with the existing methodologies has been also presented to prove the importance of the design presented.

References

1. Salameh, D., Linton, D.: Novel Wide Bandwidth GAAS Sampling MMIC using Microstrip Based Nonlinear Transmission Line (NLTL) and NLTL Shock Wave Generator Design, IEEE 28th European Microwave Conference, pp. 18–23 (1998)
2. Dwivedi, A.K., Urma, K.A., Islam, A.: Trigger pulse generator using proposed buffered delay model and its application. *Act. Passive Electron. Compon.* Article ID 920508, 9 p, 2015. doi:[10.1155/2015/920508](https://doi.org/10.1155/2015/920508)
3. Dwivedi, A.K., Urma, K.A., Kumar, A., Islam, A.: Robust Design of CNFET Based Buffered Delay Model and Microwave Pulse Generator, In: IEEE International Conference on Devices, Circuits and Communications (ICDCCom—2014), BIT, Mesra, Ranchi, India, September 12–13, 2014

4. Mahapatra, N.R., Tareen, A., Garimella, S.V.: Comparison and analysis of delay elements, In: The 2002 45th Midwest Symposium on Circuits and Systems, IEEE press, vol. 02, pp. II-473-II-476, Aug (2002)
5. Rabaey, J.M.: Digital integrated circuits: A design perspective, Prentice-Hall Book Company, 1st edn, ISBN #0-13-178609-1. Upper Saddle River, NJ (1996)
6. Mahapatra, N.R., Garimella, S.V., Tareen, A.: Efficient techniques based on gate triggering for designing static CMOS ICs with very low glitch power dissipation, In: Proceedings of the 2000 IEEE International Symposium on Circuits and Systems, vol. 2, pp. 537–540, Geneva, Switzerland, May 28–31 (2000)
7. Maymandi-Nejad, M., Sachdev, M.: A monotonic digitally controlled delay element. *IEEE J. Solid-State Circ.* **40**(11), 2212–2219 (2005)
8. Srivastava, P., Dwivedi, A.K., Islam, A.: Power—and Variability-Aware Design of FinFET-Based XOR Circuit at Nanoscale Regime, In: IEEE International Conference on Advanced Communications, Control and Computing Technologies (ICACCCT), Ramanathapuram, Tamilnadu, India, May 08–10, 2014, pp. 440–444
9. [Online]. Available: <http://ptm.asu.edu/>
10. Pai, C.S., Diodato, P.W., Liu, R.: A case study of RC effects to circuit performance. In: Proceedings of the IEEE 1998 International Interconnect Technology Conference, IEEE press, 1998, pp. 244–246, doi:[10.1109/IITC.1998.704911](https://doi.org/10.1109/IITC.1998.704911)

Layered-Encoding Video Multicasting in Mobile WiMAX Networks

Jamil M. Hamodi and Ravindra C. Thool

Abstract The accessibility of good smartphones in 4G is encouraging mobile users to use video services. This huge increase in the usage of video communication over wireless networks poses many challenges according to fluctuations of wireless channel conditions along with user mobility. The challenge in providing video multicasting over wireless networks would most likely be the channel condition of each user in the same multicast group which is probably not identical due to location of the user and/or congestion of the network. In this paper, an adapted multicasting scheme has been proposed, which gracefully adapts the video quality to heterogeneous receivers and varying network conditions. Results obtained from simulation indicate that an enhancement of video quality under numerous channel conditions was through graceful degradation of these channels in terms of average throughput and packet loss ratio.

Keywords IPTV · QoS · Modulation and coding · WiMAX · OPNET · Performance study · Multicast

1 Introduction

Next-generation broadband access networks (i.e., 4G) are required to generate a more achievable information rate, and offer Ethernet speed along with range effectiveness. LTE in addition to WiMAX are a couple of completely different technologies to produce 4G recognized speeds. Completely new traditional proposals or even lets out beyond 4G are listed in traditional systems, for instance,

J.M. Hamodi (✉) · R.C. Thool
Shri Guru Gobind Singhji Institute of Engineering and Technology (SGGSIE&T),
Nanded, India
e-mail: jamil_hamodi@yahoo.com

R.C. Thool
e-mail: rcthool@yahoo.com

3GPP, WiMAX community or even ITU-R. 4G companies are better for user mobility in comparison to WiFi systems, to produce real-time video services. The International Telecommunication Institute (ITU) mandates that if user has a 4G technology, he should use a peak data rate of 100 Mb/s pertaining to high-mobility users in addition to 1 Gb/s pertaining to low-mobility users [1, 2]. These types of rising bandwidth help users to earn real-time video services, including IPTV services, live video streaming, as well as Internet-based telecast regarding sports.

Multicast is definitely a productive technique of profiting distributed dynamics involving wireless range to produce targeted prospects to be available to a number of consumers concurrently, although minimizing wireless useful resource consumption. Wireless multicast comes with a productive process to help these kinds of services as a direct result-distributed dynamics inside the wireless medium. The challenge in offering video multicasting over wireless networks would certainly link condition of each user inside the same multicast group and might not be the same as the position from the individual and/or congestion of the network. A new challenge when using the adapted modulation and code scheme in wireless multicast services will be the fact that a number of receivers experience heterogeneous channel condition (e.g., because of their distance to the mobile towers or even the locations), any effective modulation and code scheme (MCS) that is certainly and simply acquired successfully by almost all wireless consumers inside multicast group is usually useful to accommodate almost all receivers. For this reason, the multicast data rate and video quality usually are restricted to end users when using the worst channel condition [4]. In multicast scenario, receivers may very well be different in line with latency requirements, visual quality requirements, processing features, power constraints (wireless or wired), and bandwidth constraints. The actual heterogeneous requirements involving receivers make it tough to generate a powerful multicast process. To deal with the issue, which involves low-data rate according to the end users inside the worst channel quality, different modulation and coding techniques in different scalable video layers have been applied such as end users inside very good channel conditions who have been given more enhancement layers (EL) and have absolutely better video quality, and also the end users inside bad channel condition who have been given a lesser amount of ELs and have absolutely basic video quality.

Thus, in this paper, an adapted multicasting scheme has been proposed by applying layered-encoding of video streaming that will gracefully adapt the video quality of heterogeneous users besides varying network conditions. The property associated with scalable coding is valuable using the enhanced ability of user's mobility. This technique takes advantage of IP multicast, which is the actual way of one-to-many transmission in excess of the IP infrastructure in different networks. This may be used to multicast to quite a few customers without demanding beforehand knowledge of the number of receiving subscribers (SS) there are. OPNET provides comprehensive improvement associated with multilevel designs which include all the essential parameters that must be mirrored from the layout method associated with PHY and/or MAC layers. The investigation

work and results offered within this paper emphasis mainly upon the implementation of real-time scalable video coded (SVC) streaming for modeling along with simulation multicasting video layered-encoding deployment in mobile WiMAX. It mainly aims to establish some sort of efficient study associated with layered-encoded video multicasting IPTV transmission in mobile WiMAX underneath heterogeneous receivers as a way to take a look at and also to analyze user behavior with this model.

The entire content of this paper is usually structured as follows: Sect. 2 allows applicable background along with preliminaries in WiMAX PHY layer modulation and coding techniques, and also scalable video coding (SVC). Usually, Layered-encoded video streaming over wireless networks is typically pointed out inside Sect. 3. The architecture of layered video multicasting IPTV over mobile WiMAX networks is developed in Sect. 4. Also, Sect. 4 provides SVC-dependent adaption scheme for multicast over mobile WiMAX networks. Section 5 explains the particular useful steps to get recent analyzed by the recent ALE simulation. Simulation outcomes along with analysis usually are acquired in Sect. 6. Section 7 is usually, of course, the final outcomes of the findings as well as a summation for this modest research endeavor.

2 Background

This section offers related historical past preliminaries intended for physical layer on WiMAX modulation and coding schemes, as well as SVC.

2.1 *WiMAX Preliminaries*

WiMAX supports numerous modulations and coding schemes in addition to make it possible for each scheme to be changed even on a burst-by-burst basis per link that will be determined by channel conditions, which are : QPSK, 16-QAM, and 64-QAM modulation, however, 64-QAM is optional in the uplink channel. Also, WiMAX supports adaptive modulation and coding schemes which significantly raise the system capacity by allowing real-time trade-off between throughput and robustness on each link. Unlike uplink, the base station (BS) may definitely evaluate the channel quality when using the delivered signal quality. The BS scheduler can find the channel quality of every user's uplink and downlink in addition to delegating a modulation and coding scheme that will maximize the throughput for the signal-to-noise ratio (SINR). This depends on the indicator of the channel quality feedback which helps the mobile to provide the BS with feedback over the downlink channel quality. OFDMA system architecture is shown in Fig. 1, which contains one transmitter BS along with M downlink mobile receivers (users) with

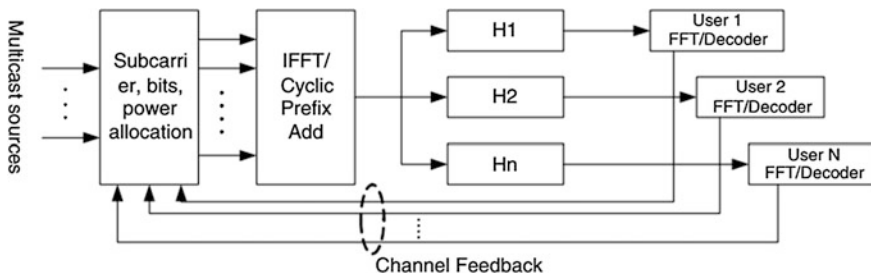


Fig. 1 OFDMA downlink multicast transmission [5]

channel feedback. However, OFDMA technology is normally introduced to handle frequency selective channel fading [5].

2.2 H.264/SVC Video Compression

SVC is definitely the right attractive method to fix that drawbacks triggered in the attributes of modern video transmission systems. Typically, the word “scalability” represents the removal of components from the video bit stream of approach that enables user to adjust it to the number of must haves and also preferences from consumers along with assorted final features, and also network conditions. SCV is specially made for multicast that it helps delivering video streaming to varying numbers of users with heterogeneous channel capacities [6, 7].

The typical modes of scalability are generally temporal, spatial, and quality scalability. The part of the stream representing lower picture size (spatial resolution) is called spatial scalability, whereas if this part represents frame rate (temporal resolution) then it is called temporal scalability. With quality scalability, the sub-stream supplies the same spatio-temporal resolution as the overall bit stream, though with a lesser fidelity, where fidelity is commonly and informally named as signal-to-noise ratio (SNR) [6]. The two common scalabilities which described the quality of scalability are fidelity and SNR scalability. A spatial scalability is used to encode video stream before injecting into OPNET modeler. All layered video streams, base layer (BL), ELs used in this simulation are obtained from Arizona State University [8].

Spatial Scalability: A relevant video stream is spatially scalable when anyone can dynamically vary the pixel resolution of the transmitted video frames. Thus, anyone can transmit or extract a stream of lower spatial resolution for video streaming on handheld devices including mobile phone devices and PDAs; while higher spatial resolutions are usually employed for video conferencing, gaming, along with applications on portable laptop computers. Again, one can discard all spatial layers S_k in ways that $k \geq K+1$, and the remainder of the layers constitutes a

valid video stream along with a lower spatial resolution compared to one that includes all of the spatial layers [3], description of all mode types are omitted due to lack of space.

3 Layered Video Stream Over Wireless Networks

Supporting a video layered-encoding multicasting in a network is not an alternative concept. Huge numbers of existing works in [9–18] are carried out to study distinct issues in this area, all of these were not discussed in video layered-encoded multicasting through graceful adaptation. The works [9–11] are really among the other ones which usually propose techniques of adjusting the radio resource allocation scheme for layer-encoded video streams multicasting over wireless networks. Kim et al. [9] proposed a new method which usually intends to increase the total utility by adjusting the resource allocation level. However, each layer has been assigned by different modulation scheme. Liu et al. [10] proposed a method to increase the total utility of all users in networks by finding out the optimal quantity of layers along with level of bandwidth for each layer. According to users' channel condition, the BS could not adjust the transmission rates along with the data rate for each receiver. A utility-based method for layer-encoded IPTV multicasting has been proposed in [11] that is dependant on each layer's utility value along with users' channel condition.

On the other hand, there has been extensive work for the situation associated with learning resource allocations intended for scalable video multicast in wireless networks in [12–18]. A technique of layered-encoded video to wireless multicasting services has been presented in [12]; they have used only two MCSs of their scheduling system. An optimal bandwidth allocation algorithm has been proposed in [13] that assumes that each user in wireless networks possesses a limit in capacity that appeared on the quantity of channels the user can receive. Kim et al. [14] proposed a greedy algorithm called MP-AMC algorithm that greedily determines how much period of slots were used on each and every video sessions along with the MCS for every single individual layer within a number of video sessions. Even though, this algorithm treats how much layers every end user may receive once this individual user joins the multicast group. Kuo et al. [15] proposed a suboptimal algorithm called U-LEM, which determines how much layers are given by every end user dynamically prior to their channel conditions and probable community bandwidth in order to enhance the full utility. In [16, 17], optimum system utility seems to be attained under the predictions of the final result rate of each and every video layer which is normally randomly altered. Within useful video streaming systems, the BS cannot customize the particular layer rate while considering the particular video encoding which is actually performed inside the video server. A genetic algorithm has been proposed in [18], which adaptively changes the modulation techniques of every video layer to minimize a typical distortion of all clients.

4 System Architecture and the Adaption Scheme for Layered-Encoding Multicast

4.1 System Model

A testbed is developed for the investigated performance of layered-encoded video multicasting IPTV transmission over mobile WiMAX networks. The testbed, as shown in Fig. 2, is definitely an end-to-end multicast architecture, that can evaluate video layered-encoding as well as multicast broadcast service (MBS) in WiMAX systems. It has two main components: a video server section and mobile WiMAX network. The video server component stores video content which is here encoded by H.264 SVC encoder whose encoded video content is divided into multiple layers: BL and multiple ELs. BL, which is called quarter common intermediate format (QCIF)/quarter video graphic array (QVGA), provides a simple video quality, frame rate, and video resolution, as well as enhancement layer (HL), which is called common intermediate format (CIF)/video graphic array (VGA), that will elaborate the video quality, frame rate, and video resolution. This model works as follows: all video layers will be transmitted to the BS that happens when an individual user subscribed a movie. Then, BS determines more clients to be scheduled for each video layer by adopting modulation techniques. Figure 2 shows the video layer-encoding multicasting architecture in mobile WiMAX networks. This model contains one multicast group containing heterogeneous receivers with various distance from the BS. In order to include almost all MSSs for the multicast group, BS chooses QPSK, possibly the robust MCS, to transmit BL to all MSSs in multicasting group.

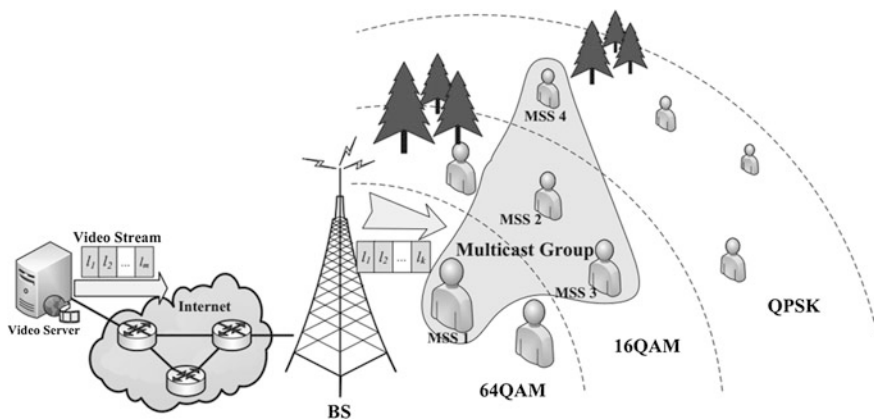


Fig. 2 Video layered-encoding multicasting architecture in mobile WiMAX [19]

4.2 *Adaption Scheme for Layered-Video Multicast*

To the best of the authors' knowledge, particular individual feedbacks for mobile subscriber station (MSS) are certainly not available in multicast group. Usually, clients on the BS are suspected to suffer from good channel conditions. However, clients in cell edge are suspected to suffer from poor channel conditions. A simple, but useful adaption scheme for layered-encoding video has been proposed. This scheme enhances video quality by improving throughput when considering reduction in bit rate. The adaptation scheme is briefly described below: to guarantee the basic quality of video to all users in group, BS transmits the BL to all users after ordering them in decreasing order depending on their channel condition. Then the other EL sending is as follows:

- Step 1: All the two temporal layers (L1 and L2) tend to be provided for users close to the BS.
- Step 2: Users faraway from the BS tend to be served together with only temporal layer 1, which contains only the Intra-coded frames (I Frames) along with Predictive coded frames (P frames).

This approach improves throughput by minimizing the involving loss of I and P frames.

5 Simulation Model

A simulation model is presented in this section. It is used to analyze the effect of multicasting layered-encoding video (IPTV) over mobile WiMAX networks. As a result of differences in geographical location, channel noise/interference, fading and path loss, different receivers experience, and different channel conditions, the simulation is performed to gauge the performance study of multicast layered video during the mobile WiMAX networks under different heterogeneous receivers that are able to investigate and analyze the behavior and gratification of the model.

The main goal of this paper is to establish multicast model for layer-encoding video multicasting under time-varying channel conditions. In traditional multicast scheme, BS broadcasts packets to all users in any multicast group that uses the same data rate. However, users could possibly be based at assorted locations or perhaps user mobile at different speeds in the cellular. Consequently, the ability of multicast service is throttled in the way of a person with the worst channel state. Our multicast model differs from that state, where video contents may be decomposed into two layers: BL and multiple ELs. Essentially, the most related aspects of video sequence usually are included in the BL, even though fewer related waste details usually are put during the entire EL. As a result, video traces of Tokyo Olympics of spatial scalability based on SVC scheme along with quarter common intermediate format (QCIF), BL (L0), in addition to common intermediate format (CIF), ELs

(L1, L2) have been of our own simulation. The video traffic injected into OPNET simulation has been obtained from Arizona State University and includes these specifications: a 532×288 video resolution; with size selected for group of picture (GOP) as 16 due to this video, also it will be encoded with 30 frames per seconds (fps). To ensure basic streaming for all those multicast users, the BL flow is going to be effectively carried to users with worst channel quality, and also to almost all users using the higher channel quality. The ELs flow is actually carried purely toward users using higher channel condition that is contingent in distance from the BS, then, toward the modulation, and coding scheme. Numerous data from ELs might be received by number of users according to the wireless channel, which is generally frequency selective.

Initially, Fig. 3 shows that the topology is evaluated, a relative server with a movie encoder able of transmitting multilayer video to all subscriber stations (SSs). It is also assumed that BL (L0) video stream is transmitted to each user in multicasting group. Moreover, it is assumed that the EL (L1) video stream is also transmitted by users having 16-QAM MCS, also it is assumed that users on the BS have better modulation that is 64-QAM, and would be able to receive all video layers, to guarantee better video quality.

According to simulation, the favored MIL3 OPNET modeler simulation must be used. Here the OPNET modeler may be used in order to expedite any usage of inbuilt models of commercially accessible network elements with reasonably accurate emulation of various true to life network topologies. Independent cases from the video conferencing application utilized the particular individual and distinctive stream, one for BL and the other two for enhancement layers within the Tokyo Olympics movie. One of the key parameters in this configuration will be the

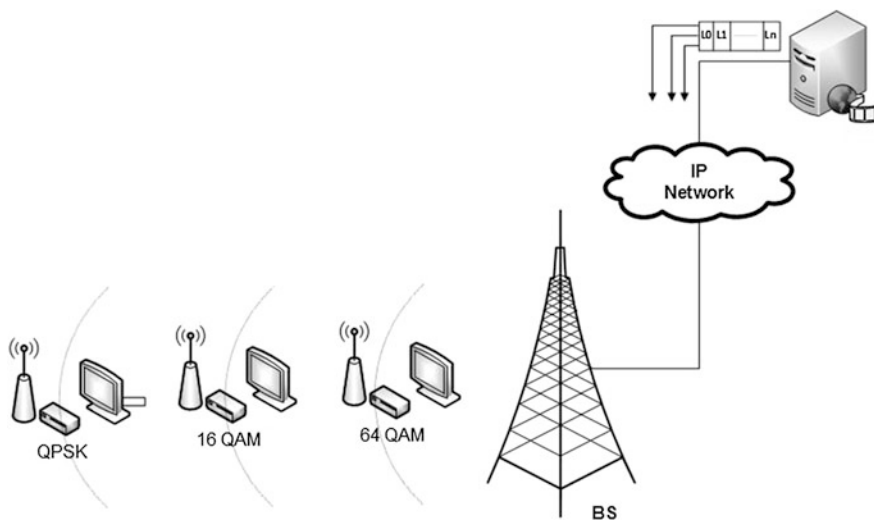


Fig. 3 Topology of layered-video multicasting over WiMAX

interarrival frame time period and frame size. The video frame rate has been configured as 30 (fps) in the parameter of incoming inter-arrival times. However, the outgoing interarrival time has been set to “none” to get unidirectional streaming of video server. Moreover, the substream video layer traces have been injected as script file in the frame size parameters of video conference application.

Three independent profiles are configured to work simultaneously during this model. In the first profile, the BL (L0) video stream is configured to be transmitted to all users in multicasting group. The second profile is configured to transmit (L1) as a first EL to users who have 16-QAM modulation and coding schemes, and the third profile transmit a second EL (L2) to users who have higher modulation scheme 64-QAM MCS.

6 Simulation Results and Discussion

Average packet jitter, end-to-end delay, WiMAX delay, and throughput of mobile node are evaluated through this simulation under different heterogeneous receivers, which represent the channel condition and distance from (BS). A free space model is configured in this simulation as a path loss model along with rtPS as service class, these are thought to be constant.

Figure 4a, b presents the average packet jitter, along with average E2E delay according to various heterogeneous receivers. As in authors’ knowledge, video quality is better when the jitter is zero. Therefore, as shown in Fig. 4a, average layer video multicasting IPTV is approximately zero for users near the BS, which have higher modulation and coding scheme (64 QAM), whereas users near the cell edge had a worse average variation of jitter. Average end-to-end delay for layer encoding video multicasting IPTV under heterogeneous receivers is shown in Fig. 4b, as it is

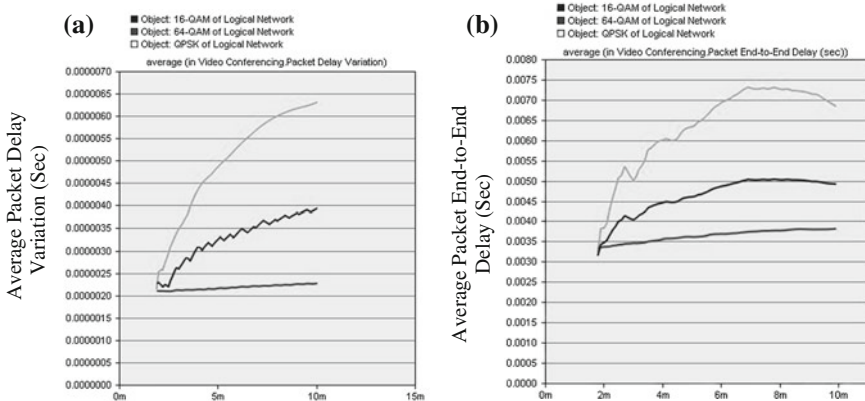


Fig. 4 a Average jitter of user node and b average packet E2E delay of user node

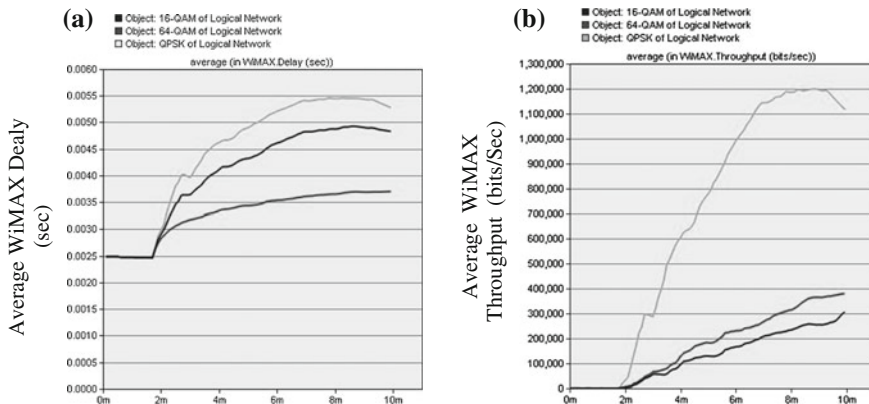


Fig. 5 a Average WiMAX delay for user node and b average WiMAX throughput for user node

possible to observe that the average E2E delay of layer video multicasting IPTV gives lower packet E2E delay for users near the BS.

Based on the proposed method, almost all practical temporal layers along with BL are usually delivered to clients near BS, whereas only BL can be delivered to clients near cell edge. Therefore, as shown in Fig. 5a, the WiMAX delay is significantly higher at QPSK modulation and coding scheme, and turns out to be less in higher modulation and coding scheme (QAM). WiMAX delay will be higher that affects the throughput and will undoubtedly be higher as shown in Fig. 5b. From Fig. 5a, it is clear that the delay may be very low for 64-QAM modulation and coding scheme. However, another modulation and coding scheme has higher delay. The packet loss ratio is reduced for users close to BS that seems to be 64-QAM modulation and coding scheme. As a result, the throughput will be improved when the delay is less. The results as with Figs. 5a, b, indicate that video quality for clients near BS has been gracefully degraded in part of reduction in average throughput along with packet loss ratio besides acceptable PSNR value. There are also clients modulation by 16-QAM that require significantly smaller amount of data rate (0.38 Mbps) in comparison to different modulation QPSK. Table 1 summarized the results and outcomes from simulation. Furthermore, scalable video representation is a nice method for heterogeneity injury in multicast case.

Table 1 Quality of services of SVC video under different modulation and coding techniques

Parameters	QPSK	16 QAM	64 QAM
Jitter	6.3 μ s	3.9 μ s	2.2 μ s
End-to-End delay	6.8 ms	4.9 ms	3.8 ms
WiMAX delay	5.2 ms	4.8 ms	3.7 ms
Throughput	1.1 Mbps	0.3 Mbps	0.38 Mbps

7 Conclusions and Perspective Work

This specific paper proposes a particular adaptation multicasting scheme using layered-encoding video streaming to enhance quality of video service of heterogeneous and has varying network conditions through graceful adaptation. Layer encoding of video stream is the process of encoding the playback quality file at different bitrates and streaming it through different ports. IP multicast has been used in this system, which is a way to transmit data from one-to-many over an IP infrastructure in a network. A case study was presented for deploying layered encoding video multicasting IPTV over mobile WiMAX networks, in addition to OPNET simulation scenario was developed to check out numerous numbers of performance metrics including: packet jitter delay, E2E delay, WiMAX delay of mobile node, and WiMAX throughput of mobile node. Outcomes obtained from simulation indicate that a perceptual good video quality is gracefully degraded while considering numerous channel conditions, which are part of average throughput and packet loss ratio. Further work needs to study the effect of video-layered encoding over broadband wireless networks on part of resource allocation.

References

1. Costa, J.M.: More frequencies needed for mobiles—Terrestrial spectrum sought for IMT. *ITU News*. **3**, 18–21 (2007)
2. Fagbohun, O.: Comparative studies on 3G, 4G and 5G wireless technology. *IOSR J. Electron. Commun. Eng.* **9**(3), 88–94 (2014)
3. Chilamkurti, N., Zeadally, S., Chaouchi, H.: *Next-Generation Wireless Technologies: 4G and Beyond*. Springer, London (2013)
4. Peilong, Li., Honghai, Zh., Baohua, Zh., Sampath, R.: Scalable video multicast with adaptive modulation and coding in broadband wireless data systems. *IEEE/ACM Trans. Netw.* **20**(1) (2012)
5. Yuedong, Xu, Xiaoxin, Wu, John, C.S.: Cross-Layer QoS scheduling for layered multicast streaming in OFDMA wireless networks. *Wirel. Pers. Commun.* **51**, 565–591 (2009)
6. Heiko, Sc., Detlev, M., Thomas, W.: Overview of the scalable video coding extension of the H.264/AVC Standard. *IEEE Trans. Circ. Syst. Video Technol.* **17**(9), pp. 1103–1120 (2007)
7. Vladimir, V., György, D.: Multicast Scheduling for Scalable Video Streaming in Wireless Networks. In: *MMSys'10*, Phoenix, Arizona (2010)
8. Video Trace Library, Arizona State University, <http://trace.eas.asu.edu>
9. Kim, J., Cho, J., Shin, H.: Resource allocation for scalable video multicast in wireless cellular networks. In: *WiMob* (2005)
10. Liu, J., Li, B., Hou, Y.-T., Chlamtac, I.: On optimal layering and bandwidth allocation for multi-session video broadcasting. *IEEE Trans. Wirel. Commun.* **3**(2), 656–667 (2004)
11. Kuo, W.-H., Liu, T., Liao, W.: Utility-based resource allocation for layer encoded multicast streaming in IEEE 802.16 (WiMAX) wireless networks. In: *IEEE ICC* (2007)
12. Hwang, C.S., Kim, Y.: An adaptive modulation method for multicast communications of hierarchical data in wireless networks. *IEEE ICC* **2**, 896–900 (2002)
13. Liu, J., Li, B., Hou, Y.T., Chlamtac, I.: Dynamic layering and bandwidth allocation for multisection video broadcasting with general utility functions. *IEEE INFOCOM* **1**, 630–640 (2003)

14. Kim, J., Cho, J., Shin, H.: Resource allocation for scalable video broadcast in wireless cellular networks. *IEEE WiMob* **2**, 174–180 (2005)
15. Kuo, W. H., Liu, T., Liao, W.: Utility-based resource allocation for layer-encoded IPTV multicast in IEEE 802.16 (WiMAX) wireless networks. In: *IEEE ICC*, pp. 1754–1759 (2007)
16. Kim, J., Cho, J., Shin, H.: Layered resource allocation for video broadcasts over wireless networks. *IEEE Trans. Consum. Electron.* **54**(4), 1609–1616 (2008)
17. Shi, J., Qu, D., Zhu, G.: Utility maximization of layered video multicasting for wireless systems with adaptive modulation and coding. *IEEE ICC* **11**, 5277–5282 (2006)
18. Chi, H., Lin, C., Chen, Y., Chen, C.: Optimal rate allocation for scalable video multicast over WiMAX. In: *IEEE ISCAS*, pp. 1838–1841 (2008)
19. Tsai, Ch.-T., Jan, R.-H., Chen, Ch.: Optimal modulation and coding scheme allocation of scalable video multicast over IEEE 802.16e networks. *EURASIP J. Wirel. Commun. Netw.* **33** (2011)

Intuitionistic Fuzzy Similarity and Information Measures with Physical Education Teaching Quality Assessment

Arunodaya Raj Mishra, Divya Jain and D.S. Hooda

Abstract Information and similarity measures have a vital place in the fuzzy set theory. It has been investigated by many researchers with different aspects. In this paper, new intuitionistic fuzzy similarity and information measures based on sine function are proposed. Comparison of proposed information measures with the existing ones is listed. Numerical results clearly indicate the efficiencies of these measures over others. New technique for multi-criteria decision-making (MCDM) quandaries to rank the alternatives is introduced. This technique is developed on the application intuitionistic fuzzy information measure and weighted averaging operator (IFWAO). A case of five colleges ranking of a district region is studied and discussed.

Keywords Intuitionistic fuzzy set · Intuitionistic fuzzy information measure · Similarity measure · MCDM · TOPSIS · Physical education teaching quality

1 Introduction

Fuzziness is the imperfection and skepticism, if the limits and margins of the set under the concern is not well defined, i.e., whether the element belongs to or does not belong to a set. The incipient approaches and theories treating imprecision and dubiousness were introduced by Zadeh [1]. Grzegorzewski [2] discussed the promising and indispensable inclusion of IFS followed [3] who introduced the

A.R. Mishra (✉) · D. Jain · D.S. Hooda
Department of Mathematics, Jaypee University of Engineering and Technology,
Guna 473226, M.P., India
e-mail: arunodaya87@outlook.com

D. Jain
e-mail: divya.jain@juet.ac.in

D.S. Hooda
e-mail: ds.hooda@juet.ac.in

concept of intuitionistic fuzzy set (IFS). Mishra et al. [4, 5] proposed exponential and trigonometric intuitionistic fuzzy information measures and applied these to assessment of service quality and rating of township development. Castineira [6] recognized the incompatibility measure in the framework of the IFS of Atanassov in the form of axioms.

While Liu and Wang [7] introduced incipient methods to solve multi-criteria decision-making quandaries in an intuitionistic fuzzy environment. An approach in group decision-making was developed by Xu [8, 9] on intuitionistic preference relations and also proposed clustering algorithm for IFS. The distance between IFS was investigated by Szmidt and Kacprzyk [10], Wang and Xin [11], and many other researchers further studied the distance measures in IFS and applied it to many real-life problems of pattern recognition.

The degree of homogeneous attribute or similarity of fuzzy sets is known as similarity measures. It is a consequential concept in the fuzzy set theory. It was introduced by Wang [12] who also provided computation formula. Thereafter it occupied the main focus of the various research findings. Many investigations have been done and extensively applied in fuzzy clustering, image processing, fuzzy reasoning, and fuzzy neural networks. Comparative study of similarity measures was conducted by Wang et al. [13]. Many studies fixate on the similarity measure and the information measure for IFS [11, 14] and the relationship between them, particularly, on the organized conversion of the information measure to the similarity measure and similarity measure to information measure for IFSs.

In today's scenario, it is common to make decision in the presence of multiple conflicting criteria, which is known as multiple criteria (or attribute) decision-making (MCDM/MADM) quandaries. For example, if one has to buy a refrigerator, he has to look at many parameters like prize, capacity, safety, power consumption, technology, etc. A decision in multi-criteria decision-making quandaries can be taken by selecting the best suitable alternatives based on quantitative and qualitative assessment. Required option can be chosen by providing the preference set in the terms of numerical value or interval. But, in real-life situation the data set or preference values are not always precise, it can be vague or imprecise, which can be great dispute for the decision taking authorities and hence was a leading factor toward MCDM techniques in fuzzy scenarios.

The technique for ordering the preferable option by similarity to an ideal elucidation (TOPSIS) was developed by Hwang and Yoon [15]. In this technique, the concept is that the chosen option should be the closest from the positive ideal elucidation and the farthest from the negative ideal elucidation. Joshi and Kumar [16] proposed intuitionistic fuzzy information and distance measure-based TOPSIS technique for multi-criteria decision-making (MCDM). Likewise various authors introduced IFS with TOPSIS to give a hybrid technique for MCDM problems. Some methods were based on IFS using entropy weights and linear programming given by Liu and Wang [7, 17, 18].

The present communication is arranged as follows: Sect. 2 introduces the fundamental conceptions of fuzzy sets, intuitionistic fuzzy sets, intuitionistic fuzzy aggregation operator, and intuitionistic fuzzy information and similarity measures,

followed by the study of new similarity and intuitionistic fuzzy information measures and their validity in Sect. 3. In Sect. 4, empirical illustration is listed, and comparison is tabulated between new and existing intuitionistic fuzzy information measures, followed by the application of the proposed intuitionistic fuzzy information measure in physical education teaching assessment in Sect. 5.

2 Prerequisite

In this section, authors discuss some fundamental conceptions of fuzzy sets and intuitionistic fuzzy sets, and information and similarity measures for IFSs. Throughout this paper, FSs and IFSs represent the fuzzy sets and intuitionistic fuzzy sets in X , respectively.

Definition 2.1 ([1]) Let $X = \{x_1, x_2, \dots, x_n\}$ be a finite universe of discourse and $\tilde{A} \subset X$. Then \tilde{A} is a fuzzy set defined by

$$\tilde{A} = \{(x_i, \mu_{\tilde{A}}(x_i)) : \mu_{\tilde{A}}(x_i) \in [0, 1]; \forall x_i \in X\}, \tag{1}$$

where $\mu_{\tilde{A}}(x_i)$ is the membership function $\mu_{\tilde{A}}(x_i) : X \rightarrow [0, 1]$ such that

$$\mu_{\tilde{A}}(x_i) = \begin{cases} 0, & \text{if } x_i \notin \tilde{A} \text{ and there is no ambiguity} \\ 1, & \text{if } x_i \in \tilde{A} \text{ and there is no ambiguity} \\ 0.5, & \text{there is maximum ambiguity whether } x_i \notin \tilde{A} \text{ or } x_i \in \tilde{A}. \end{cases}$$

Definition 2.2 ([3]) The intuitionistic fuzzy set $A \subset X = \{x_1, x_2, \dots, x_n\}$ is defined as

$$A = \{ \langle x, \mu_A(x), \nu_A(x) \rangle : x \in X \}, \tag{2}$$

where $\mu_A : X \rightarrow [0, 1]$ is the degree of membership and $\nu_A : X \rightarrow [0, 1]$ is the degree of nonmembership of $x \in X$ in A , respectively, such that

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1, \forall x \in X.$$

The intuitionistic index (or hesitancy degree) of an element $x \in X$ in A is as

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x).$$

It implies $0 \leq \pi_A(x) \leq 1, \forall x \in X$ [3].

If $\pi_A(x) = 0$ then IFSs can formed FSs, i.e., $A = \{ \langle x, \mu_A(x), 1 - \mu_A(x) \rangle : x \in X \}$, with $\pi_A(x) = 0, \forall x \in X$.

The complement set of A is A^c and defined as

$$A^c = \{ \langle x, v_A(x), \mu_A(x) \rangle : x \in X \}. \tag{3}$$

[19] represents $\alpha = \langle a, b \rangle$ an intuitionistic fuzzy number (IFN), if $0 \leq a + b \leq 1$.

Definition 2.3 Let $\alpha_1 = \langle a_1, b_1 \rangle$ and $\alpha_2 = \langle a_2, b_2 \rangle$ be two IFNs, three intuitionistic fuzzy aggregation operators are defined by

$$\alpha_1 \oplus \alpha_2 = \langle a_1 + a_2 - a_1a_2, b_1b_2 \rangle, \alpha_1 \otimes \alpha_2 = \langle a_1a_2, b_1 + b_2 - b_1b_2 \rangle \text{ and} \tag{4}$$

$$w\alpha_1 = \langle 1 - (1 - a_1)^w, b_1^w \rangle, w > 0.$$

Definition 2.4 ([20]) Let $h : IFS(X) \rightarrow [0, 1]$ be an information measure defined on IFSs. Then it is valid, if it holds the given postulate:

- (P1). $h(A) = (\text{minimum}) \Leftrightarrow A$ is a crisp set;
- (P2). $h(A) = 1$ (maximum) $\Leftrightarrow \mu_A(x_i) = v_A(x_i)$ for any $x_i \in X$;
- (P3). $h(A) \leq h(B)$ and if A is less fuzzy than B , i.e., $\mu_A(x_i) \leq \mu_B(x_i)$ and $v_A(x_i) \geq v_B(x_i)$ for $\mu_B(x_i) \leq v_B(x_i)$ or $\mu_A(x_i) \geq \mu_B(x_i)$ and $v_A(x_i) \leq v_B(x_i)$ for $\mu_B(x_i) \geq v_B(x_i)$ for any $x_i \in X$;
- (P4). $h(A) = h(A^c)$.

Definition 2.5 The function $Sim : IFS(X) \times IFS(X) \rightarrow [0, 1]$ is said to be a valid measure of similarity on $IFS(X)$, if Sim holds

- (S1). $Sim(A, A^c) = 0$ if A is a crisp set;
- (S2). $Sim(A, B) = 1$, iff $A = B$;
- (S3). $Sim(A, B) = Sim(B, A)$;
- (S4). $Sim(A, C) \leq Sim(A, B)$ and $Sim(A, C) \leq Sim(A, B)$, For all $A, B, C \in IFS(X)$ and $A \subseteq B \subseteq C$.

3 Similarity and Information Measures for IFSs

Here, similarity and information measures for IFSs are proposed and their validity are proved.

Let $A = \{ \langle x, \mu_A(x), v_A(x) \rangle : x \in X \}$ and $B = \{ \langle x, \mu_B(x), v_B(x) \rangle : x \in X \}$ be two IFSs in X defines

$$Sim(A, B) = 1 - \frac{1}{n} \sum_{i=1}^n \sin \left[\frac{\{ \max(|\mu_A(x_i) - \mu_B(x_i)|, |v_A(x_i) - v_B(x_i)|) \} \pi}{2} \right]. \tag{5}$$

Theorem 1 *The mapping Sim, defined by (5) is a similarity measure for IFSs.*

Proof It is obvious that $Sim(A, B)$ holds the postulate (S1)–(S3). Hence only (S4) has to be proved. For $A \subset B \subset C$, we have $\mu_A(x_i) \leq \mu_B(x_i) \leq \mu_C(x_i)$, and $v_A(x_i) \geq v_B(x_i) \geq v_C(x_i), \forall x_i \in X$. It implies

$$\begin{aligned} & \max(|\mu_A(x_i) - \mu_B(x_i)|, |v_A(x_i) - v_B(x_i)|) \\ & \leq \max(|\mu_A(x_i) - \mu_C(x_i)|, |v_A(x_i) - v_C(x_i)|), \forall x_i \in X. \\ \sin & \left[\frac{\{\max(|\mu_A(x_i) - \mu_B(x_i)|, |v_A(x_i) - v_B(x_i)|)\} \pi}{2} \right] \\ & \leq \sin \left[\frac{\{\max(|\mu_A(x_i) - \mu_C(x_i)|, |v_A(x_i) - v_C(x_i)|)\} \pi}{2} \right], \forall x_i \in X. \end{aligned}$$

These imply

$$\begin{aligned} & 1 - \frac{1}{n} \sum_{i=1}^n \sin \left[\frac{\{\max(|\mu_A(x_i) - \mu_B(x_i)|, |v_A(x_i) - v_B(x_i)|)\} \pi}{2} \right] \\ & \geq 1 - \frac{1}{n} \sum_{i=1}^n \sin \left[\frac{\{\max(|\mu_A(x_i) - \mu_C(x_i)|, |v_A(x_i) - v_C(x_i)|)\} \pi}{2} \right], \forall x_i \in X. \end{aligned}$$

This means that

$$Sim(A, B) \geq Sim(A, C). \tag{6}$$

Similarly,

$$Sim(A, B) \geq Sim(A, C). \tag{7}$$

Let $A = \{(x, \mu_A(x), v_A(x)) : x \in X\}$ be an IFS in X defines

$$h(A) = \frac{1}{n} \sum_{i=1}^n \sin \left\{ \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)} \right) \frac{\pi}{2} \right\}. \tag{8}$$

Theorem 2 *The mapping h, defined by (8) is information measure on IFSs.*

Proof Let $h_i(A) = \sin \left\{ \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)} \right) \frac{\pi}{2} \right\}$, for $i = 1, 2, \dots, n$. From $0 \leq \mu_A(x_i) \leq 1, 0 \leq v_A(x_i) \leq 1$, we have $0 \leq \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)} \right) \frac{\pi}{2} \leq \frac{\pi}{2}$. Thus $0 \leq h_i(A) \leq 1$.

(P1). For $\mu_A(x_i) = 0, v_A(x_i) = 1$ or $\mu_A(x_i) = 1, v_A(x_i) = 0$ for any $x_i \in X$. It implies $h_i(A) = 0$. Hence $h(A) = 0$.

Conversely, suppose that $h(A) = 0$ for $i = 1, 2, \dots, n$. Since $0 \leq h_i(A) \leq 1$, therefore, it follows that $h_i(A) = 0$. Also since $0 \leq \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)} \right) \frac{\pi}{2} \leq \frac{\pi}{2}$, we have

$$\left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)}\right) \frac{\pi}{2} = 0 \text{ or } \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)}\right) \frac{\pi}{2} = \frac{\pi}{2}.$$

Hence A is a crisp set.

(P2). Let $\mu_A(x_i) = v_A(x_i)$ for each $x_i \in X$. Applying this condition to (8), we easily obtain $h(A) = 1$.

Conversely, suppose that $h(A) = 1$. From (8) and $0 \leq h_i(A) \leq 1$, we obtain that $h_i(A) = 1$ for each $x_i \in X$. Also from $0 \leq \left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)}\right) \frac{\pi}{2} \leq \frac{\pi}{2}$, we have $\mu_A(x_i) = v_A(x_i)$ for each $x_i \in X$.

(P3). Let $\mu_A(x_i) \geq \mu_B(x_i)$ and $v_B(x_i) \geq v_A(x_i)$ for $\mu_B(x_i) \geq v_B(x_i)$, i.e., $\mu_A(x_i) \geq \mu_B(x_i) \geq v_B(x_i) \geq v_A(x_i)$. Thus

$$\left(\frac{\mu_A(x_i) \wedge v_A(x_i)}{\mu_A(x_i) \vee v_A(x_i)}\right) \frac{\pi}{2} \leq \left(\frac{\mu_B(x_i) \wedge v_B(x_i)}{\mu_B(x_i) \vee v_B(x_i)}\right) \frac{\pi}{2}. \tag{9}$$

It follows that $h_i(A) \leq h_i(B)$.

Similarly, when $\mu_A(x_i) \leq \mu_B(x_i)$ and $v_B(x_i) \leq v_A(x_i)$ for $\mu_B(x_i) \leq v_B(x_i)$, i.e., $\mu_A(x_i) \leq \mu_B(x_i) \leq v_B(x_i) \leq v_A(x_i)$. we can also prove that $h_i(A) \leq h_i(B)$.

Hence we have

$$h(A) \leq h(B). \tag{10}$$

(P4). For $A^c = \{(x_i, v_A(x_i), \mu_A(x_i)) : x_i \in X\}$, we can easily get that

$$h(A^c) = h(A). \tag{11}$$

Hence $h(A)$ is valid intuitionistic fuzzy information measure.

4 Numerical Comparisons

In present section, the efficiency of information measure through comparisons with some existing information measures for IFSs is demonstrated.

Example 1 Let

$$A_1 = \{(x, 0.2, 0.4) : x \in X\}, A_2 = \{(x, 0.3, 0.4) : x \in X\},$$

$$A_3 = \{(x, 0.4, 0.5) : x \in X\}, A_4 = \{(x, 0.5, 0.5) : x \in X\}.$$

We first recall these information measures.

Vlachos and Sergiagis [21] proposed an entropy measure E_{VS} according to cross entropy measure:

Table 1 Comparisons with existing entropies

	A_1	A_2	A_3	A_4
h	0.7071	0.9239	0.9511	1
E_{VS}	0.9510	0.9897	0.9920	1
E_{WZ}	0.9749	0.9927	0.9898	1

$$E_{VS} = -\frac{1}{n \ln 2} \sum_{i=1}^n [\mu_A(x_i) \ln \mu_A(x_i) + \nu_A(x_i) \ln \nu_A(x_i) - (1 - \pi_A(x_i)) \ln(1 - \pi_A(x_i)) - \pi_A(x_i) \ln 2]. \tag{12}$$

Wei and Zhang [22] developed information measure based on cosine function for IFSs:

$$E_{WZ}(A) = \frac{1}{n} \sum_{i=1}^n \cos \left\{ \frac{(\mu_A(x_i) - \nu_A(x_i))\pi}{2(1 + \pi_A(x_i))} \right\}. \tag{13}$$

These two information measures satisfy the set of requirements in Definition 2.4.

From Table 1, it can be seen that h and E_{VS} , the nearer the degree of membership and nonmembership, the higher the information measure. Particularly, when $\mu_A(x_i) = \nu_A(x_i)$, the information measure reaches its maximum value. The results obtained by using the information measures h and E_{VS} are in accordance with the intuition.

Table 1 shows that $h(A_1) < h(A_2) < h(A_3) < h(A_4)$ and $E_{VS}(A_1) < E_{VS}(A_2) < E_{VS}(A_3) < E_{VS}(A_4)$, which are consistent with the intuition. But, from the above table it is visible that $E_{WZ}(A_1) < E_{WZ}(A_2) > E_{WZ}(A_3) < E_{WZ}(A_4)$. It implies that the information measure E_{WZ} is not in accordance with the intuition. And hence our measure is more efficient and authentic than the information measure E_{WZ} .

5 Physical Education Teaching Quality Assessment

So far, researchers concentrated on the empirical as well as theoretical evidences for the reform and development of the teaching quality assessment to deal with confounding quandaries of expansion and quality improvement of physical education. The association of teaching quality system and the proposal for the fundamental supervision is favorable to the growth and the suitable progress of the undergraduate physical education majors [23, 24].

Xu [19, 25] dealt with triangular intuitionistic fuzzy information measure in the multi-criteria attribute decision-making quandaries, while solving the problems of physical education teaching quality in the universities and colleges. Here, in this section, we develop an approach on intuitionistic fuzzy information measure to

evaluate the physical education teaching quality in the universities and colleges. The developed IFMWAO method is as follows:

- Step 1. Let $C = \{C_1, C_2, \dots, C_m\}$ be a discrete set of colleges of a district region, and $L = \{L_1, L_2, \dots, L_n\}$ be the set of options. Let $\tilde{D} = (\tilde{r}_{ij})_{m \times n} = (\mu_{ij}, \nu_{ij})_{m \times n}$ be the intuitionistic fuzzy judgment matrix, where μ_{ij} represents the extend up to which C_i holds L_j and ν_{ij} reflects C_i does not support the option L_j .
- Step 2. By using information measure formula (8), we can easily compute the information of each intuitionistic fuzzy value in the judgment matrix and obtain the information matrix of this judgment matrix as $D = (h_{ij})_{m \times n}$, where $h_{ij} = E(\tilde{r}_{ij})$.
- Step 3. Normalize the information values in the above decision matrix by the following equation

$$\bar{h}_{ij} = \frac{h_{ij}}{\max h_{ij}}, j = 1(1)n; i = 1(1)m. \tag{14}$$

And expressed it as $\bar{D} = (\bar{h}_{ij})_{m \times n}$.

- Step 4. Compute weight vector by applying the given formula

$$w_i = \frac{1 - \sum_{j=1}^n \bar{h}_{ij}}{m - \sum_{i=1}^m \sum_{j=1}^n \bar{h}_{ij}}, i = 1(1)m. \tag{15}$$

- Step 5. Chen and Tan [26] defines the score function δ of x as

$$\delta(x) = \mu_A(x) - \nu_A(x), \tag{16}$$

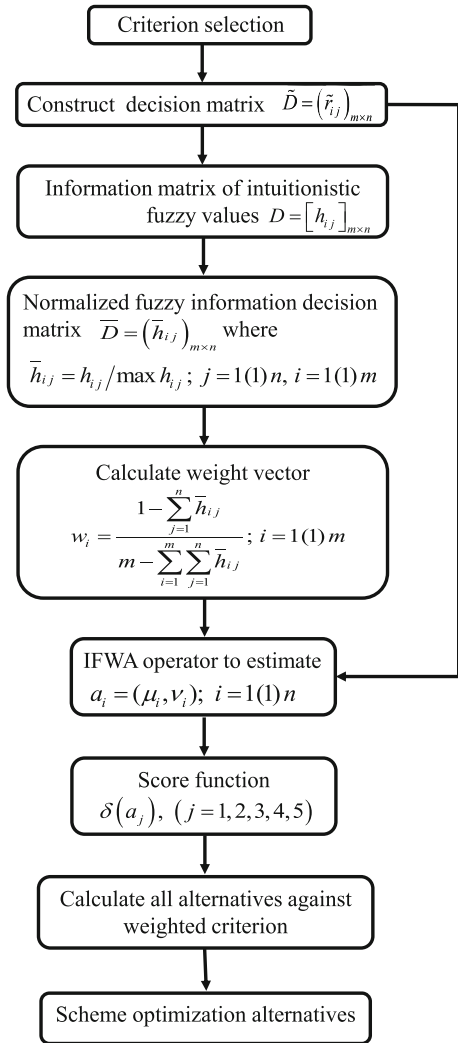
where $\delta(x) \in [-1, 1]$. For higher values of deviation $\delta(x)$ the membership value $\mu_A(x)$ should be high, while the nonmembership value should be low. The accuracy function h defined by Hong and Choi [27] to compute the degree of precision of the intuitionistic fuzzy value is

$$\eta(x) = \mu_A(x) + \nu_A(x), \tag{17}$$

where $\eta(x) \in [0, 1]$. The greater value of $\eta(x)$ results in the superior degree of membership of IFS. Let $\alpha_1 = (\mu_1, \nu_1)$, $\alpha_2 = (\mu_2, \nu_2)$ be two intuitionistic fuzzy values, then (Fig. 1)

- 1. for $\delta(\alpha_1) < \delta(\alpha_2)$, we obtain $\alpha_1 < \alpha_2$;
- 2. for $\delta(\alpha_1) = \delta(\alpha_2)$, we get if $\eta(\alpha_1) < \eta(\alpha_2)$, then $\alpha_1 < \alpha_2$; otherwise if $\eta(\alpha_1) = \eta(\alpha_2)$, then $\alpha_1 = \alpha_2$.

Fig. 1 Implementation flowchart of IFMWAO method



Step 6. Wu and Zhang [28] proposed the intuitionistic fuzzy weighted averaging operator as

$$IFWA_{\omega}(\alpha_1, \alpha_2, \dots, \alpha_n) = \left(1 - \prod_{i=1}^n (1 - \mu_i)^{\omega_i}, \prod_{i=1}^n \nu_i^{\omega_i} \right), \quad (18)$$

where ω_i is the weight of α_i , $\omega_i \in [0, 1]$ and $\sum_{i=1}^n \omega_i = 1$.

5.1 A Real Case Study

The above-defined method is illustrated mathematically in this subsection. Let there be a panel of five colleges of a district region $C_j(j = 1, 2, \dots, 5)$. The education institution has to take a decision based on the following four options: L_1 is the sports basic theoretical knowledge evaluation; L_2 is the physical fitness and sports skill evaluation; L_3 is the learning attitude evaluation; L_4 is the effective expression. The ranking of five colleges of a district region $C_j(j = 1, 2, \dots, 5)$ are calculated by the triangular intuitionistic fuzzy set on the above four options, and the decision matrix is constructed as (Table 2)

Step 1. The information measure is computed using formula (8) in the above judgment matrix and get the following information matrix:

$$D = (h_{ij})_{4 \times 5} = \begin{bmatrix} 0.7071 & 0.8090 & 0.7071 & 0.4343 & 0.6228 \\ 0.9511 & 0.7071 & 0.5090 & 0.8090 & 0.8655 \\ 0.6228 & 0.9511 & 0.5090 & 1.0000 & 0.8655 \\ 0.8090 & 1.0000 & 0.9511 & 0.7071 & 0.8090 \end{bmatrix}.$$

Step 2. Using (14), the above information matrix is transformed to the normalized information matrix as

$$\bar{D} = (\bar{h}_{ij})_{4 \times 5} = \begin{bmatrix} 0.7840 & 1.0000 & 0.8740 & 0.5386 & 0.7698 \\ 1.0000 & 0.7435 & 0.5352 & 0.8506 & 0.9100 \\ 0.6228 & 0.9511 & 0.5090 & 1.0000 & 0.8655 \\ 0.8090 & 1.0000 & 0.9511 & 0.7071 & 0.8090 \end{bmatrix}.$$

Step 3. Compute weight vector by utilizing the normalized information matrix and by the formula (14), we have

$$w_i = \frac{1 - \sum_{j=1}^n \bar{h}_{ij}}{m - \sum_{i=1}^m \sum_{j=1}^n \bar{h}_{ij}}, i = 1(1)m.$$

Hence, obtained as $W = (0.2455, 0.2475, 0.2401, 0.2669)$.

Table 2 Weighted intuitionistic fuzzy decision matrix

	C_1	C_2	C_3	C_4	C_5
L_1	(0.6, 0.3)	(0.5, 0.3)	(0.3, 0.6)	(0.7, 0.2)	(0.7, 0.3)
L_2	(0.5, 0.4)	(0.6, 0.3)	(0.6, 0.2)	(0.5, 0.3)	(0.4, 0.6)
L_3	(0.7, 0.3)	(0.5, 0.4)	(0.6, 0.2)	(0.4, 0.4)	(0.6, 0.4)
L_4	(0.5, 0.3)	(0.5, 0.5)	(0.4, 0.5)	(0.6, 0.4)	(0.5, 0.3)

Table 3 Ranking order alternatives for different methods

Method	Ranking	Best alternative
TOPSIS proposed by Grzegorzewski and Mrowka [29]	$C_1 \succ C_4 \succ C_3 \succ C_5 \succ C_2$	C_1
Intuitionistic fuzzy TOPSIS proposed by Joshi and Kumar [16]	$C_1 \succ C_3 \succ C_4 \succ C_5 \succ C_2$	C_1
Proposed IFMWAO method	$C_1 \succ C_3 \succ C_4 \succ C_5 \succ C_2$	C_1

Step 4. Obtain score function

By applying $IFWA_{\omega}$ operator (18), the set of intuitionistic fuzzy values $\alpha_j = (\mu_j, \nu_j)$, where $j = 1(1)n$, is obtained as

$$\alpha_1 = (0.5812, 0.3222), \alpha_2 = (0.4770, 0.3684), \alpha_3 = (0.5458, 0.3345), \\ \alpha_4 = (0.4596, 0.3142), \alpha_5 = (0.4941, 0.3816).$$

By (16), score function $\delta(\alpha_j)$, ($j = 1, 2, 3, 4, 5$) is as

$$\delta(\alpha_1) = 0.2590, \delta(\alpha_2) = 0.1086, \delta(\alpha_3) = 0.2113, \delta(\alpha_4) = 0.1454, \delta(\alpha_5) = 0.1125.$$

Step 5. Desirable ranking

The options are ranked in the descending order by comparing the score values. Here, we obtained the following ordering as $C_1 \succ C_3 \succ C_4 \succ C_5 \succ C_2$. Thus the most desirable college is C_1 .

The ranking of these five alternatives are also shown in the following Table 3.

It is found that there is no conflict in the preference ordering of all the alternatives by intuitionistic fuzzy TOPSIS method and developed technique. There is only one conflict found in deciding the preference ordering of C_3 and C_4 .

6 Conclusions

In this work, new intuitionistic fuzzy similarity and information measures are developed and analyzed the features of developed measures. Numerical results illustrate the proposed similarity and information measures are sensible and effective supplement of measures of IFSs. An intuitionistic fuzzy information method for MCDM problems has been proposed. A real case study to the rank five colleges of a district region based on four attributes using the proposed method is done.

Moreover, the results of the developed technique and the already existing method are compared. Consequently, it is concluded that our method reduces the complexity, while evaluating the results and hence it is more effective and efficient.

References

1. Zadeh, L.A.: Fuzzy sets. *Inf. Control* **8**, 338–353 (1965)
2. Grzegorzewski, P.: On possible and necessary inclusion of intuitionistic fuzzy sets. *Inf. Sci.* **181**, 342–350 (2011)
3. Atanassov, K.: Intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **20**, 87–96 (1986)
4. Mishra, A.R., Jain, D., Hooda, D.S.: Exponential intuitionistic fuzzy information measure with assessment of service quality. *Communicated* (2015)
5. Mishra, A.R., Jain, D., Hooda, D.S.: Intuitionistic fuzzy information with application in rating of township development. *Communicated* (2015)
6. Castineira, E.E., Cubillo, S., Montilla, W.: Measuring incompatibility between Atanassovs intuitionistic fuzzy sets. *Inf. Sci.* **180**, 820–833 (2010)
7. Liu, H.W., Wang, G.J.: Multi-criteria decision-making methods based on intuitionistic fuzzy sets. *Eur. J. Oper. Res.* **179**, 220–233 (2007)
8. Xu, Z.H.: Intuitionistic preference relations and their application in group decision making. *Inf. Sci.* **177**, 2363–2379 (2007)
9. Xu, Z.H., Chen, J., Wu, J.J.: Clustering algorithm for intuitionistic fuzzy sets. *Inf. Sci.* **178**, 3775–3790 (2008)
10. Szmidt, E., Kacprzyk, J.: Distances between intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **114**, 505–518 (2000)
11. Wang, W.Q., Xin, X.L.: Distance measure between intuitionistic fuzzy sets. *Pattern Recogn. Lett.* **26**, 2063–2069 (2005)
12. Wang, P.Z.: *Fuzzy Sets and Its Applications*. Shanghai Science and Technology Press, Shanghai (1983)
13. Wang, X.Z., De Baets, B., Kerre, E.: A comparative study of similarity measures. *Fuzzy Sets Syst.* **73**, 259–268 (1995)
14. Mitchell, H.B.: On the Dengfeng-Chuntian similarity measure and its application to pattern recognition. *Pattern Recogn. Lett.* **24**, 3101–3104 (2003)
15. Hwang, C.L., Yoon, K.S.: *Multiple Attribute Decision Making: Methods and Applications*. Springer, Berlin (1981)
16. Joshi, D., Kumar, S.: Intuitionistic fuzzy entropy and distance measure based TOPSIS method for multi-criteria decision making. *Egypt. Inf. J.* **15**, 97–104 (2014)
17. Lin, L., Yuan, X.H., Xia, Z.Q.: Multicriteria fuzzy decision-making methods based on intuitionistic fuzzy sets. *J. Comput. Syst. Sci.* **73**, 84–88 (2007)
18. Yue, Z.: Extension of TOPSIS to determine weight of decision maker for group decision making problems with uncertain information. *Exp. Syst. Appl.* **39**, 6343–6350 (2012)
19. Xu, Z.S.: Intuitionistic fuzzy aggregation operators. *IEEE Trans. Fuzzy Syst.* **15**(6), 1179–1187 (2007)
20. Szmidt, E., Kacprzyk, J.: Entropy for intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **118**, 467–477 (2001)
21. Vlachos, I.K., Sergiagis, G.D.: Intuitionistic fuzzy information—application to pattern recognition. *Pattern Recogn. Lett.* **28**, 197–206 (2007)
22. Wei, C., Zhang, Y.: Entropy measures for interval-valued intuitionistic fuzzy sets and their application in group decision making. *Math. Probl. Eng.* **2015**, 1–13 (2015)
23. Fan, H.: An approach to evaluating the physical education teaching quality of higher education institution with uncertain information. *JDCTA* **6**, 419–425 (2012)

24. Pang, M.: An approach to evaluating the physical education teaching quality of higher education institution with intuitionistic fuzzy information. *JCIT* **7**, 164–170 (2012)
25. Chen, Z.P., Yang, W.: A new multiple attribute group decision making method in intuitionistic fuzzy setting. *Appl. Math. Model.* **35**, 4424–4437 (2011)
26. Chen, S.M., Tan, J.M.: Handling multi-criteria fuzzy decision making problems based on vague set theory. *Fuzzy Sets Syst.* **67**, 163–172 (1994)
27. Hong, D.H., Choi, C.H.: Multicriteria fuzzy decision making problems based on vague set theory. *Fuzzy Sets Syst.* **114**, 103–113 (2000)
28. Wu, J.Z., Zhang, Q.: Multicriteria decision making method based on intuitionistic fuzzy weighted entropy. *Expert Syst. Appl.* **38**, 916–922 (2011)
29. Grzegorzewski, P., Mrowka, E.: Some notes on (Attnassov's) intuitionistic fuzzy sets. *Fuzzy Sets Syst.* **156**, 492–495 (2005)

Molecular Level Insight into the Interactions of SoxC and SoxD from Epsilonproteobacteria *Sulfurimonas denitrificans*: A Biomolecular Computational Approach

Sujay Ray, Arundhati Banerjee and Angshuman Bagchi

Abstract In deep seabeds, sulfide oxidation is essential for metabolism by microbes all through the seafloor. For the purpose, the organism *Sulfurimonas denitrificans* utilizes the gene cluster-*sox* (sulfur oxidizing). It comprises two units: *soxXYZAB* and *soxZYCD*. SoxCD complex formation is paramount for entire oxidation of sulfide and thiosulfate to sulfate. Herein for a computational molecular-level analysis, 3D models of SoxC and SoxD proteins were constructed by discrete molecular modeling techniques. Protein–protein docking generated SoxCD complex. Few stability calculating parameters revealed the simulated final protein complex as a highly stable one. Solvent accessibility value of the MD-simulated complex also disclosed it as the most interactive one. Identification of responsible amino acids for protein–protein interactions investigated that Asn145 from SoxD and His8 from SoxC played pivotal roles for the interactions to turn out stronger. Current study thereby provides a proposal for molecular mechanism and biophysical analysis of sulfur oxidation process to render a safe biota.

Keywords Sulfur oxidation · Molecular modeling · Sox operon · Protein–Protein interaction · Molecular docking · And energy optimization

S. Ray · A. Bagchi (✉)
Department of Biochemistry and Biophysics, University of Kalyani,
Kalyani, Nadia, India
e-mail: angshuman_bagchi@yahoo.com

S. Ray
e-mail: raysujay@gmail.com

A. Banerjee
Department of Biotechnology, National Institute of Technology,
Mahatma Gandhi Avenue, Durgapur, West Bengal, India
e-mail: arundhati.92star@gmail.com

1 Introduction

Hydrogen sulfide (H_2S) is one of the harmful, toxic pollutants in the environment. Therefore, the sulfur anion formed in the environment is generally converted to elemental form of sulfur (S^0) by oxidation process. The oxidation states of sulfur are -2 , 0 , $+2$, $+4$, and $+6$. Therefore, sulfur can perform different important bioinorganic reactions. Generally, microorganisms which survive in the deep ocean beds use compounds of sulfur that is reduced and inorganic in nature, like sulfides, for chemolithoautotrophy [1]. *Sulfurimonas denitrificans* is such a microorganism which is also an epsilonproteobacteria [1]. The molecular biochemistry for the oxidation process of sulfur anion is well established for organisms like *Paracoccus pantotrophus*, *Allochroamatium vinosum*, and others [2–5]. The metabolic reactions of sulfur anions by bacteria are generally classified into two different groups: (1) reverse sulfate reduction and (2) Sox multienzyme complex organization [6]. Among the two, the widely disseminated one is the “Sox multienzyme complex organization.” The operon called the *sox* operon encodes the Sox multienzyme system which comprises a cluster of sulfur oxidation (*sox*) genes [7–10].

However, it is noteworthy that the organism *Sulfurimonas denitrificans* (*S. denitrificans*) has a *sox* multienzyme complex, comprising *soxXYZAB* and *soxCD* [11]. Interestingly, the entire sulfur anion metabolism requires the active participation of proteins, namely SoxYZ complex, SoxXA complex, SoxCD complex, and the protein-SoxB [12]. In case of other microorganism, the SoxCD comprises 4 subunits by $\alpha_2\beta_2$, which acts as sulfur dehydrogenase [2]. SoxC protein is a molybdenum cofactor-containing protein. SoxD is a protein which has double heme groups (D_1 and D_2) and is a c-type cytochrome protein [13]. In order to know how the SoxCD protein complex functions in the *S. denitrificans* from epsilonproteobacteria molecular modeling studies of the two proteins SoxC accompanied by SoxD were performed. Using homology modeling (also known as comparative modeling) technique, the three-dimensional (3-D) structures of SoxC and SoxD proteins were prepared. The 3-D structures of the SoxCD protein complex were formed by biomolecular docking techniques. From the molecular docking analysis, the residual participation from the SoxCD protein complex interactions was explored. All the structures underwent overall minimization of energy. The structural strength or firmness was also studied after the consideration of their stereochemical properties. Therefore, this helps to study the genres of binding and the array of interaction in the SoxCD protein in *S. denitrificans* from epsilonproteobacteria.

2 Materials and Methods

2.1 *Sequence Analysis and Template Search for Homology Modeling*

The respective amino acid sequences of SoxC and SoxD proteins of *S. denitrificans* were obtained from NCBI (Accession No.: WP_011373674.1 for SoxC and WP_011373673.1 for SoxD). The identities of the proteins were identified using Uniprot KB for validation. Results from BLASTP [14] against PDB [15] and HH-Pred [15] inferred that the templates for building homology models were from the X-ray crystal structure of *Paracoccus pantotrophus* (PDB code: 2XTS_B with 52 % query coverage sharing 39 % sequence identity for SoxD and 2XTS_A with 85 % query coverage and sharing 45 % sequence identity for SoxC). It is known that HH-Pred is a very sensitive method. The method identifies the similar proteins and draws inclusions from remote sequence homologs [16].

2.2 *Molecular Modeling of SoxC and SoxD Monomers*

The models were built following homology modeling technique. Swiss Model software [17] in ExPasy web server was utilized for the purpose. After building the models they were superimposed on each of their crystal templates (i.e., B chain of 2XTS for SoxD, A chain of 2XTS for SoxC). 0.073 Å and 0.064 Å were the root mean squared deviations (RMSD) of the back bone superimpositions which was done by PyMOL for SoxD and SoxC respectively.

2.3 *Loop Optimization Using ModLoop*

The errors due to conformational disparity in the loop regions of the proteins were rectified using ModLoop. ModLoop remodeled the each of the structures via optimization [18].

2.4 *Refinement of the Modeled Protein Structures*

The structures of the modeled monomer proteins were further refined by ModRefiner [19]. The tool, ModRefiner helps to clear the unfavorable geometries in the prepared models.

2.5 Validation of the Models of the Proteins

The stereochemical profiles of the modeled structures were checked employing PROCHECK [20], Verify3D [21], and ERRAT [22]. The protein models were qualitatively good according to the predictions. Zero amino acid residues were observed in the unfavored regions of the Ramachandran Plots for each of the SoxC and SoxD proteins [23].

2.6 Molecular Docking Simulations

Cluspro2.0 [24] docking server aided to generate the 3-D model structure of SoxCD protein complex. Cluspro2.0 provided a total 10 docked SoxCD complexes. The structure having the best cluster size was opted for further analysis.

2.7 Energy Minimization of the Complex

With the use of Chiron Energy Minimization tool, the overall energy of the modeled SoxCD complex structure was minimized [25]. It follows CHARMM force field to minimize the overall energy of the docked complex [26]. The tool Chiron utilizes a high heat exchange rate of the individual proteins with the bath in simulations by Molecular Dynamics (MD). Therefore, Chiron not only removes the steric clashes rapidly, but also produces minimal disconcertion of the protein backbone.

2.8 Stability of the SoxCD Complex on Energy Minimization

As the overall optimization of energy would result into a better structural stability and interaction pattern of the protein complex, the determination of the structural stability of the modeled complex was inspected using FastContact server [27, 28]. Estimation of the net solvent accessibility area additionally analyzed to deduce a comprehensive result regarding the most stable complex (that is, complex structure pre- and post-energy minimization).

2.9 Calculation of Interaction Patterns and Binding Modes in the Complex

Protein Interaction Calculator (PIC) web server [29] was utilized to investigate and discern the protein–protein interaction pattern in SoxCD protein complex.

2.10 Electrostatic Surface Potential Assessment

The surface potential for the SoxCD complex was calculated pre- as well as post-energy minimization. The electrostatic potential on the surface of the proteins was generated in vacuum electrostatics with the aid of PyMOL.

3 Results

3.1 Model Description for the Monomer of SoxC

The functional SoxC protein from *S. denitrificans* after homology modeling was observed to be akin to the SoxC protein from the SoxCD protein complex belonging of *P. pantotrophus* (PDB Code: 2XTS; A chain). The protein contains α -helices (amino acid residues 3–6, 40–42, 67–69, 102–107, 124–127, 135–139, 153–159, 184–190, 219–221 and 356–363), β -sheets (amino acid residues 88–90, 98–100, 111–118, 142–151, 166–172, 179–183, 192–197, 211–213, 227–232, 247–249, 252–255, 264–269, 285–292, 299–304, 326–332, 339–347, 377–380 and 383–387) and coil regions. The structure is well presented in Fig. 1 with β -sheets and α -helices in magenta and cyan shades respectively with interspersing pink shaded coils.

3.2 Model Description for the Monomer of SoxD

The functional SoxD protein from *S. denitrificans* after homology modeling was observed to be akin to the SoxD protein from the SoxCD protein complex from *P. pantotrophus* (PDB Code: 2XTS; B chain). The protein monomer is made of mostly of α -helices (amino acids: 3–9, 27–42, 86–95, 107–121, 135–140, and 193–196) and coil regions. Two β -sheets, antiparallel in nature (amino acids: 24–26 and 132–134) were observed in this functional model of the protein. The structure is presented in Fig. 1 with blue and yellow shades illustrating α -helices and β -sheets, respectively, with interspersing red shaded coils.

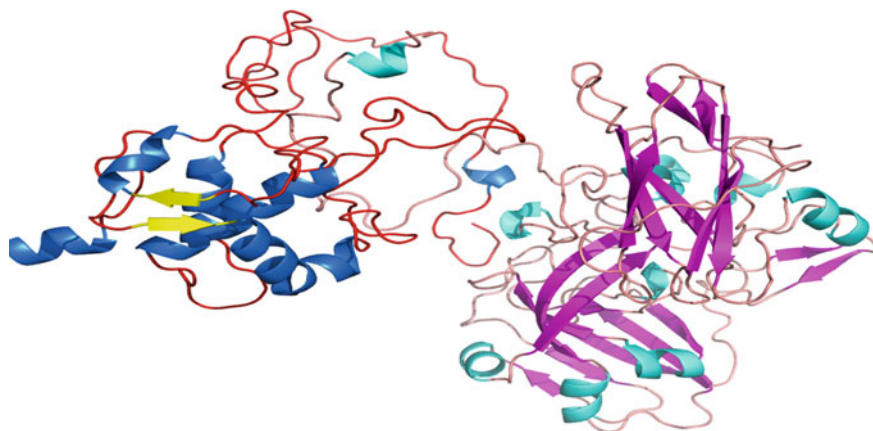


Fig. 1 The SoxCD docked complex with SoxC (*right*) showing α -helices (*cyan*), β -sheets (*magenta*) with interspersing *pink* coils and SoxD (*left*) showing α -helices (*blue*), β -sheets (*yellow*) with interspersing *red* coils (colour figure online)

3.3 Determination of the Stability of SoxCD Protein Complex

The change in net interaction energy values of the SoxCD protein complex and the energy minimized complex with MD simulations are well depicted in Table 1. It is clear from the table that the interaction energies at the interfaces between the proteins became stronger after the overall energy optimization of the modeled SoxCD complex. Moreover, from Table 1, the decrement in the value of net solvent accessibility for the complex supports the fact that the complex structure appeared to be more interactive after energy optimization.

3.4 Interactions in SoxCD Complex

The modeled complex structure of the SoxCD protein is depicted in Fig. 1. The analysis using PIC web server shows SoxC and SoxD intermingle strongly with one another via non-covalent forces. The principal interactive bonding were observed to

Table 1 Stability calculation of the SoxCD complex after Chiron simulation

Parameters/SoxCD complex structures	Before Chiron MD simulation	After Chiron MD simulation
Total interaction energy (kcal/mol)	-1115.102	-1190.738
Net solvent accessibility (\AA^2)	31618.33	31186.2

Table 2 Main chain–side chain bindings in the SoxCD complex

Position	Chain	Residue	Atom	Position	Chain	Residue	Atom
8	A	HIS	ND1	148	B	GLY	O
11	A	TRP	NE1	63	B	LEU	O
81	A	VAL	N	203	B	SER	OG
203	A	ARG	NH1	197	B	ALA	O
203	A	ARG	NH1	197	B	ALA	O
61	B	LYS	NZ	4	A	ASN	O
82	B	TYR	OH	13	A	THR	O
82	B	TYR	OH	14	A	SER	O
145	B	ASN	ND2	12	A	GLY	O
145	B	ASN	ND2	12	A	GLY	O
203	B	SER	OG	81	A	VAL	O

Table 3 Side chain–side chain bindings of the SoxCD complex

Position	Chain	Residue	Atom	Position	Chain	Residue	Atom
8	A	HIS	NE2	145	B	ASN	OD1
8	A	HIS	NE2	145	B	ASN	ND2
145	B	ASN	OD1	8	A	HIS	NE2
145	B	ASN	OD1	8	A	HIS	NE2
145	B	ASN	ND2	8	A	HIS	NE2
145	B	ASN	ND2	8	A	HIS	NE2

be accomplished by hydrogen bonding (H-bonding) and ionic–ionic interactive bonding. The main–chain side chain and side chain–side chain hydrogen bonds are predominant between the protein complex. Tables 2 and 3 represent the protein–protein different H-bonding interactions between the SoxCD protein complex.

3.5 Electrostatic Potential Calculation

Interestingly, the electrostatic potential calculations using vacuum also presents the energy minimized complex structure to be highly stable and interactive one. Figure 2 depicts the pictorial view for the comparative study of the electrostatic potentials for the complex structures before and after energy minimization. Blue areas on the complex surface represent the regions that are electrostatically positive, whereas red areas represent the regions which are electrostatically negative ones in either of the two cases.

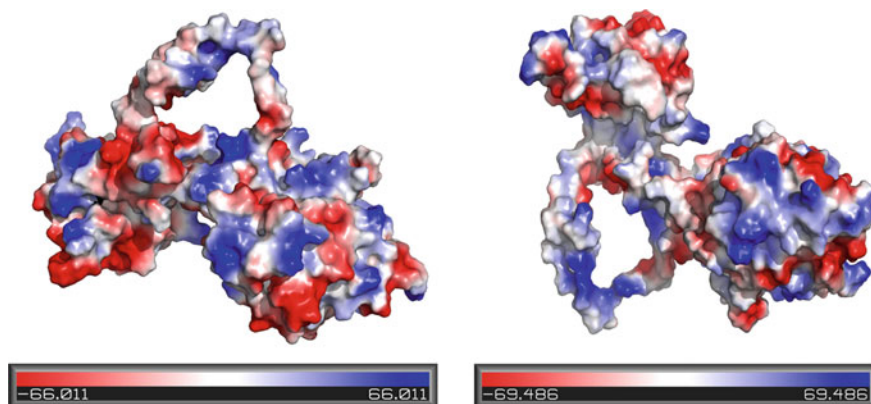


Fig. 2 Comparative view of the electrostatic potential change on the surfaces before (*left*) and after (*right*) energy minimization and simulation of SoxCD protein complex

4 Discussion

In the extant work, the modeled functional, 3D modeled protein structures of the SoxC and SoxD were built and analyzed. The different interaction pattern in SoxCD complex were demonstrated, evaluated, and analyzed. After energy optimization, the interaction energies were observed to get declined from -1115.102 kcal/mol (prior to energy optimization) to -1190.738 kcal/mol. As a support, the net solvent accessibility value also got a descend from 31618.33 to 31186.2 Å² after minimizing the complex structure. The electrostatic potential values also ensure that the complex structure after minimization and simulation to be an exceedingly firm, steady, and interactive one. From the hydrogen bond interactions, it was revealed that His8 from SoxC and Asn145 from SoxD play vital roles in forming the SoxCD protein complex structure.

Therefore, the extant probe provides acquaintance in the interactions between SoxC and SoxD proteins from *S. denitrificans* which is an epsilonproteobacteria. There were previous documentations by investigations for the collaborations of SoxCD proteins from several classes of bacteria but this serves as the first molecular and computational exploration from *S. denitrificans*, an epsilonproteobacteria which targets directly the sulfur oxidation phenomenon beneath the sea—occupying the bulk part on earth's surface. It thereby renders a toxic-free environment.

5 Conclusion

The cooperation through binding between SoxC and SoxD was analyzed in this work. The proteins were taken from *S. denitrificans*, which is an epsilonproteobacteria. The functional modeled tertiary structures of SoxC and SoxD were

modeled and examined. Their interactions along with the interacting residues were analyzed and demonstrated. The stability and most interactive state for the binding in SoxCD complex were also inspected. This is a study for the disclosure of the participation of SoxC and SoxD proteins from *S. denitrificans*, an epsilonproteobacteria. Though there are prior documentations for the cooperation of SoxCD proteins from other classes of bacteria, but this probe provides the first residue-level approach to promulgate the SoxCD complex participation from *S. denitrificans* which is an epsilonproteobacteria.

Acknowledgments The authors would like to thank the DST-PURSE programme 2012–2015 going on in the Department of Biochemistry and Biophysics, University of Kalyani for providing different instrumental and infrastructural support. The author is also thankful to the DBT sponsored Bioinformatics Infrastructure Facility in the Department of Biochemistry and Biophysics, University of Kalyani for the necessary support.

References

1. Nancy, H., Akerman, D., Butterfield, A., Huber, J.A.: Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subsurface *Epsilonproteobacteria* in diffuse hydrothermal vent fluids. *Front Microbiol.* **4**, 185 (2013)
2. Friedrich, C.G.: Physiology and genetics of sulfur-oxidizing bacteria. *Adv. Microb. Physiol.* **39**, 235–289 (1998)
3. Kletzin, A., Urich, T., Muller, F., Bandejas, T.M., Gomes, C.M.: Dissimilatory oxidation and reduction of elemental sulfur in thermophilic archaea. *J. Bioenerg. Biomembr.* **36**, 77–91 (2004)
4. Friedrich, C.G., Bardischewsky, F., Rother, D., Quentmeier, A., Fischer, J.: Prokaryotic sulfur oxidation. *Curr. Opin. Microbiol.* **8**, 253–259 (2005)
5. Frigaard, N.U., Dahl, C.: Sulfur metabolism in phototrophic sulfur bacteria. *Adv. Microb. Physiol.* **54**, 103–200 (2009)
6. Kappler, U., Dahl, C.: Enzymology and molecular biology of prokaryotic sulfite oxidation. *FEMS Microbiol. Lett.* **203**, 1–9 (2001)
7. Friedrich, C.G., Quentmeier, A., Bardischewsky, F., Rother, D., Kraft, R., Kostka, S., Prinz, H.: Novel genes coding for lithotrophic sulfur oxidation of *Paracoccus pantotrophus* GB17. *J. Bacteriol.* **182**, 4677–4687 (2000)
8. Mukhopadhyaya, P.N., Deb, C., Lahiri, C., Roy, P.: A *soxA* gene, encoding a diheme cytochrome *c*, and a *sox* locus, essential for sulfur oxidation in a new sulfur lithotrophic bacterium. *J. Bacteriol.* **182**, 4278–4287 (2000)
9. Rother, D., Henrich, H.J., Quentmeier, A., Bardischewsky, F., Friedrich, C.G.: Novel genes of the *sox* gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in *Paracoccus pantotrophus* GB17. *J. Bacteriol.* **183**, 4499–4508 (2001)
10. Appia-Ayme, C., Berks, B.C.: SoxV, an orthologue of the CcdA disulfide transporter, is involved in thiosulfate oxidation in *Rhodovulum sulfidophilum* and reduces the periplasmic thioredoxin SoxW. *Biochem Biophys Res Commun.* **296**, 737–741 (2002)
11. Ghosh, W., Dam, B.: Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiol. Rev.* **33**(6), 999–1043 (2009)

12. Friedrich, C.G., Rother, D., Bardischewsky, F., Quentmeier, A., Fischer, J.: Oxidation of reduced inorganic sulfur compounds by bacteria: emergence of a common mechanism? *Appl. Environ. Microbiol.* **67**, 2873–2882 (2001)
13. Zander, U., Faust, A., Klink, B.U., de Sanctis, D., Panjikar, S., Quentmeier, A., Bardischewsky, F., Friedrich, C.G., Scheidig, A.J.: Structural basis for the oxidation of protein-bound sulfur by the sulfur cycle molybdohemo-enzyme sulfane dehydrogenase SoxCD. *J. Biol. Chem.* **286**(10), 8349–8360 (2011)
14. Altschul, S.F., et al.: Basic local alignment search tool. *J. Mol. Biol.* **25**, 403–410 (1990)
15. Berman, M.H., et al.: The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000). doi:[10.1093/nar/28.1.235](https://doi.org/10.1093/nar/28.1.235)
16. Söding, J., Biegert, A., Lupas, A.N.: The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005). Web Server issue. doi:[10.1093/nar/gki408](https://doi.org/10.1093/nar/gki408)
17. Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T.: SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **42**(W1), W252–W258 (2014)
18. Fiser, A., Sali, A.: ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **19**(18), 2500–2501 (2003)
19. Xu, D., Zhang, Y.: Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophys. J.* **101**, 2525–2534 (2001). doi:[10.1016/j.bpj.2011.10.024](https://doi.org/10.1016/j.bpj.2011.10.024)
20. Laskowski, R.A., et al.: PROCHECK: a program to check the stereochemistry of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993)
21. Eisenberg, D., et al.: VERIFY3D: Assessment of protein models with three-dimensional profiles. *Methods Enzymol.* **277**, 396–404 (1997)
22. Colovos, C., Yeates, T.O.: Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* **2**, 1511–1519 (1993)
23. Ramachandran, G.N., Sashisekharan, V.: Conformation of polypeptides and proteins. *Adv. Protein Chem.* **23**, 283–438 (1968)
24. Comeau, S.R., et al.: ClusPro: An automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* **20**, 45–50 (2004)
25. Ramachandran, S., Kota, P., Ding, F., Dokholyan, N.V.: PROTEINS: Structure. *Funct. Bioinform.* **79**, 261–270 (2011)
26. Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M.: CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**(2), 187–217 (1983)
27. Maleki, M., Vasudev, G., Rueda, L.: The role of electrostatic energy in prediction of obligate protein-protein interactions. *Proteome Sci.* **11**, S11 (2013). doi:[10.1186/1477-5956-11-S1-S11](https://doi.org/10.1186/1477-5956-11-S1-S11)
28. Camacho, C.J., Zhang, C.: FastContact: rapid estimate of contact and binding free energies. *Bioinformatics* **21**(10), 2534–2536 (2005)
29. Tina, K.G., Bhadra, R., Srinivasan, N.: PIC: protein interactions calculator. *Nucleic Acids Res.* **35**, W473–W476 (2007)

“Jugaad”—The Creativeness for Selection of Software Development Methodology Advisory System—Fuzzy Expert System

Kardile Vilas Vasantao and Chitra G. Desai

Abstract Globalization and technical revolution are raising several challenges to the software development sector. Over the past 50 years, software has evolved as a specialized problem-solving and information analysis tool to an industry, but now it is facing multiple challenges. The objective of this study is to landscape current knowledge, in terms of productivity and find out its impact on the software development. To resolve such problems, “Software Development Practitioner” needs to find out a flexible way (Jugaad) for development. “**Jugaad**” is the Indian common man’s philosophy to achieve the dream goal within the available resources. The researcher has made earnest attempts to study the steps to be undertaken to make the flexible module. The goal of this paper is to rectify the present hurdles and hassles in development approaches by representing “Selection of Software Development Methodologies Advisory System” on the basis of reference of fuzzy expert system.

Keywords Productivity of software development • Factor affecting productivity • JUGAAD • Fuzzy expert system • Uncertainty • Selection of software development advisory system

1 Introduction

Since 1960 software has become a part of society and has been bringing revolutions in the sector. But still today, software development organization is facing the problem of failures [1–3]. It is quite natural that one can recognize the question

K.V. Vasantao (✉)

Computer Science Department, Tuljaram Chaturchand College, Baramati (Pune), Maharashtra, India
e-mail: vilaskardile@gmail.com

C.G. Desai

Department of Computer Science, NDA, Khadakwasala, Pune, Maharashtra, India
e-mail: chitragdesai@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_40

411

‘Why is the development productivity trip down?’ With this inspiration, the study is carried out to rectify the failure part. The researcher has made workaholic ardent efforts by the way of distinctive permutations and combinations of variables in technological factors. Researcher has reviewed the relevant literature of software development productivity and analyze its impact, and reformulate the customized software on the basis of “Jugaad.”

This paper is organized in 5 sections.

- Section 1 deals with the introduction of title of this paper.
- Section 2 deals with software development productivity, affecting factors, and strategies for enhancing the productivity process.
- In Sect. 3, it deals with methodology and observation.
- The Sect. 4 deals with discussions supported by implications, interpretations, and suggestions.
- The conclusion and future scope of the study are depicted in Sect. 5.

2 Software Development Productivity

2.1 Constraints in Software Development Productivity

The basic concept of the system is the *output*, which depends mainly on two aspects, one is ‘participant’ and another is ‘processes.’ So it is natural that, if ‘software project productivity’ is the ‘output,’ then it must depend on the ‘projects participant’ and ‘its ‘process’(developing methodology). The study reveals that every project is unique and certainly has successful option (process). But it is difficult to define, which process is fit for productivity, since each one has pitfalls [4, 5]. Such type of environment increases the level of uncertainty and complexity. Hence it is one of the strong causes and constrain for trip down software project.

2.2 Challenges for Software Development Process

It is a necessity to utilize suitable methodology to enhance software development productivity [4]. Because, ‘Technology never fails but it will fails due to utilization of its, in opponent opportune’ [6]. In such situation, there may be an inaccurate handling of design approach or hasty decision, which may prove to be wrong later. In the ratio of challenging state, primary survey indicates that such inaccurate occurrence takes place in software development stage.

In such an uncertain situation, a million dollar question is “How development practitioner would formulate the development methodology on the basis of suitability or requirement?”

2.3 Opportunity for Software Development Process

Based on this, the study suggests the development practitioner to be more flexible like “Jugaad”—the philosophy used by people in India, Pakistan, and South Asian countries. The word originated as **Jugaad**, (Hindi: जुगाड़) is a Hindi-Urdu term. These people applied this term “*Jugaad*” to a creative or innovative idea providing a quick, alternative way of solving or fixing a problem with available resources on the basis of feasible and flexible considerations [7]. There are various definitions available in different manner and reflect the underlying scientific philosophical way of thinking and disciplines, but here the researcher take it as ‘**Just Utilization Gaining Approach And Deploy**’ to solve the problems.

Thus our hypothesis is “Appropriate formation of the project development and methodology on the basis of suitability is not puzzling.”

3 Methodology and Observation

In this section, the methodology of the study has been discussed. The inductive approach with qualitative solution is utilized with prototype model “Selection of Software Development Methodology Advisory System” on the basis of published literature of fuzzy expert system and its application [7–20, 22]. Purpose of this module is to understand causes of uncertainty and intensity of the project. And there formation of suitable approaches modules to fit as the best practice (Figs. 1 and 2).

3.1 Prepare Input Set for SoSDMAS

Initialization with knowledge, stability, and time (Fig. 3).

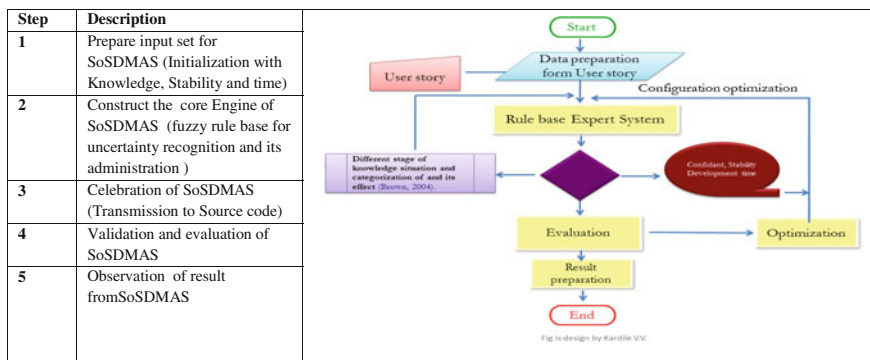


Fig. 1 Algorithm and flowchart of SoSDMAS

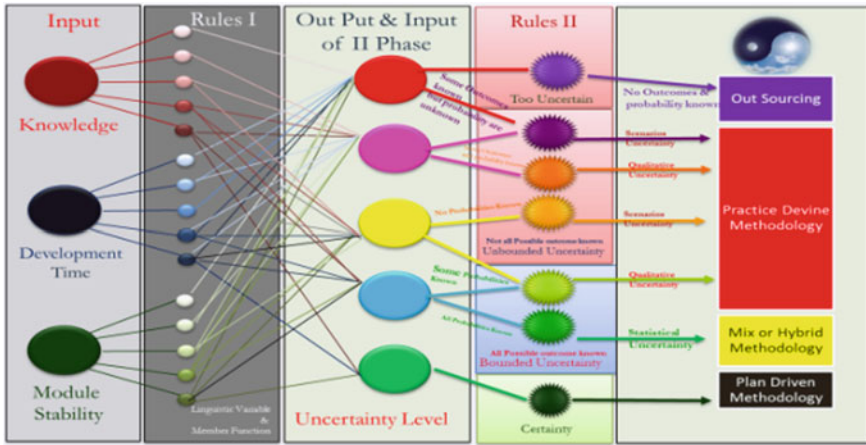


Fig. 2 Algorithm and flowchart of SoSDMAS

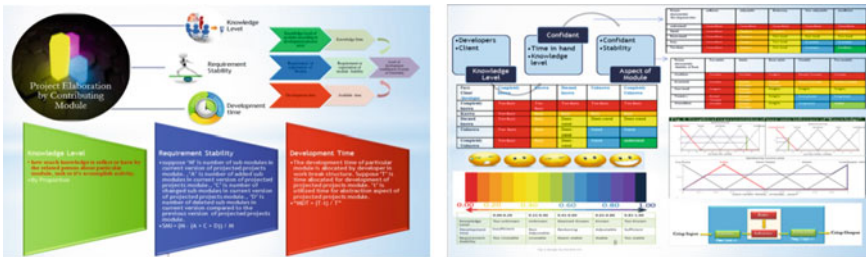


Fig. 3 Input set for SoSDMAS—Fuzzy expert system (initialization with knowledge, stability, and time)

3.2 Construct the Core Engine of SoSDMAS

Fuzzy rule base for uncertainty recognition and its administration (Fig. 4).

3.3 Celebration of SoSDMAS

Transmission to Source code: This prototype module is constructed by using platform (Trial version of Visual Basic 6.0) and implemented only for projected study purpose (Fig. 5).

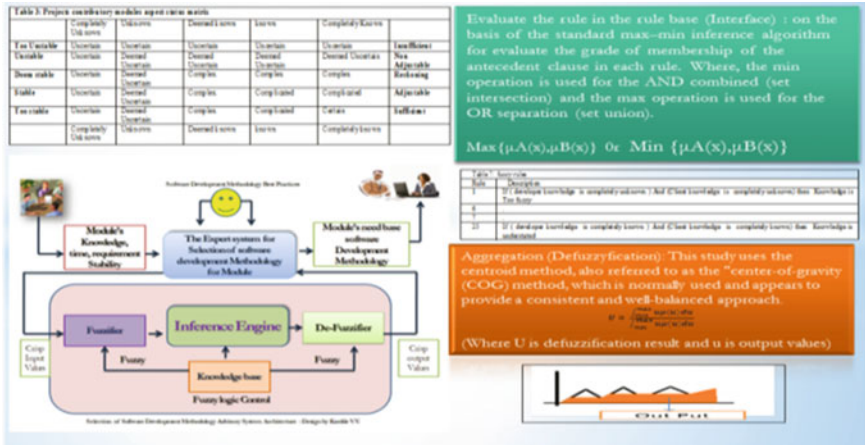
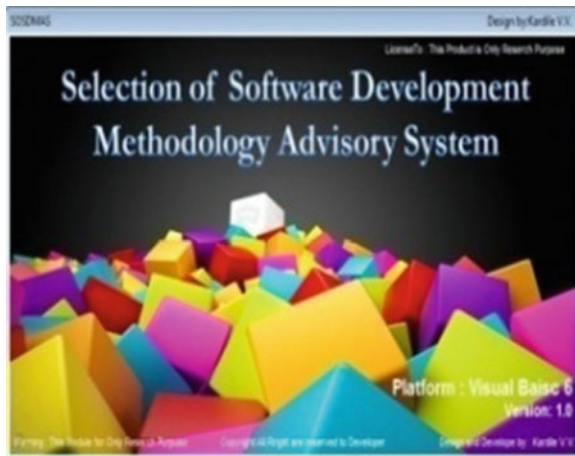


Fig. 4 Construct the core engine of SoSDMAS

Fig. 5 SoSDMAS screen



3.4 Validation and Evaluation of SoSDMAS

This module validation and evaluation are done by the utility of the system according to the easiness of interaction, capabilities, efficiency, speed, reliability, and accuracy of results which are verified with mathematical solution by implementer and two experts of fuzzy expert system with ethical consideration.

3.5 Experimental Case Study

To simulate and validate **SoSDMAS** let us consider, User specifies a requirement of “Log in” screen module application that generates and maintains user authentication of administration. It collects the password of user and validates it. It allows the new user to add informations in detail. Provides facility to recover the forgotten password by one- or two-step verification like the following screen as a project (Fig. 6).

This module has three subactivities as: Login facility, sign up, and forgot password with own aspect. As per requirement, development practitioner surly celebrated this module with practice-driven approach with outward appearance of agile development methodologies. Due to its subactivity that facilitates to recover the forgotten password by one- or two-step verification. One-step verification celebrates some hint questions or alternative mailing address. In two-step verification, it will require alternative service like mobile number, or it can regenerate randomly and received by the head of department. It is not confirm, but when we concentrate over remaining two subactivities, it evidently celebrates certain aspects.

As a resultant, project may be success in development with chances of overrun due to two modules having certain aspects that will also be developed by practice-driven instead of plan-driven.

Let us evaluate the above case with “Selection of software development methodology advisory system (SoSDMAS).” We elaborate this module in its contributory activity as login facility, sign up, and forgot password.

3.5.1 Preparation of Input

As per the flow of SoSDMAS, input data preparation from user stories gather by fact finding techniques (as per knowledge level, stability, and development time) on the basis of uncertainty aspect.

(A) Knowledge Level of Module: In this consideration, knowledge level of the module prepared by proportion values of user’s story and data collection sources like project sponsor, starring committee, project manager, and other personal associations and/or affected by the project.

In the present example, researchers collect data from 10 respondents with 8 questions and 6 options. The exhaustive worksheet is showed in Table 1.

The image shows a login form with a light gray background. It contains two input fields: 'User Id' and 'Password', each with a white box and an orange border. Below the input fields are three buttons: 'Forgot Password' (white with red border), 'New User' (white with red border), and 'LOG IN' (red with white text).

Fig. 6 Illustrative project require screen

Table 1 Master data

Sr. no	Respond	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
1	1	2	2	1	2	2	2	2	2
2	2	2	1	2	4	2	1	2	2
3	3	2	1	1	1	2	1	1	2
4	4	2	1	1	2	1	1	2	1
5	5	2	1	2	3	1	2	2	2
6	6	2	1	1	1	3	1	1	2
7	7	2	2	2	2	2	1	2	1
8	8	2	1	1	2	3	1	1	2
9	9	2	1	1	3	1	1	2	2
10	10	2	2	2	4	1	2	2	1

Table 2 A question and its optional proportion response

Sr. no	Total respond	Question	Actual respond						Desired respond	Total
			O1	O2	O3	O4	O5	O6		
1	10	1	0	1	0	0	0	0	O2	1
2	10	2	0.7	0.3	0	0	0	0	O1	1
3	10	3	0.6	0.4					O1	1
4	10	4	0.2	0.4	0.2	0.2			O2	1
5	10	5	0.3	0.4	0.3				O2	1
6	10	6	0.7	0.3					O1	1
7	10	7	0.3	0.7					O2	1
8	10	8	0.3	0.7					O2	1

Each question’s propositional response and desired option (calculate formula of Proportion = n/N , where n is number of response for option, N is total number of response for question) (Table 2).

Modulewise question and its optional response (Table 3)

Modulewise knowledge level (Table 4)

(B) Requirements Stability of the Module: The requirement stability of the module describes itself how much requirement is stable. In this concern, there is already a defined “Software Maturity Index model” [17] (A metric in IEEE 982.1-1988, see http://www.standards.ieee.org/reading/ieee/std_public/description/se/982.1-1988_desc.html). This model measures the software product stability. On the basis of this model, if project is replaced in subproject or module then it will measure the module stability. So, modified model is as follows:

In this module, suppose ‘ M ’ is the number of submodules in the current version of projected projects module, ‘ A ’ is the number of added submodules in the current version of projected projects module, ‘ C ’ is the number of changed submodules in the current version of projected projects module, ‘ D ’ is the number of deleted

Table 3 Modulewise question and its optional proportion response respond

Sr. no	Module	Question	O1	O2	O3	O4	O5	O6	Total
1	M1	1	0	10	0	0	0	0	10
2		2	7	3	0	0	0	0	10
3	M2	3	6	4					10
4		4	2	4	2	2			10
5		5	4	4	3				10
6	M3	6	7	3					10
7		7	3	7					10
8		8	3	7					10

Table 4 Modulewise knowledge level

Sr. no	Module	Total	Knowledge level
1	M1	0.85	Completely known
2	M2	0.47	Deemed known or unknown
3	M3	0.7	Known

submodules in the current version compared to the previous version of projected projects module.

$$SMI = (M - (A + C + D))/M \tag{1}$$

Or More perspicuously, $SMI = 1 - N/M,$ (2)

where M is the total number of submodules in the current version of the projected projects module and N is the number of submodules added, changed, or deleted between the previous version and this one.

SMI can be a measurement of product stability, when SMI approaches 1.0, the product is completely stable and 0.0 is completely unstable. It can be classified in the range as 0.0–0.20 too unstable, 0.21–0.40 unstable, 0.41–0.60 deemed stable, 0.61–0.80 stable, and 0.81–1.00 too stable.

Modulewise stability level (Table 5)

When correlated with the development to accomplish the complete project, developer identifies modules aspect.

(C) Development Time: The development time of particular module is allocated by the developer in work break structure.

Table 5 Modulewise stability level

Module	Description	A	C	D	A + C + D	(M - (A + C + D))/M	Stability
Module 1	Log in	0	0	0	0	(1-0)/1 = 1	Too stable
Module 2	New user	0	1	1	2	(3-2)/3 = 0.35	Unstable
Module 3	Forgot password	1	0	0	1	(2-1)/2 = 0.50	Deemed stable

Suppose ‘ T ’ is the time allocated for development of projected projects module, ‘ t ’ is the utilized time for abstraction aspect of projected projects module. Then time in hand for development of projected projects module or module development time is

$$MDT = (T - t)/T. \tag{3}$$

MDT can be a measurement of development time, when MDT approaches 1.0, the development time is too sufficient and 0.0 is insufficient. It can be classified as 0.0–0.20 too insufficient, 0.21–0.40 nonadjustable, 0.41–0.60 revoking, 0.61–0.80 adjustable, 0.81–1.00 sufficient.

Modulewise development time in hand level (Table 6).

Furthermore, with extracting above consideration in state of knowledge and its stability for solving problems, and development time availability, it generates a rule-based matrix as follows (Table 7).

3.5.2 SOSDMAS Evaluation

Furthermore, above matrix table is extracted in three states separately for constructing fuzzy expert system on the basis of type I method (Fig. 7). As

Table 6 Modulewise stability level

Module	Allocated time	Utilized time	Module development time is	Development time in hand
1	10	0	$(10-0)/10 = 1$	Sufficient
2	15	7	$(15-7)/15 = 5.33$	Revoking
3	15	5	$(15-5)/15 = 0.66$	Adjustable

Table 7 Projects contributory modules aspect status matrix

	Completely unknown	Unknown	Deemed known	Known	Completely known	
Too unstable	Uncertain	Uncertain	Uncertain	Uncertain	Uncertain	Insufficient
Unstable	Uncertain	Deemed uncertain	Deemed uncertain	Deemed uncertain	Deemed uncertain	Non adjustable
Deem stable	Uncertain	Deemed uncertain	Complex	Complex	Complex	Reckoning
Stable	Uncertain	Deemed uncertain	Complex	Complicated	Complicated	Adjustable
Too stable	Uncertain	Deemed uncertain	Complex	Complicated	Certain	Sufficient
	Completely unknown	Unknown	Deemed known	known	Completely known	

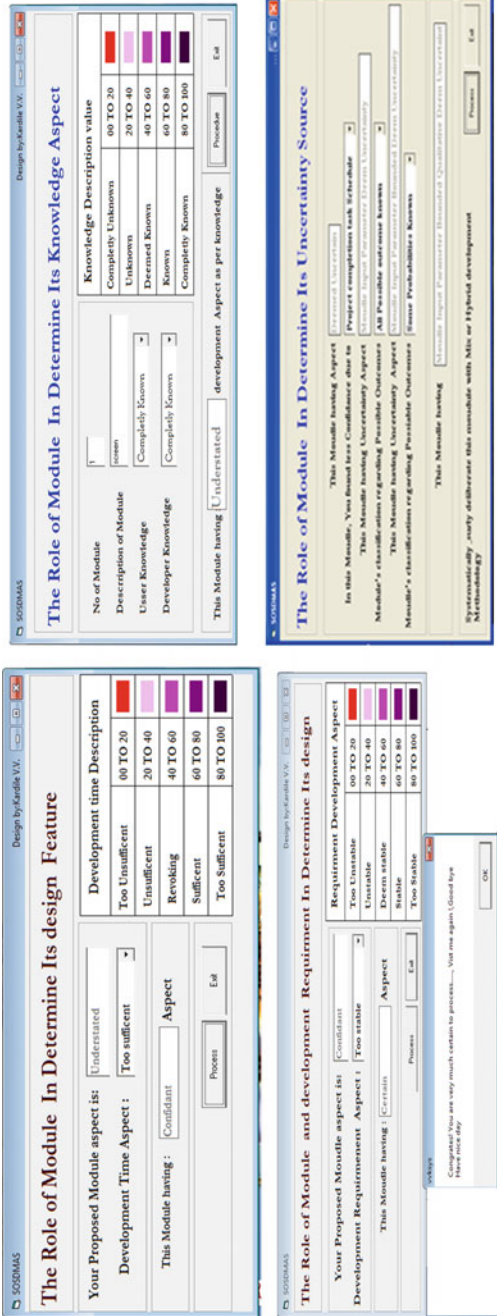


Fig. 7 SOSDMS screen

3.6 Observation of Result from SoSDMAS

Result by exploiting above “Selection of software development methodologies Advisory system” model specifies development approach as per its best practice and project’s uncertainty level. The pilot test is performed for this module with 10 postgraduate students from master’s in computer science and five software professionals.

4 Discussion

4.1 Implications

The study had three contributory modules as shown in Table 8, which have been tested on ‘Uncertainty Level’ along with other variables as shown in the above table. The analysis depicts that, the high level of ‘uncertainty’ project may not have better productivity rather it is time-consuming and expect stability in functioning. The module No. 1 above is the best example which works on plan drive rather than practice. Likewise, the other two modules also has the ‘Uncertainty Level’ but the intensity is ‘uncertain and deemed uncertain’ because of this peculiar characteristic the productivity in the project development is better since it is practice-driven and other one is the mix of plan and practice drive. Overall, the flexible approach will give better productivity since it is practical-oriented.

4.2 Interpretation

The study reveals that the project always has diverse aspects and uniqueness. But that can be customized (developed) as per the need of the customers and best practice of software for better productivity, by using various modules and distinctive permutations of variables in the technology (Fig. 8).

Table 8 Modulewise knowledge level

Sr. no	Module	Knowledge level	Stability level	Development time	Uncertainty level	Suggested development methodology as per best practice
1	M1	0.85	1	1	Certain	Plan driven
2	M2	0.47	0.35	0.53	Uncertain	Practice driven
3	M3	0.7	0.50	0.66	Deemed uncertain	Mix approach
Complete project		$2.02/3 = 0.67$	$1.85/3 = 0.61$	$2.19/3 = 0.73$	Complex	Plan driven

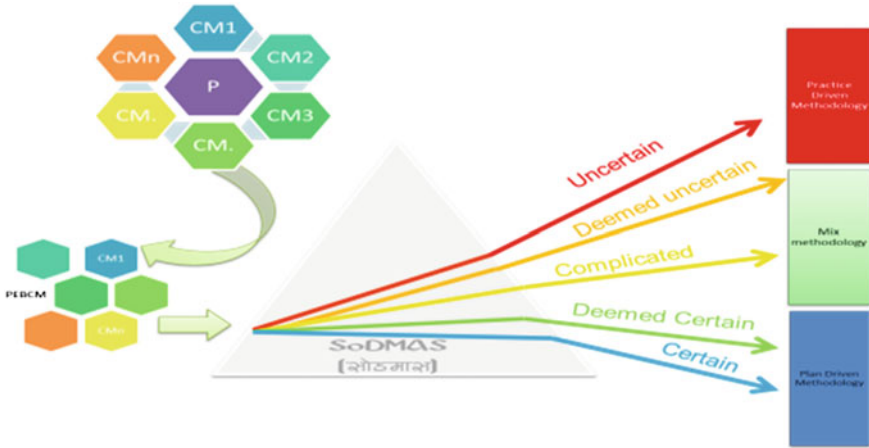


Fig. 8 SoSDMAS output

4.3 Suggestions

The configuration of ‘Project Modules’ and ‘Software Development Methodology’ should be on the need base, meaning flexible. Application of such project or software development certainly enhances the productivity. But to establish such kind of configuration, it requires opportunity of deemed uncertain environment as discussed in Table 8.

5 Conclusion

In this section, this study hope and trust that SoSDMAS is very useful to avoid confusion in formation of project and its developing methodology as per project aspect and best practice. It should assist the development practitioner for needbase formation by taking into account the certainty or uncertainty aspect separately.

5.1 Scope of Further Work

As per literature review, it is yet to publish notification, which can explore the policy for decreasing the ratio of project challenging and cost overrun/time element. This study found that there is a need to recognize uncertainty before modeling process that will assist the developer to choose and design appropriate software development methodology for efficient development.

References

1. Kardile, V.V.: Enhance accuracy in software development’s planning & estimation process by adopt “uncertainty analysis and assessment” in the system modeling process a review (2012). ISBN:978-1-4244-8677-9
2. Kardile, V.V.: Need to understand uncertainty in system development modelling process. IEEE (2011). ISBN:978-0-7695-4437-3
3. Ambler, S.W.: Defining Success, by. Dr. Dobb’s Journal. Source :2014 IT project success survey (2014). www.ambaysoft.com/surveys/success2014.html
4. Boehm, B.W.: Improving software productivity. IEEE Comput. **20**(8), 43–58 (1987)
5. Kardile, V.V.: Understanding need of “uncertainty analysis” in the system design process. Int. J. Soft. Eng. Appl. (IJSEA), vol. 2, 3 July 2011. ISSN: 0975-9018
6. Kardile, V.V.: Understanding the need of flexible software development approach using economic model IEEE: CFP1195-PRT (2011). ISBN:978-1-4244-8677-9
7. Radjou, N., Prabhu, J., Ahuja, S.: Jugaad Innovation. Jossey-Bass, A Wiley Imprint (2012)
8. Chen, J.: An analytical theory of project investment: a comparison with real option theory. Int. J. Manag. Financ. (2006)
9. Ngai, E.W.T., Wat, F.K.T.: Design and development of a fuzzy expert system for hotel selection (2003)
10. Klauer, B., Brown, J.D.: Conceptualising imperfect knowledge in public decision making: ignorance, uncertainty, error and ‘risk situations’. Env. Res. Eng. Manag. **27**(1), 124e128 (2004)
11. Kumru, M.: Assessing the visual quality of sanitary ware by using fuzzy logic. Appl. Soft Comput. **13**(8), 3646–3656 (2013)
12. Ayman Al Ahmar, M.: Rule based expert system for selecting software development methodology. J. Theor. Appl. Inf. Technol. © 2005–2010, jait&lls
13. Goodarxi, M.H., Rafe, V.: Educational advisor system implemented by web based fuzzy expert system. Asian J. Inf. Technol. **11**(2), 77–82 (2012). J. Soft. Eng. Appl. (Medwell Journals 2012) **5**, 500–507. doi:10.4236/jsea.2012.57058. Published July 2012. <http://www.SciRP.org/journal/jsea>
14. Fard, M.T.T., Jafari, H.R., Shojaie, S.E.: Analysis expert system factors in IT project selection using Fuzzy AHP. Interdisc. J. Contemp. Res. Bus. copy right © 2012. Institute of Interdisciplinary Business Research 512, **4**(8) Dec 2012. jcrb.webs.com
15. Fenton, N.E., Neil, M.: Software metrics: successes, failures and new directions. J. Syst. Soft. **47**, 149–157 (1999)
16. Refsgaard, J.C., van der Sluijsb, J.P., Højberga, A.L., Vanrolleghemc, P.A.: Uncertainty in the environmental modelling process, a framework and guidance. Environmental Modeling & Software **22** (2007)
17. Software Maturity Index: www.c2.com/cgi/wiki?SoftwareMaturityIndex. A metric in IEEE 982.1-1988, see http://www.standards.ieee.org/reading/ieee/std_public/description/se/982.1-1988_desc.html
18. Walker, W.E., Harremoës, P., Rotmans, J., Van der Sluijs, J.P., Van Asselt, M.B.A., Janssen, P., Kreyer von Krauss, M.P.: Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. Integrated Assessment **4**(1), 5e17 (2003)
19. Thompson, K.: Agile journal productivity report. www.agilejournal.com (2011)
20. Kardile, V.V.: Understanding of software effort estimation at the early software development of the lifecycle—a literature view. Int. J. Eng. Res. Appl. **2**(1), 848–852 (2012)
21. Anselmo, D., Ledgard, H.: Measuring productivity in the software industry. Communications of the ACM, vol. 46, (Nov 2003)
22. Scacchi, W.: Understanding software Productivity (1994). Google search last access 12/12/2012

A Framework for Data Clustering of Large Datasets in a Distributed Environment

Ch. Swetha Swapna, V. Vijaya Kumar and J.V.R. Murthy

Abstract The chief motivation is to develop a framework for handling clustering of large datasets in a distributed manner. The proposal presented in this work addresses both numerical and categorical data with effective noisy information handling approach. Two basic models are developed known as primary and connected model to design the distributed approach. After forming clusters separately based on numerical and categorical features, an evolutionary approach is suggested to merge the clusters for optimization. A modification of multiple kernel-based FCM algorithm (MKFCM) Chen et al. (A multiple kernel fuzzy c-means algorithm for image segmentation 41:1263–1274, 2011) is used to implement the proposal. A comprehensive view of the designed method and algorithm is presented in this paper. Comparison of the results on few sample datasets shows the effectiveness of the proposed approach over existing one.

Keywords Clustering · Categorical and numerical data · Large dataset

1 Introduction

Data comes from different sources and in every possible form. For example, from social media sites, banking data, images, videos, web logs. The data also can be of different types, formats, and may contain several attributes. The organizations and

Ch.S. Swapna (✉) · J.V.R. Murthy
Department of Computer Science and Engineering, JNTU, Kakinada,
Andhra Pradesh, India
e-mail: swetha_swapna@yahoo.com

J.V.R. Murthy
e-mail: mjonnalagedda@gmail.com

V.V. Kumar
Department of Computer Science and Engineering, Anurag,
Group of Institutions, JNTU, Hyderabad, India
e-mail: vijayvakula@yahoo.com

enterprises generate data in volumes ranging from terabytes to petabytes stored as databases. Therefore, processing such large and complex data becomes difficult. The rate at which the data flows is also fast in certain applications where rapid processing is required.

Data mining is the extraction of potential and hidden information in the databases which can be used for promoting the company financially. Here various techniques are used like data summarization, eliminating unwanted data and noise, clustering and classification.

Clustering [1] is a process of grouping objects into clusters. The objects within the clusters share similar properties and are dissimilar to the objects of other clusters. Nowadays data comes in wide varieties and dealing with such miscellaneous data is certainly a hard task. There are numerous ways for dealing with numerical and categorical data. There are certain drawbacks which restricts the traditional clustering techniques from handling large datasets. They are sensitive to initialization problems, thereby giving suboptimal solutions which are apparent for mixed type of data. The algorithm can be run multiple times with different prototype initializations to overcome the above problem. But keeping in mind the constraints such as amount of computation and the time taken have led to the study of other approaches. Evolutionary algorithms are one such promising alternatives which guarantee to give global optima.

It is difficult to acquire both high accuracy and efficiency in a clustering algorithm of large data. The two targets always conflict with each other. The power of a single computer is not enough to wield the massive datasets. Parallel and distributed clustering is the key technique which is highly scalable and less expensive to do clustering in a distributed environment. A new framework is suggested to address the mixed data issues of large datasets while clustering. A provision for eliminating the outlier (noisy) data is also incorporated to enhance the cluster accuracy. An already existing algorithm MKFCM [2] is implemented as a basis for clustering and comparisons are made. The results reveal the effectiveness of the proposed approach.

The rest of this paper is organized as follows. The second section includes the recap of recent works. The third section provides detailed description about the present work and methodologies. The results are analyzed in the fourth section, and we conclude with the fifth section.

2 Review of Related Works

Literature presents several techniques for distributed clustering. Here, we review some of the techniques presented for literature. Inderjit and Modha [3] presented a parallel implementation of the k-means clustering algorithm based on the message passing model. The inherent data parallelism in the k-means algorithm is being exploited. As the number of data points increase, they analytically showed that the scale up and speed up of their algorithm approaches the optimal. Here, both the

computation and the memory used on distributed computers were parallelized by them. The analysis of test results on various datasets showed that the presented algorithm can effectively handle large dataset problems.

Jin et al. [4] have designed a method called Fast and Exact K-means Clustering (FEKM). Only one or a small number of passes on the entire dataset was required by the designed method and provably produced the same cluster centers as reported by the original k-means algorithm. Here, the cluster centers were altered by taking one or more passes over the entire datasets, before this the designed algorithm created initial cluster centers by sampling. Also, here they described and evaluated a distributed version of FEKM which was called as DFEKM. It was best for analyzing data that was distributed across loosely coupled machines. The DFEKM provided better result. If there was a significant load imbalance the designed method outperformed parallel k-means.

Ji and Ling [5] have designed a distributed clustering method to find global clustering patterns based on ensemble learning. The distributed data sources were examined and exploited by the showcased method. Distributed clustering has two stages, clustering being done at local sites foremostly and then in global site. The server site receives the local clustering results which form an ensemble that is merged with schemes of ensemble learning for yielding global clustering results. These outcomes were mathematically converted to be a combinatorial optimization problem. For the model a novel distributed clustering algorithm called DK-means was implemented. The experimental results achieved by DK-means showed similar results to K-means which clusters centralized dataset at a time.

In order to remove the single computing resource constraint Beaumont et al. [6] tried executing the task on a set of resources. Their aim was to design a distributed method for a large set of resources which enables to build clusters. They explained about a generic 2-phases method which was based on resource augmentation and whose approximation ratio was $1/3$.

3 Proposed Approach for Distributed Clustering

The recent advancement in digital world creates very large data to do their relevant process in various domains. Due to this uncontrollable growth of data, clustering played major role to partition data into small sets to do relevant processes within the small sets. But, again, the additional challenge of cluster identification problem is how to treat the large data since most of the algorithms are suitable only for small data. The usual way of handling large data in clustering is to solve clustering problem with parallel algorithm. The important assumption here is that the parallel algorithm can do better in terms of time consumption but the effectiveness should be also satisfactory. It means that the recent cluster methods should be applicable to do with large databases and performance should increase linearly with growth in data size.

Also, when developing a large data clustering, the additional challenges like, (i) different data format (numerical and categorical data), and (ii) noisy data or outlier data handling should be taken into account without much computational complexity. A few of the existing clustering algorithms of large data either can handle both data types but are not efficient when clustering large datasets or can handle large datasets efficiently but are limited to numeric attributes. These issues formulate a problem of developing an ideal parallel algorithm that is suitable to do large data clustering for both types of data (numerical and categorical) even if it contains noisy information. The objective of the work is to formulate a distributed data clustering that is suitable to two different types of data while handling noisy information effectively.

The aim is to develop a distributed clustering algorithm by altering the multiple kernel-based FCM algorithm [2]. MKFCM is an effective algorithm developed for clustering. This algorithm does not address the mixed data and noisy information effectively. A modification of KFCM (MFKCM) is employed in a distributed way by considering the issues discussed above. After the clusters are formed an evolutionary algorithm known as genetic algorithm [7] is applied to merge the clusters optimally.

Initially, the data is clustered with MFKCM algorithm to handle the numerical data. Later on the categorical data is dealt using the data-based measure. The clustered data is passed to the genetic algorithm for cluster optimization. Then a distance-based measure is applied to calculate the outliers in the clusters and then removed. The detailed explanation of the approach is furnished below.

3.1 Distributed System

A distributed system with n number of nodes is considered which is provided with large database. The aim is to cluster the large database effectively as the data contains both categorical data and numerical data. In a distributed clustering model the clusters are classified as primary model and connected model. The primary model represents clusters that are selected as major clusters. Connected model are the clusters that will be joined to the primary model.

Consider a dataset with m number of data and n number of attributes. The data is a multidimensional data represented as $m \times n$ in a distributed system, represent the data in nodes. The total $m \times n$ data is distributed over the nodes present.

$$D = \{(m_1, m_2, \dots, m_i), (n_1, n_2, \dots, n_j)\}. \quad (1)$$

Here, D represents the dataset, which consists of m data and n attributes. There are a total of “ i ” number of data m and “ j ” number of attributes n . Assume that we are dealing with a distributed system with number of nodes, $N = 3$. So the input data

has to be split over to the nodes. The data is partitioned into 30:30:40 proportions. Accordingly, the data contained in each node be,

$$\begin{aligned}
 N1 &= \{(p_1, p_2, \dots, p_i), (n_1, n_2, \dots, n_j)\}; p \times n. \\
 N2 &= \{(q_1, q_2, \dots, q_i), (n_1, n_2, \dots, n_j)\}; q \times n. \\
 N3 &= \{(r_1, r_2, \dots, r_i), (n_1, n_2, \dots, n_j)\}; r \times n.
 \end{aligned}
 \tag{2}$$

Let us discuss the distributed system in detail through scenario. Consider an input data of ten thousand data that is provided with four attributes. This input data D has to be supplied to the three defined nodes. The data is selected and given to each node as per the defined proportions. Figure 1 represents the distributed clustering model.

3.1.1 Clustering Initial Nodes

First the data is distributed among the nodes and then cluster process is subjected. One of the major issues considered by the approach is to handle categorical data and numerical data together. The categorical data clustering is processed in association with the numerical data clustering. We start with numerical data initially.

Fig. 1 Distributed clustering architecture

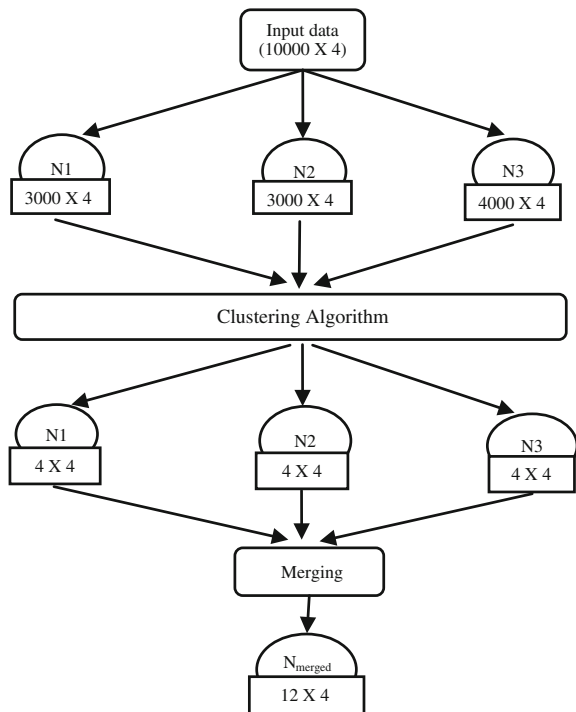
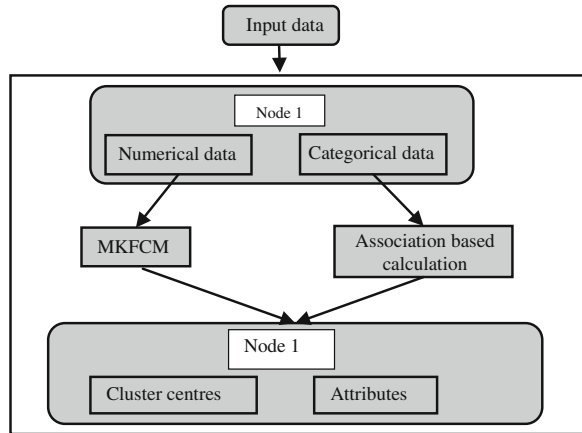


Fig. 2 Initial block of the large data clustering



The numerical data columns are chosen and clustered using the multiple kernel fuzzy c-means clustering (MKFCM). Then columns with categorical data are selected. The categorical data are processed in such a way that its major contribution in the total data is identified and that data is supplied to a particular cluster which has most association to the categorical data.

Figure 2 describes the initial phases where the clustering of numerical data and categorical data at a single node are dealt with. As represented in the figure, the numerical data is selected and clustered with the help of MKFCM algorithm. Then the categorical data is selected and their association to the clusters of numerical data is assessed and then grouped to a cluster.

3.1.2 Numerical Data Clustering

The initial process in handling the mixed data regarding the large dataset is numerical data clustering. The data distributed over the nodes are selected for MKFCM clustering, which is the recent method for efficient data clustering. The main problem regarding the MKFCM is, it can only handle numerical data. The K in the algorithm represents the kernels we used for clustering. Only two kernels are used for processing the clustering. Thus, 2-KFCM algorithm is employed here for the numerical data clustering. The data distributed over the nodes are selected and given as input to 2-KFCM algorithm. The MKFCM is basically an advanced version of the FCM algorithm. The data in the nodes are given as input with number of clusters and the kernel function.

3.1.3 Handling Categorical Data

Numerical data clustering is comparably easy with regard to categorical data. The categorical data is bypassed through a common to best cluster approach. The set of categorical data columns are picked from the data input and process is separated. Consider we are having the data as,

$$d = l \times n. \tag{3}$$

Here, “ l ” represents the numerical data and categorical data. The categorical data may be with different values. So we select frequency of each data in the column and find its relevance to the particular cluster. Consider we have clusters listed as,

$$Cl = [cl_1, cl_2, cl_3]; \text{ suppose the categorical data is given as, } C = [m, m, l, m, l, l]. \tag{4}$$

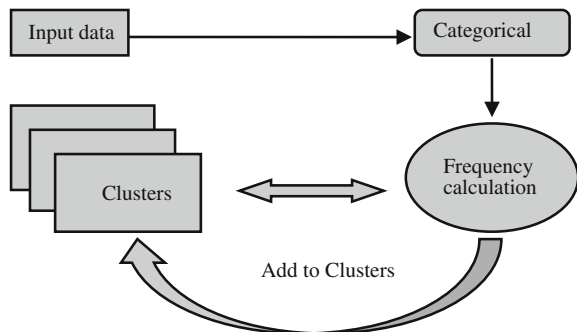
Suppose the frequency regarding “ m ” with respect to cluster cl_1 is 2 and that with “ f ” is 0. Thus according to the proposed approach’s implication, the data “ m ” will be added to the cluster cl_1 . In the next case, let the frequency of “ f ” is higher than “ m ” compared to cluster cl_2 . Thus “ f ” will be added to the cluster cl_2 . This procedure is followed until all the categorical data has been processed.

Figure 3 represents how the categorical data is treated. The processed data are then passed to the next level of processing. The next phase is detecting the outliers and noisy data from the clusters.

3.2 Removal of Outliers

The outliers are regarded as the most unwanted data belonging to the cluster. As far as a large dataset is considered there will be lot of unwanted data. So in order to deal with the problem we propose cluster centroid-based method. Initially after the numerical data and categorical data clustering, centroids of the cluster are

Fig. 3 Categorical data handling



calculated. The cluster centroids are selected and stored in a set. Each cluster and its centroids are defined as follows:

$$\begin{aligned} \text{Clusters} &= [cl_1, cl_2, \dots, cl_n]. \\ \text{Cluster centroids} &= [cc_1, cc_2, \dots, cc_n]. \end{aligned} \quad (5)$$

Once the centroids of each cluster are calculated, a distance-based calculation is subjected for calculating outliers. The distance is calculated between the cluster centroids and points in the cluster. Basic Euclidean distance is used for distance calculation and a threshold is set for finding the outliers.

The below steps represent the outlier detection model. Here, cc_i is the cluster centroid and cp_i is the cluster point. A threshold is set based on the average of maximum distance possessed by boundary points in the cluster. The outliers are points which possess distance more than the average. The detected outliers are then removed from the cluster and the clusters are subjected to optimization.

$$\text{distance} = d(cc_i - cp_i). \quad (6)$$

$$\begin{aligned} &\text{If}(\text{distance} > \text{threshold}) \\ &cp_i = \text{outlier} \\ &\text{else} \\ &cp_i = cp_i + 1. \end{aligned} \quad (7)$$

3.3 Merging the Clusters

The aim of this section is to get the efficient data from all the clusters from the other nodes. A genetic algorithm-based clustering is initiated for obtaining the merging node. Initially, clusters are selected after the outlier detection. The initial population to the genetic algorithm is derived from those clusters.

$$\text{pop} = [(cc_p \times n), (cc_q \times n), (cc_r \times n)]. \quad (8)$$

Here, pop represents the population to the genetic clustering. The elements to the population are selected from the centroid of the clusters in each node. 3 nodes are considered for processing the data, and each node has some defined number of clusters in it. The value cc_p represents the total number of cluster centers in the node 1, cc_q represents the cluster centers regarding node 2 and similarly cc_r belongs to node 3. Similar to the genetic algorithm, initially the fitness of the population is calculated using the defined fitness function.

$$fitness = \frac{1}{\sum_{i=1}^k \sum_{ch \in Nd} d(ch_i, cc_i)}. \quad (9)$$

Here the “ Nd ” represents the number of elements in the node and the ch represents the chromosomes or the data points in the cluster. The value cc represents the cluster centroid. The chromosomes of the population is derived separately as,

$$\begin{aligned} ch_1 &= [cc_p \times n]; cc_p \rightarrow cluster\ centers; n \rightarrow attributes. \\ ch_2 &= [cc_q \times n]; cc_q \rightarrow cluster\ centers; n \rightarrow attributes. \\ ch_3 &= [cc_r \times n]; cc_r \rightarrow cluster\ centers; n \rightarrow attributes. \end{aligned} \quad (10)$$

Consider the example block diagram plotted in Sect. 1. If the clusters of node 1 are selected as initial population, we can represent the chromosome as,

$$ch_1 = [4 \times 4]; cluster\ centers = 4; attributes = 4. \quad (11)$$

Here, we consider 4 clusters for node 1, thus there will be 4 cluster centers and the data totally contains 4 attributes. According to the above plotted fitness calculation, the fitness of the population is calculated with respect to each cluster. Then each cluster is selected according to their fitness value, which is minimum. Then the crossover process is applied to each chromosome for improving their performance. The crossover point is fixed and crossover rate is also fixed for each chromosome. The new population obtained after the crossover process is then subjected to the mutation process which will introduce a random data point to the cluster. Thus after the initial iteration the new population is developed. The newly generated population is again processed with the fitness function furnished above.

$$\begin{aligned} &If (new_fitness < old_fitness) \\ &\quad new_ch_i \rightarrow pop \\ &else \\ &\quad pop = old_ch_i . \end{aligned} \quad (12)$$

The procedure is recurred until the looping of the genetic algorithm is terminated. The resultant clusters are selected as the final clusters for the large dataset. The significance of the proposed approach is evaluated based on observational analysis. Considering the example plotted in Sect. 1, we subject the demonstration of the merging cluster. The merging cluster process considers input from the clusters.

Figure 4 depicts the processing of merging nodes with respect to the genetic clustering. In the figure, the 3 chromosomes are supplied to the genetic clustering algorithm. The chromosomes are initially subjected to the fitness calculation. Then after a series of genetic algorithm procedures, we finally group the data in the

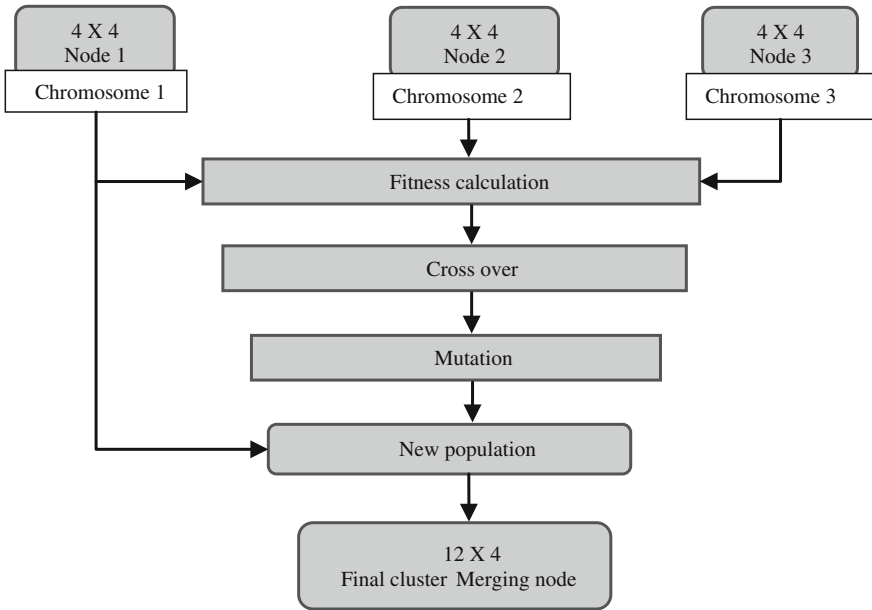


Fig. 4 Genetic clustering: merging node creation

resultant cluster. The resultant cluster is stored in a separate node, which is known as the merging node.

Algorithm: genetic clustering

Input: cluster centroids and attributes

Output: cluster centroid and attribute

Select the data from each node

Assign each “ i x j” as chromosomes

Select all chromosomes

$$ch_1 = [cc_p \times n];$$

$$ch_2 = [cc_q \times n];$$

$$ch_3 = [cc_r \times n];$$

Calculate fitness of each chromosome

$$fitness = \frac{1}{\sum_{i=1}^k \sum_{ch \in Nd} d(ch_i, cc_i)}$$

Apply crossover over the chromosomes

Apply mutation on the new off-springs

Select the new offspring and add it to merging node

End.

4 Experimental Results and Discussion

The experimental evaluation is conducted in order to evaluate the proposed approach. The work is developed in java programming language with JDK 1.7.0. The system that is used for the development is based on Intel Core i5 processor supported by 3 GB RAM and 500 GB hard disk space. The proposed system defined the problem in the distributed computing system. So the dataset relevant to the system has to be selected. In reference to the distributed computing system, Flag dataset [8] has been picked from the UCI data repository.

Flag dataset [8]: The dataset contains details of various nations and their flags. In the dataset, the contents are separated with commas and are a multivariate dataset. The dataset possess 30 attributes among which 10 attributes are numeric values and others are either Boolean or nominal valued. The dataset contains both numerical and categorical data. There are total of 194 instances.

Adult dataset [9]: The adult dataset deals with the prediction of whether an adult makes 50 K or more per year. The dataset is extracted by Barry Becker from the 1994 census database. A set of reasonably clean records are selected by providing various conditions. The dataset is a multivariate dataset with 48,842 instances. The dataset contain both categorical data and numerical data. The dataset possess a number of 14 attributes connecting social data.

The proposed distributed clustering algorithm uses mainly two evaluation matrices, the accuracy and time. The accuracy of the methods deals with the accuracy with which it clusters particular data into different clusters. The clustering accuracy depends upon input parameters and number of clusters. The cluster centroids of the produced clusters are calculated initially and are evaluated with respect to the class labels in the data. The accuracy is calculated based on the number of data relevant to a class to the total data present. Each evaluation is measured on the basis of a single cluster and the resultant values are summed together to get the overall accuracy.

$$Accuracy = \sum_{c=1}^C \frac{N_c^R}{N^T}. \quad (13)$$

Here, C represents the total number of clusters. The value N^R represents number of data relevant to a class and N^T represents the total number of data given as input. The next evaluation metric we use is the time for execution. The time for execution is evaluated by beginning and end of the program.

4.1 Performance Evaluation Based on Flag Dataset

The flag dataset is initially selected for the analysis. The analysis conducted for varying parameters: number of nodes in the distributed system, number of clusters

in the system and population size of the genetic algorithm. In the coming section, the accuracy varies based on the number of nodes subjected.

Figure 5 describes the accuracy of the proposed approach by varying the number of nodes of distributed system. The nodes are varied from 2 to 8 with interval 2. The analysis is conducted by setting the parameters, population size, and number of clusters to constants. The number of clusters is set as 8 and population size is set as 10. The analysis from the figure shows that the proposed approach has upper hand over the existing method.

In the case of existing method, as the number of nodes increases the variation in accuracy is not uniform. In the case of showcased work, the accuracy increases as the number of nodes increases. The maximum accuracy attained by the designed method is 72 %, which is very better figure as compared to the existing method, for which the maximum accuracy obtained is 40 %. All these experiments are conducted in the same environment.

Figure 6 depicts the accuracy of the method by varying the number of clusters intended by the system. The required number of clusters is varied from 4 to 16 with interval 4. The analysis is conducted by setting the parameters population size and number of nodes to constants. The number of nodes is set as 2 and population size is set as 10. The responses of the existing work are slightly different from that of the varying number of nodes. In this, the entire accuracy rate is constant when varying the number of clusters. On the other hand, the presented method showed similar

Fig. 5 Accuracy based on number of nodes

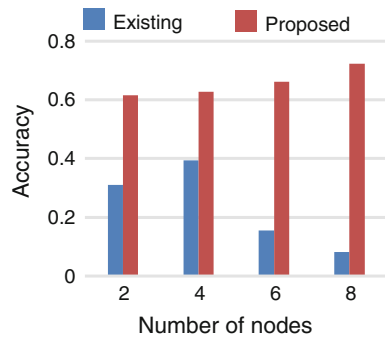


Fig. 6 Accuracy based on number of clusters

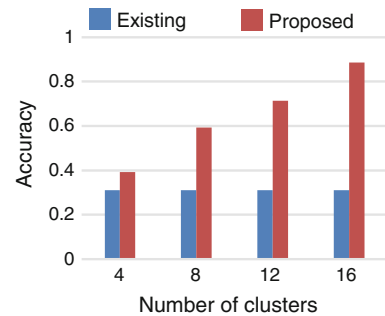
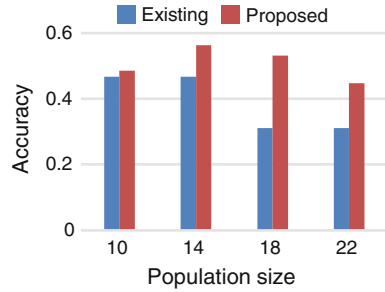


Fig. 7 Accuracy based on population size



responses as per the above analysis. Its accuracy is rising as the number of cluster increases. In this case the maximum accuracy obtained by the method is 88 %.

In Fig. 7, we have presented the accuracy of the proposed approach by varying the population size of the genetic algorithm. The population size of the genetic algorithm is varied from 10 to 22 with interval 4. The analysis is conducted by setting the parameters: number of clusters and number of nodes to constants. The responses we obtained in this regard are very different from the above two processes.

As the number of population size of the genetic algorithm increases, nonuniform responses from both the methods are obtained. Though both the algorithms provided irregular response, the proposed approach has the upper hand in the analysis also. The highest accuracy obtained for the current approach is 58 % and that for the existing approach is 44 %.

Figure 8 represents the time analysis of the intended work. The part “a” represents the time analysis based on varying number of nodes. The part “b” represents the time analysis based on the number of clusters and finally the part “c” represents

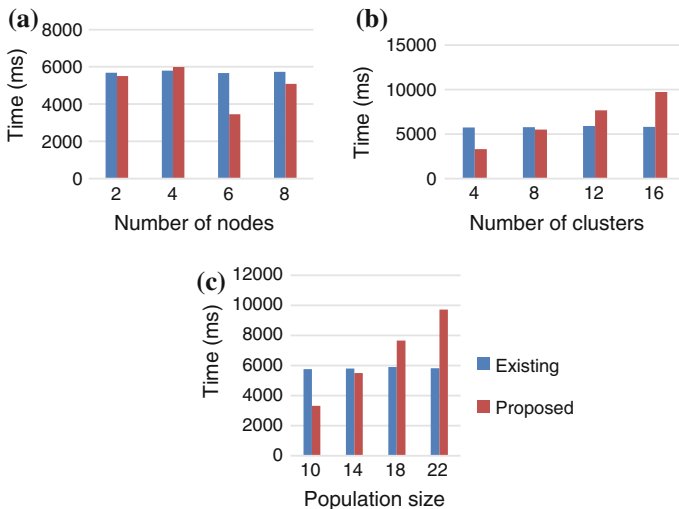


Fig. 8 Time analysis based on flag dataset

the time response based on varying the population size. The response regarding time for execution is different from that of accuracy analysis. The response is clearly irregular in manner for all the cases. Both the suggested objective and existing approach are clearly different from each other. In all the above three cases for some events the proposal shows better time efficiency and for other methods existing methods show better time efficiency.

4.2 Performance Evaluation Based on Adult Dataset

The initial analysis is based on the flag dataset with respect to the proposed approach. The flag dataset is initially selected for the analysis. The analysis was conducted for varying parameters; number of nodes in the distributed system, number of clusters in the system, and population size of the genetic algorithm. In the following section, the accuracy based on varying in number of nodes is subjected. In Fig. 9, we have presented the accuracy of the proposed approach by varying the number of nodes of distributed system. The nodes are varied from 2 to 8 with interval 2. The analysis is conducted by setting the parameters population size and number of clusters to constants. The number of clusters is set as 8 and population size is set as 10.

The analysis from the figure shows that the proposed approach has upper hand over the existing method. In the case of existing method, as the number of nodes increases the variation in accuracy is not uniform. In the case of proposed approach, the accuracy increases as the number of nodes increases. The maximum accuracy attained by the proposed approach is 72 %, which is a very better figure as compared to the existing method, for which the maximum accuracy obtained is 40 %. All these experiments are conducted in the same environment.

In Fig. 10, we have presented the accuracy of the proposed approach by varying the number of clusters intended by the system. The required number of clusters is varied from 4 to 16 with interval 4. The analysis is conducted by setting the parameters population size and number of nodes to constants. The number of nodes is set as 2 and population size is set as 10. The responses of the existing method are slightly different from that of the varying number of nodes.

Fig. 9 Accuracy based on number of nodes

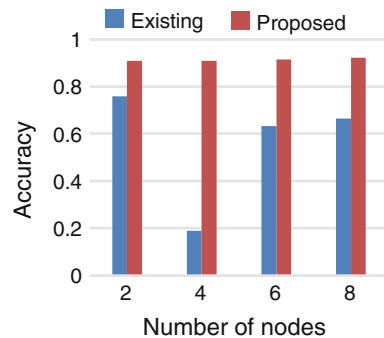
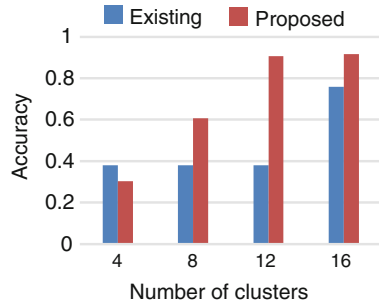


Fig. 10 Accuracy based on number of clusters



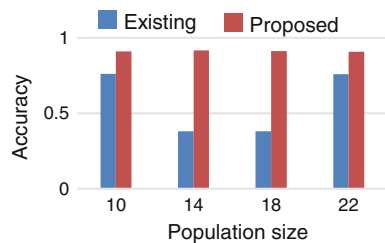
In this, all the accuracy rate is constant when varying the number of clusters. On the other hand, the proposed approach showed similar responses as per the above analysis. The accuracy of the proposed approach increases as the number of cluster increases. In this case the maximum accuracy obtained for the proposed approach is 88 %.

In Fig. 11, we have presented the accuracy of the proposed approach by varying the population size of the genetic algorithm. The population size is varied from 10 to 22 with interval 4. The analysis is conducted by setting the parameters number of clusters and number of nodes to constants. The responses we obtained in this regard is very different from the above two processes. As the number of population size of the genetic algorithm increases, we obtained nonuniform responses from both the methods. Though both the algorithms provided irregular response, the proposed approach has the upper hand in the analysis also. The highest accuracy obtained for the proposed approach is 58 % and that for the existing approach is 44 %.

Figure 12 represents the time analysis of the proposed approach. The part “a” represents the time analysis based on varying number of nodes. The part “b” represents the time analysis-based number of clusters and finally the part “c” represents the time response based on varying the population size.

The response regarding time for execution is different from that of accuracy analysis. The response is clearly irregular in manner for all the cases. Both the proposed approach and existing approach are clearly different from each other. In all the above three cases for some events the proposed approach shows better time efficiency and for other methods existing methods show better time efficiency. As considering all scenarios, we assess that the proposed approach is efficient in the time analysis.

Fig. 11 Accuracy based on population size



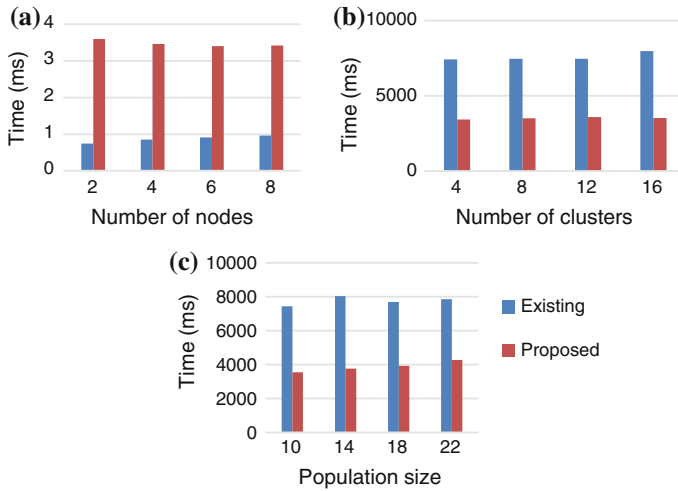


Fig. 12 Time analysis based on adult dataset

5 Conclusion

The distributed clustering acquires a special attention in the field knowledge sharing. In this paper, a method for distributed clustering has been delivered. Effective approaches to handle numerical, categorical features, and removal of outliers are taken care while clustering. Genetic algorithm is utilized to merge the clusters formed from the different distributed nodes so that finally, optimal number of clusters are obtained. The suggested method uses a modified MKFCM algorithm to handle numerical data and an association-based method to handle categorical data. The experimental analysis on a sample dataset has shown the effectiveness of the aimed objective. As further research, it remains to be seen how well it fits with wide variety of datasets and also other evolutionary approaches can be investigated to further optimize the cluster numbers while merging.

References

1. Ji, J., Pang, W., Zhou, C., Han, X., Wang, Z.: A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. *J. Knowl. Based Syst.* **30**, 129–135 (2012)
2. Chen, L., Chen, C.L., Lu, M.: A multiple-kernel fuzzy C-means algorithm for image segmentation. *IEEE Trans. Syst. Man Cybern. Part B* **41**(5), 1263–1274 (2011)
3. Inderjit, S.D., Modha, D.S.: A data-clustering algorithm on distributed memory multiprocessors. In: *Proceedings of KDD Workshop High Performance Knowledge Discovery*, pp. 245–260 (1999)
4. Jin, R., Goswami, A., Agrawal, G.: Fast and exact out-of-core and distributed K-Means clustering. *J. Knowl. Inf. Syst.* **10**(1), 17–40 (2006)

5. Ji, G., Ling, X.: Ensemble learning based distributed clustering. *Emerg. Technol. Knowl. Discov. Data Min.* **4819**, 312–321 (2007)
6. Beaumont, O., Bonichon, N., Duchon, P., Eyraud-Dubois, L., Larcheveque, H.: A distributed algorithm for resource clustering in large scale platforms. *Principles Distrib. Syst.* **5401**, 564–567 (2008)
7. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, 1st edition (1989)
8. Flag dataset: <http://archive.ics.uci.edu/ml/datasets/Fl>
9. Adult dataset: <http://archive.ics.uci.edu/ml/datasets/Adult>

Segmentation of Ancient and Historical Gilgit Manuscripts

Pinjari Hameed, Rosemary Koikara and Chethan Sharma

Abstract The Gilgit manuscripts belong to fifth century A.D. and are oeuvre of texts which deal with Buddhist work. It is one of the oldest manuscripts in the world and is considered to be a milestone in the history of Buddhist works in India. It is a collection of both official and unofficial Buddhist works which are believed to have helped in the evolution of many literatures including Chinese, Japanese, and Sanskrit. Since this manuscript is almost seventeen centuries old it has not been able to fully decipher the text yet. It has been laminated by the National Archives of India which proves it is one of the most important literatures concerning India. In this paper, we perform character-based image segmentation on Gilgit manuscript in order to simplify and to better identify character in the image of the manuscript. The employed method gives an accuracy of nearly 87 %.

Keywords Gilgit manuscripts · Segmentation · Character-based · Buddhist works

1 Introduction

Image segmentation is the process of segmenting or dividing an image into multiple parts/segments. Segmentation is done to simplify representation and to better analyze the image for future post-processing steps and make it more meaningful. Segmentation is generally the first step in the process of analyzing any image [1]. It bridges the gap between high-level and low-level image processing. The final result of any image processing task is directly related to how efficient the segmentation was done, but finding an efficient and reliable segmentation process is extremely difficult. The main use of segmentation is to identify region of interest (ROI).

P. Hameed (✉) · R. Koikara
Department of CSE, Christ University, Bangalore, India
e-mail: hameed135781@gmail.com

C. Sharma
IEEE Member, Bangalore, India



Fig. 1 Gilgit manuscript

Segmentation could also be used for image database lookup, editing images, compression, or boundary estimation. The purpose of this work is to segment Gilgit manuscripts using character-based segmentation so that they can be better analyzed to further understand historic Buddhist works.

The Gilgit manuscripts belong to fifth century A.D. and are oeuvre of texts which deal with Buddhist work. It is one of the oldest manuscripts in the world based on Buddhist work. It is a collection of both official and unofficial Buddhist works which are believed to have helped in the evolution of many literatures including Chinese, Japanese, and Sanskrit. Since this manuscript is almost seventeen centuries old it has not been able to fully decipher the text yet. It has been laminated by the National Archives of India which proves it is one of the most important literatures concerning India. These manuscripts are given the name as “Gilgit” because they were discovered in a place called Gilgit (which is now in Pakistan-occupied Kashmir) in 1931 by a group of cattle grazers.

Gilgit manuscript, as shown in Fig. 1, was originally written in Pali text and contains four sutras, one of which is the Lotus Sutra which even today is a very important scripture in Japan that influences political and cultural life of that country [2].

2 Literature Survey

The research on image segmentation is being going on since a very long time. There are many segmentation techniques which are broadly classified as Implicit, Explicit, and Holistic Segmentation [3]. Implicit segmentation is a process of searching the

images for the part of matching alphabets. It is also known as recognition-based segmentation [4–6]. Explicit segmentation is also known as pure or external segmentation in which input image text is divided into separate characters. This process is also called as dissection since a long image text is split into individual characters. Saba et al., in [7] use explicit segmentation to find consanguinity between characters in an image and the image is cut through these connections to separate out individual characters. Amer Dawoud in [8] proposed algorithm character segmentation for handwritten numerical called CSSG which is applied iteratively at equally spaced thresholds which gives an accuracy for 76.9 %. In [9] Roy et al. propose an angel-based dynamic programming method on English text which gave an accuracy of 91.36 %. In [10] David proposes a novel method to detect intersection regions and to extract character objects using circle scanning method. This paper addresses the problem of extracting character objects and deals with form, drawing, and map recognition. Roberto and Thomé [11] present a study on first sight decision tree algorithm for recognizing cursive script based on the use of histogram.

The Gilgit manuscript is basically a collection of Buddhist works and contains one of the most important texts of Maha-yana called the Samadhirajasutra. Gilgit manuscript is one of the oldest manuscripts in the world based on Buddhist work in India. Samadhirajasutra which is a popular text among Buddhist population is also called as Arya Chandrapradipasutra [2]. It has been found that this Samadhirajasutra text was written during 75–100 AD since it contains three ganas which were last used during Emperor Kanishka who lived during 75–100 AD as per Buddhist traditions [2]. Nalinaksha Dutt is the person who edited the Gilgit manuscript particularly Vol II points i–iii during 1941 and 1954. Sailendrabodhi and Dharmatasila translated these Gilgit manuscripts to Tibetan during ninth century as we note that most of the Tibetans follow Buddhism and Gilgit is a Buddhist manuscript [2]. In this paper, we perform character-based image segmentation on Gilgit manuscript in order to simplify and to better identify character in the image of the manuscript. The employed method gives an accuracy of nearly 87 %.

3 Methodology and Results

In this section, we discuss the method used to segment the Gilgit manuscript which is one of the oldest manuscripts in the world and is considered to be a milestone in the history of Buddhist works in India. The algorithm used is iterative in nature. The algorithm converts the input image to black and white image with white characters and black background. It scans the image from top left-corner of the image. When first black pixel is encountered, the complete character is found and is extracted as output image. This process is repeated until null (last character) is encountered. Flowchart of the algorithm is as in Fig. 2.

The above method is implemented using Matlab 2012a. We have tested the method on few sample images taken from the Gilgit manuscripts. Figures 3, 4, 5

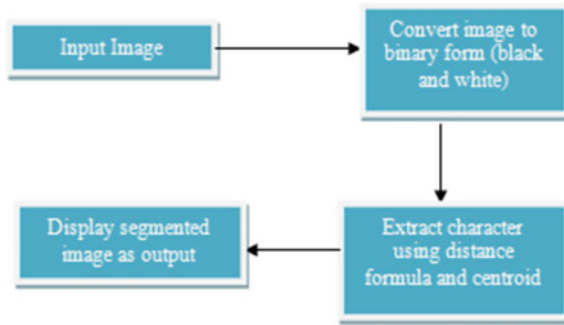


Fig. 2 Flowchart of the algorithm



Fig. 3 Sample input



Fig. 4 Binary image

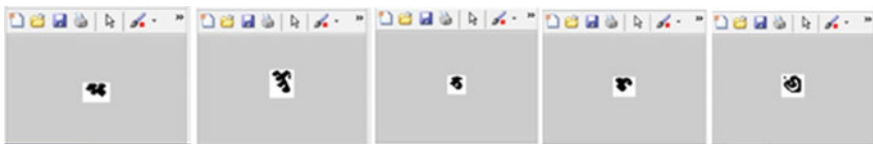


Fig. 5 Segmented characters

depict the result of the segmentation on sample of Gilgit manuscript in order to simplify and to better identify character in the image of the manuscript. The employed method gives an accuracy of nearly 87 %.

4 Conclusion

Image segmentation is the process of segmenting or dividing an image into multiple parts/segments. Segmentation is done to simplify representation and to better analyze the image for future post-processing steps and make it more meaningful. Segmentation is generally the first step in the process of analyzing any image. The Gilgit manuscripts belong to fifth century A.D. and are oeuvre of texts which deal with Buddhist work. It is one of the oldest manuscripts in the world based on Buddhist work. In this paper, we perform character-based image segmentation on Gilgit manuscript in order to simplify and to better identify character in the image of the manuscript. The employed method gives an accuracy of nearly 87 %.

References

1. http://en.wikipedia.org/wiki/Image_segmentation
2. Gilgit manuscript—piecing together fragments of history: Copyright © 2000–2004 (the-south-asian.com) Aug 2004
3. Choudhary, A.: A review of various character segmentation techniques for cursive handwritten words recognition. *Int. J. Inf. Comput. Technol.* **4**(6), 559–564 (2014)
4. Cavalin, P.R., Britto, A.S., Bortolozzi, F., Sabourin, R., Oliveira, L.: An implicit segmentation based method for recognition of handwritten strings of characters. In: *Proceedings of the ACM Symposium on Applied Computing*, 836–840 (2006)
5. Gillies, M.: Cursive word recognition using hidden Markov models. In: *Proceedings of the Fifth U.S. Postal Service Advanced Technology Conference*, pp. 557–562 (1992)
6. Cho, W., Lee, S.W., Kim, J.H.: Modelling and recognition of cursive words with hidden Markov models. *Pattern Recognit.* **28**(12), 1941–1953 (1995)
7. Saba, T., Sulong, G., Rehman, A.: Document image analysis: issues, comparison of methods and remaining problems. *Artif. Intell. Rev.* **35**, 101–118 (2011)
8. Dawoud, A.: Iterative cross section sequence graph for handwritten character segmentation. *IEEE Trans. Image Process* **16**(8), 2150–2154 (2007)
9. Roy, P.P., Pal, U., Lladós, J., Delalandre, M.: Multioriented touching text character segmentation in graphical documents using dynamic programming. *Pattern Recognit.* **45**(5), 1972–1983 (2012)
10. David, X.Z.: Extraction of embedded and/or line touching character line objects. *JPRS* **35** (2002)
11. Roberto, R.J., Thomé, A.C.G.: Cursive character recognition—a character segmentation method using projection profile based technique (2002)

Semantic-Based Approach for Automatic Annotation and Classification of Medical Services in Healthcare Ecosystem

Vijayalakshmi Kakulapati, Rishi Sayal, Ravi Aavula
and Sunitha Devi Bigul

Abstract A vast amount of related healthcare information exists over the web without any explicit semantic association. Healthcare ecosystem makes use of medical services for the services entities of publishing and classification. However, before the emergence of healthcare ecosystems, where ecosystems are generally present in the environment, medical service and healthcare information are diverse. Therefore, the first medical service is a key issue to deal with information systems in the healthcare environment. In this paper, we propose health-related clinical data annotation, classification, and interpretation of medical data in relation to the level of classification based on the existence of the frame, and for improving the customer's request to present a semantic-based Web mining. In addition, we classify medical data in relation to the level of clustering based on the use of healthcare information. Information relevant to the development of semantic information extraction can be achieved using a better phrase. Highly relevant improved information requested can be retrieved by deployment of additional medical terms. Our experimental evaluation results and the feasibility of assessing the impact of the proposed mining method show improvisation.

Keywords Healthcare · Semantic-based mining · Classification · Medical service ontology

V. Kakulapati (✉) · R. Sayal · R. Aavula
Department of CSE, Guru Nanak Institutions Technical Campus,
Ibrahimpattam, Hyderabad, India
e-mail: vldms@yahoo.com

R. Aavula
e-mail: aaularavi@gmail.com

S.D. Bigul
Department of CSE, C.M.R Institute of Technology, Kondlakoya, Hyderabad, India

1 Introduction

Data mining is popular in most visible areas such as business and web search. Meaningful and hidden patterns in data sets in the medical domain have great potential for exploration. Studies have observed that people searching the web read about health information that reveals medical information. People need reliable information accessing in a manner that is compatible to a faster workflow [1, 2]. Medical care-related information such as articles, clinical trials, news, etc., is an advantage to healthcare providers and to individuals.

Healthcare Information and communications technology has evolved along with increased medical care, health, environment, recognizing the nature reserves [3, 4]. Medical services, such as the services of health conditions, from the perspective of drugs industry, health insurance, pharmacy, diagnosis, clinics, hospitals and specialized medical treatments, such as personal services, including diversity and geographical dispersal [1, 5, 6], with features that have health conditions in the environment are involved in services and provide services to the service provider to have service requester these two characters, you can play. Delivery service knowledge bases are stored in a service provider, which is a service company, and a healthcare environment by publishing system enters the environment. Service provider of medical services and health conditions of the service allow the company's environment, the functional elements of the group. When a service provider publishes a service organization such as the Resource Description Framework or the Web Ontology Language for the Semantic Web markup languages [7] oblivion change, and health care is provided in the context of domain-specific thinking can be categorized.

Due to the purpose of the question, denoting users and vaccines and health care information retrieval problem, which are not be filled to the question of health care information, the lack of support for the classification of semantic. Consequently, the lack of discovery and annotation of semantically entities who serve the health of the ecosystem as a whole in order to ensure efficient and effective work, there is an important issue for the health of the ecosystem. Moreover, the existing literature, especially the accumulation of these entities, which affect the performance of the ever-present medical care entities, classifying a method designed.

Therefore, automated and ontology-based services are being needed for classifying methods. Such a method has been shown to serve the health of the ecosystem environment entity collection, management, and retrieval of an improvement, resulting in health care entities and may be used to categorize as shown in Fig. 1.

2 Background Study

According to the existing literature, the emerging semantic-based data mining is categorized into two basic categories, metadata based and Ontology-based data mining [8, 9]. In the following section, we discuss the features of the two categories

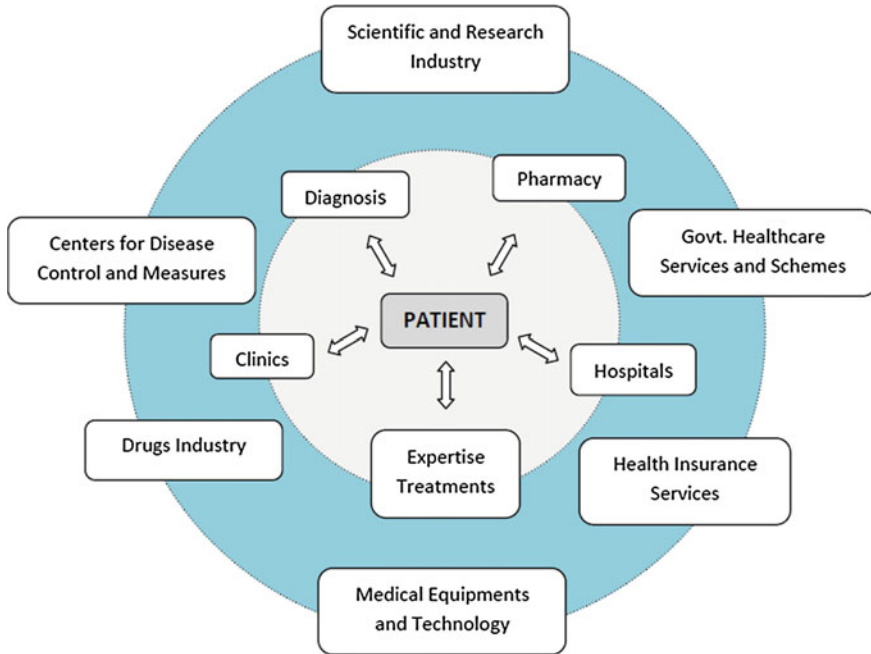


Fig. 1 Healthcare ecosystem

of semantic-based data mining and data mining systems for the healthcare environment and examine their potential problems [10–12].

2.1 *Ontology-Based Data Mining*

Ontology-based data mining [13, 14] is used for filtering and categorizing Web documents, related ontology concepts that refer to a group of mining methods, with links to Web documents [15]. Rosario and Hearst [16] use hidden Markov models and maximum entropy models to perform the tasks of entity recognition and relation discrimination for diseases and treatments. Their representation techniques are based on words in context, part of speech information, phrases, and medical lexical ontologies. Ray and Craven [17] worked on the relation extraction or relation identification of biomedical tasks, and Srinivasan and Rindflesch [5] worked on the diseases and drugs relation.

Yuvarani et al. [18] describe the existence of semantic similarities between the URL used to evaluate the contents of web pages and describe ontology-based data mining proposed to focus on crawler. Topics in the form of ontology Ontologies are stored inside the base. Query is sent to commercial search engines, and the related URL is retrieved from the search engine. Web pages are crawled, and then the URL

is extracted from web pages and their surrounding texts. The texts to determine the relevance of the query URL are matched with the concept of a consistent ontology [19].

Tane et al. [20], using Semantic Web technologies, proposed an ontology management system based on ontology crawler. It allows the user to assign a weight to some of the concepts in the ontology, which defines the traffic where you can specify individual preference. Other ontology concepts can be obtained by considering the weight of the interrelations between the ontology of concepts. After earning a web page crawler, its text and URL in order to compute the similarity between the concepts of ontology descriptions URL are matched to the weight of the ontology concepts. Comparison of weight, coupled with the concept of ontology calculates the weight of the web page. In terms of the aforementioned process, the Web pages are categorized and ranked into ontology concepts.

The crawler is near the lessons of the URL in the URL to get the descriptive texts and the texts and ontology concepts based on the URL to compute similarity between the observed. However, sometimes the best lessons are in or near enough for crawlers on the ontology-based similarity computing to limit the rate of increase in the wrong URL, and cannot be used to describe it.

2.2 Metadata-Based Data Mining

Metadata-based mining crawler abstracts meaningful information from the web pages of the mine and the existence of markup languages annotate information [8, 21].

Francesconi and Peruginelli [22] proposed crawler-based metadata management system portal for legal documents. Metadata generator provides meaningful information to change the metadata from web documents downloaded. Metadata format is in extensible markup language (XML) and two algorithms implement Naive Bayes and multiclass support vector machine (SVM) for metadata-based document classification.

Giles et al. [23] described a metadata-based e-business information search engine crawler. Lessons from the Web documents are analyzing and parsing texts crawler-based product CiteSeer metadata-running technique. Metadata-based document classification support vector machine for its (SVM) algorithm is employed.

Supervised classification approaches [19, 24] for classification of the document as the MSVM, SVM, the document metadata-computing distance-based web mining, which is described in scripture is plain enough without the support of semantic classification before use. But without the support more than half the performance will be reduced to plain classification document classification and classification of web documents based on the knowledge of the domain name cannot fulfill the needs of customers, leading to a limit of metadata-based mining, and may not match the actual domain knowledge of the structure of the search.

2.3 Semantic Mining Issues in Healthcare Ecosystems

Apart from the limitations of each category of semantic data mining approaches [25, 26], which is described as a healthcare service range of ecosystems, a common shortcoming is the following requirement of the mining share.

Healthcare ecosystems have found an urgent need for the service of the company and its divisions, including the categorization of what a particular service should be. However, the existing semantic-based crawler is crawling on the health service, medical service due to lack of domain knowledge. Therefore, the design work of the health service domain knowledge in healthcare ecosystems found crawling a challenge in the health services research.

In ontology-based and metadata-based data mining, we combine characteristics of both sides to overcome the limitations of data mining. Ontology-based classification and classification of the document for web-based abstraction of the characteristics of the effective metadata are explored for mining. Classification and web documents, and more importantly, matches in the domain knowledge of the structure of the classification error rate reduction as a result of matching the appropriate interpretation of the support [27–29], is characterized in the past. The description of the semantics of the following URL is to enrich and enhance the effect of the classification of the URL instead of just getting near the lessons of the descriptive metadata.

We present a novel semantic-based data mining methods for healthcare extraction, annotation, and classification in Healthy Ecosystems which combines the features of ontology-based and metadata-based data mining.

3 Semantic-Based Mining Approach for Healthcare Ecosystem

We have the functionality, architecture, components, machinery, data format, and the corresponding mathematical models of health from the point of view of a semantic-based data mining framework for data mining is proposed.

3.1 System Framework

The framework in Fig. 2 describes the configuration by the user brings the policy is based on the popular web service discovery functionality Healthcare ecosystems, which corresponds to the Web, extracts information related to the health service entities, which explains the name of the semantic-based data mining. The existence of this knowledge base to support the ability to annotate semantic information is based on the classification of the service. Policy Framework brings operational

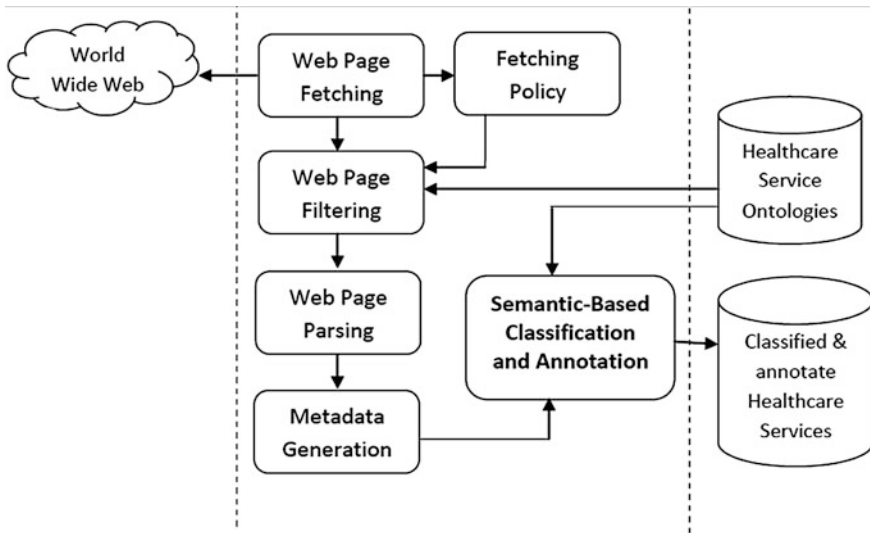


Fig. 2 Framework of semantic-based data mining for healthcare ecosystem

base, bringing in a webpage, webpage, parsing and semantic-based classification and annotation of metadata generation, has six functions.

Fetching Policy: Fetching policies can be configured by the user to control the behavior of the webpage brings. It is necessary to visit a web page, who defines the initial URL. It controls the maximum depth of a web page has been visited in order to avoid overloading. Again and again to avoid visiting the same URL, this is overseen by a visit to the URL.

Webpage Fetcher: The operation of webpage fetcher is to get the policies configured to download Web pages using the multithreading technology. In addition, it extracts the URL from the web pages and the web page URL filtering to bringing policy and sends Web pages.

Webpage Filtering: Filter defined in the policies of the webpage brings health, pathology and relevance of the web page is designed to filter through which the Web documents. It ontologies health checks for the Association of the web pages

Webpage Parser: Webpage parser is to parse snippets to extract meaningful information from the content of the webpage content is the application of web documents. Text processing is realized set of rules to deal with the diversity of information for the number of web pages, information, Web sites, Service, and in the format of a regular service metadata define the rules by referring to the original web page of the website layout is usually carried out in a consistent style.

Metadata Generation: Hypertext markup tags applied to the webpage for the metadata information Metadata Generator extracts information from the metadata. For semantic-based classification and annotation process sends the collected metadata.

Semantic-Based Classification and Annotation: The existence of a structured knowledge base of health-semantic annotation tools based classification and pre-existence of the health service by associating metadata with metadata to classify the product service. Based on the similarities between metadata and the terms of each of the community health service are termed.

The existence of the ecosystem service healthcare service organizations who share some common features of each word in terms of service, which is an abstraction, there is a hierarchy. Terms sense, an abstraction of its subconcepts in which the relationship, relate to it. All the metadata associated with a higher level concept has its subconcepts only the bottom level of service for the classification of metadata concepts, have the privilege of associating metadata service.

3.2 Healthcare Metadata Concept Similarity Computation for Classification and Annotation

In order to comment on the health service information service metadata generated metadata generator. Health metadata description and service provider of a service entity refers to the URL of the link. Description of a service, because there may be more than one entity in a real environment, this property can be extended in accordance with the number of snippets of information related to service descriptions. In addition, this property in service metadata, and can be implemented for calculating the similarity between the terms of service. According to the knowledge of the terms of service for a variety of common formats suitable for the service to which the service metadata and observed, and therefore it can be modified.

Is a service metadata, semantic-based classification metadata and positive health service entity for each service will compute the similarity between the words feeling. Here, we describe the concepts to achieve this goal, a run similar to the model.

Concept similarity considers the terms as a concept and finds its relevance to the phrase or metadata. It also related with other concepts in the domain it belongs. Let's consider a domain (D) which consists of set of terms,

$$D = k_1, k_2, \dots, k_n \tag{1}$$

where, D —is the domain and k_i —is terms that belongs to the domain.

To find the relation between the terms and metadata we find the relation between the terms and the metadata of the domain. We find the highest frequent term as $freq(x)$; and then dependency of this term to other terms to this using the following equation as,

$$freq(x) = \frac{X_n}{N_k}, X \in k. \tag{2}$$

where, X — x is the element and N_k —is the total number of elements.

After finding the most frequent terms, we have to find whether it belongs to a metadata in the metadata map. We find the inter relation between the term and other terms and calculates the dependency.

The dependency $dep(x : y)$ is used to extract the concept, it calculated using the following equation.,

$$dep(x : y) = \frac{P(x | y)}{P(y)}, x, y \in D \tag{3}$$

where,

$$P(x | y) = \frac{P(y \cap x)}{P(x)}$$

$P(.)$ —Probability of each term present in the domain.

For example, a term likes “*Malaria Antibiotics*”. We need to do the following for finding the concept similarities,

- Find the frequency of “*Malaria*”, $P(Malaria)$
- Find the frequency of “*Antibiotics*”, $P(Antibiotics)$
- Find the frequency of “*Antibiotics*” and “*Malaria*”, $P(Antibiotics | Malaria)$
- Find $\frac{P(Antibiotics \cap Malaria)}{P(Antibiotics)}$ i.e. $P(Antibiotics | Malaria)$
- Find $\frac{P(Antibiotics \cap Malaria)}{P(Malaria)}$ i.e. $dep(Antibiotics: Malaria)$

“*dep*” values are passed to $T \times M$ formation.

• **T × M matrix formation**

T is the number of terms and M is the number of Metadata.

	d_1	d_2	...	d_n	$\sum values$
c_1	$dep(c_1, d_1)$	$dep(c_1, d_2)$...	$dep(c_1, d_n)$	$\sum_{n=1}^N dep(c_1, d_n)$
...
c_n	$\sum_{n=1}^N dep(c_n, d_n)$

The $\sum Value$ are sorted in descending order and those results which are higher than the thresholds are selected and those information are annotate to the metadata.

4 System Evaluation

To evaluate the system, we take a look at our semantic-based data mining framework with three traditional information retrieval processes to perform calculations to run a series of experiments to measure, Precision, Recall, and Fallout Rate [30].

Precision: It is used to measure the preciseness of an information retrieval system. Precision for a term is the proportion of the relevant metadata associated by this term in all the metadata associated by this concept, which can be represented as:

$$\text{Precision} = \frac{\text{Number of Associated and Relevant Metadata}}{\text{Number of Associated Metadata}}$$

Recall: Refers to the measurement of the effectiveness of the recall of information retrieval system in question. Recall for a word in the collection of metadata related to the concept of metadata associated with this item, the percentage of the corresponding metadata, which can be represented as:

$$\text{Recall} = \frac{\text{Number of Associated and Relevant Metadata}}{\text{Number of Relevant Metadata}}$$

Fallout Rate: Fallout rate for a term production metadata nonrelevant for the concept of metadata associated with this item in the collection, the percentage of the concept of nonrelevant, which can be represented as

$$\text{Fallout Rate} = \frac{\text{Number of Associated and Nonrelevant Metadata}}{\text{Number of Nonrelevant Metadata}}$$

Other performance measures to compare the value of not less good performance in the Fallout Rate value.

4.1 Experiments

Experimental Environment Setup: It is, at least to start work on the health service domain ontology, semantic-based data mining is known as the Service Ontology need to be installed into the base. However, such an entity is available now. Therefore, we sought out a variety of health care services in a five-tier hierarchical structure of the ontology to build the healthcare service. Each item is defined by a certain number of explanations in terms of the concept, and concepts related to the concept and sub-concept relationship. The Healthcare Service Ontology can be seen in Fig. 3. We implement semantic-based data mining using JAVA technologies.

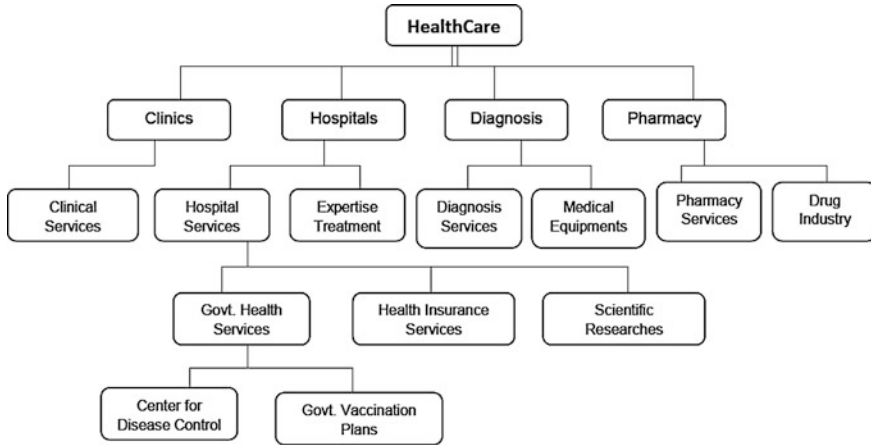


Fig. 3. Healthcare service ontology

Experimental Results: We are publishing the web site and the product is different from the web of the health service metadata, semantic-based data mining to download Web pages from the implementation of healthcare. In order to get the correct entry for data mining, we have a range of 50–100, based on the introduction of the three measures calculated at the time of the research, which vary for each increment. The test results are show below.

The Precision increases from 35 to 98 %, it clearly indicates that the effects of the variation of the initial value on Precision as shown in Fig. 4. Exceeds the threshold value of 80, however, remains relatively high precision of the above when the initial value is greater than 80. This is because some of the nonrelevant metadata filters. This result proves that the semantic-based mining can work precisely when its threshold value is varies.

Threshold values, as shown in Fig. 5 enhances the responses to recall that the observed decreases continuously. However, the diversity of the recall interval is relatively narrow. The initial value is not heavily affected by the recall; the

Fig. 4. Rate of precision ratio versus threshold

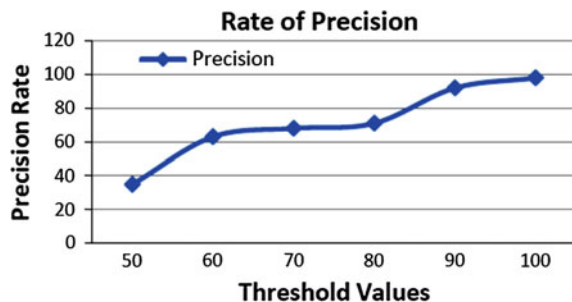


Fig. 5. Rate of recall versus threshold

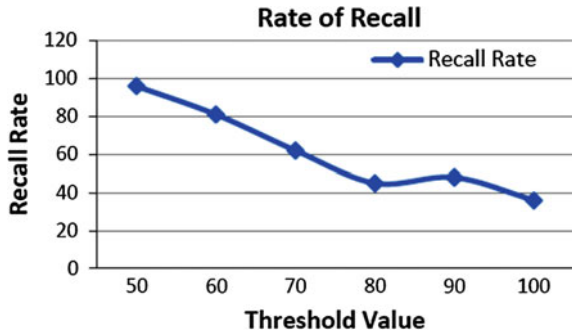
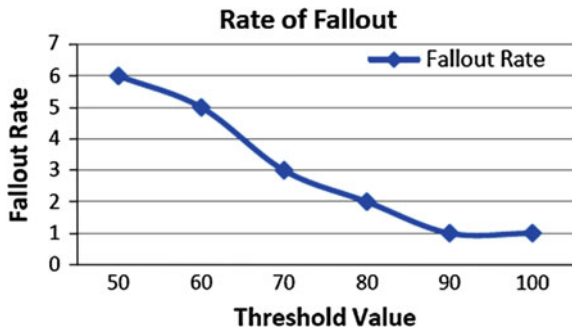


Fig. 6. Rate of fallout ratio versus threshold



relatively high threshold values do not affect the nonrelevant this is because with the exception of metadata. The average score of the crawler recall preliminarily proves the effectiveness of the proposal is more than 96 %.

As shown in Fig. 6 Fallout Rate of the rate of growth is positively affected by the initial value. Moreover, the rate of change of Fallout is from 6 to 0.09 %, which is obvious. Moreover, the initial value of 80 is reached, the higher rate threshold values fall in the semantic-based mining can maintain a low error rate is close to zero indicates.

5 Conclusion and Future Work

In this paper, we have designed a system of semantic-based data mining for health Ecosystem, which combines ontology-based metadata classification from data mining ontology-based and metadata-based data mining from metadata abstraction, in order to achieve the goal of discovery service, information and classification in health Ecosystems.

The main functions of the system, the Web, parsing, annotations, and have to find the information stored in the health service. We compute the similarity of the

concept of metadata and the terms of the relationship between function and provides a common format for service metadata and define the concept of service. From the experiments, we reduce the size of the initial value of the metadata associated with growth and nonrelevant, and a relatively high threshold values can benefit observed in the overall performance of semantic-based mining. Our research will concentrate more on specific ontology design for others in domains comprehensive health care to get a broader and more accurate search scope of semantic-based mining and strengthen the concept similarity algorithm to achieve better performance in classification and description metadata.

References

1. Winn, J., Bishop, C.: A Unified Modeling Approach to Data Intensive Healthcare. Health And Wellbeing, Microsoft Research (2007)
2. Li, J., Zhang, Z., Li, X., Chen, H.: Kernel-based learning for biomedical relation extraction. *J. Am. Soc. Info. Sci. Technol.* **59**(5), 756–769 (2008)
3. Yakushiji, A., Tateisi, Y., Miyao, Y., Tsujii, J.: Event extraction from biomedical papers using a full parser. *Proc. Pacific Symp. Biocomput.* **6**, 408–419 (2001)
4. Donaldson, I., et al.: PreBIND and Textomy: mining the biomedical literature for Protein-Protein interactions using a support vector machine. *BMC Bioinf.* **4** (2003)
5. Srinivasan, P., Rindfleisch, T.: Exploring text mining from medline. In: Proceedings of the American Medical Informatics Association (AMIA) Symposium, pp. 722–726 (2002)
6. Pustejovsky, J., Castan'o, J., Zhang, M.K., Cochran, B.: Robust relational parsing over biomedical literature: extracting inhibit relations. *Proc. Pacific Symp. Biocomput.* **7**, 362–373 (2002)
7. Lastra, J.L.M., Delamer, M.: Semantic web services in factory automation: fundamental insights and research roadmap. *IEEE Tran.* **2**(1), 1–11 (2006)
8. Dong, H., Hussain, F.K., Chang, E.: State of the art in semantic focused crawlers. In: International Conference on ICCSA, pp. 890–904. Yongin, Korea (2009)
9. Johnson H.L., Cohen K.B., Baumgartner, W.A., et al.: Evaluation of lexical methods for detecting relationships between concepts from multiple ontologies. *Proc. Pacific Symp. Biocomput.* (2006)
10. Frunza, O., Inkpen, D., Thomas, T.: A machine learning approach for identifying disease treatment relations in short texts. *IEEE Trans. Knowl. Data Eng.* **23**(6), 801–814 (2011)
11. Thomas, J., Milward, D., Ouzounis, C., Pulman, S., Carroll, M.: Automatic extraction of protein interactions from scientific abstracts. *Proc. Pacific Symp. Biocomput.* **5**, 538–549 (2000)
12. Hunter, L., Cohen, K.B.: Biomedical language processing: what's beyond PubMed? *Mol. Cell* **21–5**, 589–594 (2006)
13. Ganesh, S., Jayaraj, M., Kalyan, V., Aghila, G.: Ontology-based web crawler. In: Proceedings of the ITCC: Coding Computing, pp. 337–341. Las Vegas, NV (2004)
14. Hunter, L., Lu, Z., et al: OpenDMP: an open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. *BMC Bioinf.* **9**(78) (2008)
15. Dong, H., Hussain, F.K., Chang, E.: A survey in semantic web technologies inspired focused crawlers. In: 3rd International Conference of ICDIM, pp. 934–936. East London, UK (2008)
16. Rosario, B., Hearst, M.A.: Semantic relations in bioscience text. In: 42nd—Annual Meeting on Association for Computational Linguistics, vol. 430 (2004)

17. Ray, S., Craven, M.: Representing sentence structure in hidden Markov models for information extraction. *Int. J. CAI* (2001)
18. Yuvarani, M., Iyengar, N.C.S.N., Kannan, A.: LSCrawler: a framework for an enhanced focused web crawler based on link semantics. In: *Proceedings of the IEEE/WIC/ACM International Conference WI*, pp. 794–800 (2006)
19. Antonie, M.L., Zaane, O.R.: Text document categorization by term association. In: *Proceedings of the IEEE International Conference Data Mining (ICDM '02)* (2002)
20. Tane, J., Schmitz, C., Stumme, G.: Semantic resource management for the web: an e-learning application. In: *WWW 2004*, pp. 1–10. ACM, New York (2004)
21. Heb, A., Kushmerick, N.: Automatically attaching semantic metadata to web services. In: *Proceeding IJCAI Workshop Information Integration on the Web* (2003)
22. Francesconi, E., Peruginelli, G.: Searching and retrieving legal literature through automated semantic indexing, pp. 131–138. *ICAIL*, Stanford, CA (2007)
23. Giles, C.L., Petinot, Y., et al.: e-BizSearch: a niche search engine for e-business. In: *Proceeding for SIGIR*, pp. 413–414. Toronto, ON, Canada (2003)
24. Bruno, M., Canfora, G., Penta M.D., Scognamiglio, R.: An approach to support web service classification and annotation. In: *Proceeding IEEE International Conference. E-Tech, E-Comm and E-Service (EEE-05)* (2005)
25. Halkidi, M., Varlamis, N.B., Vazirgiannis, M.: THESUS: Organizing web document collections based on link semantics. *VLDB J.* **12**(4), 320–332 (2003)
26. Klusch, M., Zhing, X.: Deployed semantic services for the common user of the web: a reality check. In: *Proceeding for IEEE International Semantic Computing (ICSC)* (2008)
27. Leroy, G., Chen, H.C., Martinez, J.D.: A shallow parser based on closed-class words to capture relations in biomedical text. *J. Biomed. Inform.* **36**(3), 145–158 (2003)
28. Giuliano, C., Alberto, L., Lorenza, R.: Exploiting shallow linguistic information for relation extraction from biomedical literature. In: *Proceedings of the 11th Conference European Chapter of the Association for Computational Linguistics* (2006)
29. Xiong H., Tan, P., Kumar, V.: Mining strong affinity association patterns in data sets with skewed support distribution. In: *Proceeding for IEEE Third International Conference of Data Mining* (2003)
30. Goadrich, M., Oliphant, L., Shavlik, J.: Learning ensembles of first-order clauses for recall-precision curves: a case study in biomedical information extraction. In: *Proceedings of the 14th International Conference* (2004)

Real-Time Graphs for Communication Networks: A Fuzzy Mathematical Model

Siddhartha Sankar Biswas, Bashir Alam and M.N. Doja

Abstract For a given alive network, in many situations, its complete topology may not always be available because of the reason that few of its links may be temporarily disabled. Thus, at any real-time instant, only a subgraph, rather than the complete graph may be available to the system for executing its activities. Besides that, in most of the cases, the cost parameters corresponding to its links are not crisp numbers, rather fuzzy numbers. Under such circumstances, none of the existing algorithms on the shortest path problems or fuzzy shortest path problem can work. In this paper, the authors propose a mathematical model for such types of graphs to be called by real time graphs (RT-graphs) in which all real-time information (updated every q quantum of time) are incorporated so that the network can serve very efficiently with optimal results. Although the style of Dijkstra's Algorithm is followed, the approach is a completely new in the sense that the SPP is solved with the real-time information of the network.

Keywords RT-graphs · TBL · LS · LSC · TBN · RN · Fuzzy shortest path estimate · RT fuzzy relaxation

1 Introduction

Many problems of computer science, communication network, transportation systems, etc., can be modeled into graphs and then can be solved. Graph theory [1–3] has wide applications in several branches of Engineering, in particular in Computer Science, Communication systems, Civil Engineering, etc., and also in, Science, Social Science, Optimization, Management Science, Medical Science, Economics, etc., to list a few only out of many.

S.S. Biswas (✉) · B. Alam · M.N. Doja
Department of Computer Engineering, Faculty of Engineering & Technology,
Jamia Millia Islamia, New Delhi 110025, India
e-mail: ssbiswas1984@gmail.com

Besides that, in most of the cases of such directed graphs, the real data about the weights of the arcs are not always crisp but fuzzy numbers. In this paper, we propose a generalized notion of graphs which is a highly flexible and appropriate model as it incorporates the real-time information of the network problem to facilitate the decision maker to search for efficient and optimized results/solutions. We call such graphs by “Real-Time Graphs” or “RT-graphs.” Clearly a RT-graph is a variable representation of a network with respect to time. We then develop an algorithm to solve the Fuzzy Shortest Path Problem (FSPP) in RT-graphs. Consequently, our algorithm may produce different results at different times of start, but at each time, the results will be most efficient for that real time only.

2 Preliminaries

For details about the fuzzy set theory [2] of Zadeh, one could see any good book viz. [4]. The concept of fuzzy numbers (popularly used are triangular fuzzy numbers and trapezoidal fuzzy numbers) is of importance for quantifying an ill-known quantity. In our work here throughout, we use the operations like fuzzy addition \oplus , fuzzy subtraction \ominus , “ranking” of fuzzy numbers, etc. (we use either triangular fuzzy numbers (TFN) or trapezoidal fuzzy numbers).

Trivially, any crisp real number can be viewed as a fuzzy number [5]. There is no unique method for ranking n number of fuzzy numbers, because all the existing methods [1, 6–9] are soft-computing methods and each method has got merits and demerits depending upon the properties of the application domains and the problem under consideration. However, if $A_1, A_2, A_3, \dots, A_n$ be n fuzzy numbers sorted in fuzzy ascending order by a pre-decided method, i.e., if $A_1 \prec A_2 \prec A_3 \dots \prec A_n$, then A_1 and A_n are called, respectively, the fuzzy-min and fuzzy-max of these n fuzzy numbers.

3 Real-Time Graphs : A Fuzzy Model

We know that a graph G is an ordered pair (V, E) which consists of two sets V and E , where V is the set of vertices (or, nodes), and E is the set of edges (or, arcs). Here, we consider such graphs where no loop exists. Graphs may be of two types: undirected multigraphs and directed multigraphs. In an undirected graph, the edge (i, j) and the edge (j, i) , if exist, are obviously identical unlike in the case of directed graph.

In our work, here we consider more real situations which are actually and frequently faced by the present day communication systems. For example, consider an Adhoc Network or a MANET where there may exist a path from node u to node v , but because of some reasons, this path may be temporarily damaged and hence temporarily unavailable for transmission of packets by the node u to its neighbor

node v . This is a very useful information to the communication system if available to the sender node u in advance. We incorporate such kind of real-time situation [5] in our proposed model here to define a new structure called by Real-Time Graph or RT-graph.

3.1 “Neighbor” Node

For a given node u , the node v will be designated as a ‘neighbor’ node of u if u has a directed link from u to v .

3.2 Link Status (LS)

At any real instance of time, corresponding to every neighbor node v , there exists a Link Status i_{uv} which takes any of the two values from $\{0, 1\}$ with the following significance:

$i_{uv} = 0$, if the link uv is non-functional at that time
 $= 1$, if the link uv is functional at that time

3.3 Temporarily Blocked Link

If at any given instance of time, i_{uv} happens to be 0, i.e., the link uv is non-functional, then we say that the link uv is a Temporarily Blocked Link. It may be noted that, since we are dealing with real-time situations, therefore after q quantum of time, the situation might change, that is why the word Temporary being used.

3.4 Link Status Class (LSC) Vector

For a given node u , the collection of all LS is called “Link Status Class” (LSC) of u denoted by the vector I_u . If node u has k (≥ 0) number of neighbor nodes $x_1, x_2, x_3, \dots, x_k$, then

$$I_u = \{i_{ux1}, i_{ux2}, i_{ux3}, \dots, i_{uxk}\}.$$

3.5 Temporarily Blocked Neighbor (TBN) and Reachable Neighbor (RN)

If v is a neighbor node of a given node u and if i_{uv} is 0 at a given instant of time, then v is called a Temporarily Blocked Neighbor (TBN) of u for that instant (see Fig. 1)

If a neighbor v is not a tbn, then it is called a reachable neighbor (rn) of u (see Fig. 2).

3.6 Communicable Node

For a given node u , if $I_u \neq \varnothing$ (null vector), then the node u is called a communicable node for that instant of time.

4 Fuzzy Shortest Path in a RT-Graph

Graphs are a very important model of network. There are many real-life problems of network which are modeled into graphs and hence solved. The dynamic model “RT-graph” is a new concept on the generalization of graphs considering its huge potential for real-time applications in communication or transportation systems. In this section, we solve the FSPP for RT-graphs where we use the notion of Dijkstra’s Algorithm but with simple soft-computations with real-time data, unlike the existing algorithms for FSPP.

Consider a directed RT-graph G where the arcs are of fuzzy weights (fuzzy numbers). Suppose that v is an rn of u in G with link status i_{uv} . Suppose that the subalgorithm $FW(G)$ returns the fuzzy weight set W .

4.1 Fuzzy Shortest Path Estimate $D[V]$ of a RN Vertex V in a Directed RT-Graph

Consider a directed RT-graph G where the arcs are of fuzzy weights (fuzzy numbers). Suppose that v is an rn of u in G with link status i_{uv} . Suppose that the subalgorithm $FW(G)$ returns the fuzzy weight set W .

Fig. 1 Temporarily blocked neighbor

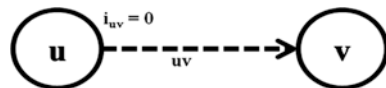
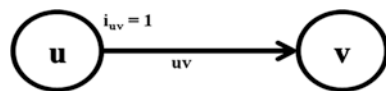


Fig. 2 Reachable neighbor



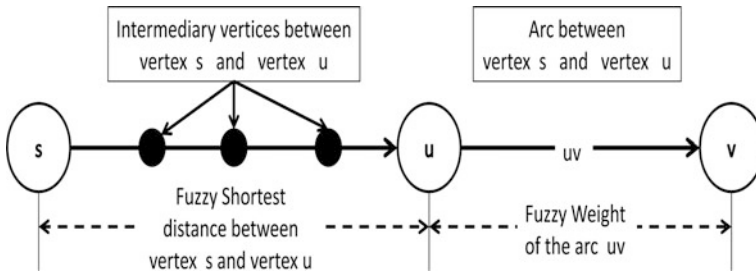


Fig. 3 Fuzzy shortest path estimate

The estimate $d[v]$ (see Fig. 3) in RT-graphs is computed on real-time data only. Suppose that s is the source vertex and the currently traversed vertex is u . At this real time, we check whether u is a communicable node or not. If yes, then we proceed further. If v is not a RN, then $d[v]$ needs not to be estimated. Otherwise, $d[v]$ of any RN vertex v , in a directed RT-graph, is computed using fuzzy addition as below:

$$\begin{aligned}
 &(\text{fuzzy shortest path estimate of } v) = (\text{fuzzy shortest path estimate of } u) \oplus (\text{fuzzy weight corresponding to the arc from the vertex } u \text{ to the vertex } v). \\
 &\text{or, } d[v] = d[u] \oplus w_{uv}.
 \end{aligned}$$

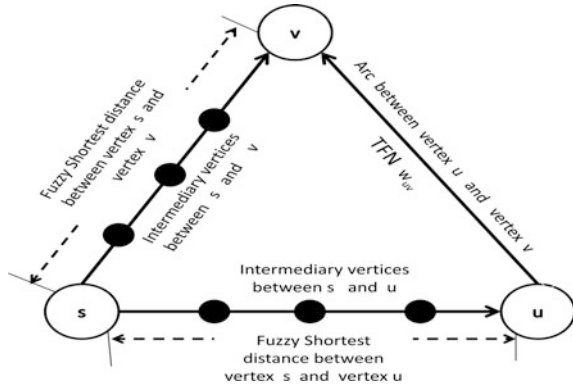
4.2 Fuzzy Relaxation of an Arc in a Directed RT-Graph

We extend the classical notion of relaxation to the case of RT-graph. At this real time, the important deviation from the classical method is that we do not attempt to do relaxation of an arc/link if it is a TBL. By fuzzy relaxation, we shall mean here the relaxation process of an arc whose weight is a fuzzy number (or at best a crisp number). For this, first of all we initialize the graph along with its starting vertex and fuzzy shortest path estimate for each vertices of the graph G . The following “FUZZY-INITIALIZATION-SINGLE-SOURCE (FISS)” algorithm will do:

- FISS (G, s)
1. For each vertex $v \in V[G]$
 2. $d[v] = \infty$
 3. $v.\pi = \text{NIL}$
 4. $d[s] = 0$

After the fuzzy initialization, the process of fuzzy relaxation of each arc begins but to be applicable at real time only, and hence, tbl cases are excluded from the execution of relaxation jobs. The subalgorithm RT-FUZZY-RELAX (see Fig. 4) plays the vital role to update $d[v]$, i.e., the fuzzy shortest distance value between the starting vertex s and the vertex v , which is an rn of the current traversed vertex u .

Fig. 4 RT-FUZZY-RELAX algorithm in a RT-graph



```

RT-FUZZY-RELAX (u, v, W)
1 IF u is a communicable node, THEN
2   IF v is a rn, THEN
3     IF d[v] > d[u] * wuv
4       THEN d[v] ← d[u] * wuv
5         v.π ← u
    
```

where, w_{uv} W is the fuzzy weight of the arc from the vertex u to the vertex v , and $v.\pi$ denotes the parent node of vertex v .

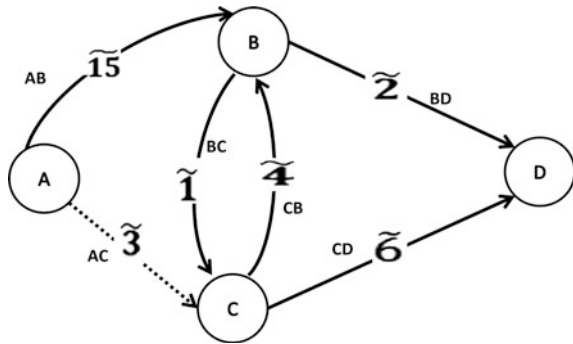
4.3 Real-Time Fuzzy Shortest Path(RT-FSP) Algorithm in a RT-Graph

We now present our main algorithm to find single-source fuzzy shortest path in a RT-graph. We name this by “Real-Time Fuzzy Shortest Path Algorithm,” i.e., in short by the title RT-FSPA. In this algorithm, we use the above subalgorithms and also the subalgorithm RT-EXTRACT-FUZZY-MIN (Q) which extracts the node u with minimum key using a pre-decided fuzzy ranking method and updates Q at real time only.

```

RT-FSPA (G, s)
1 FISS (G, s)
2 W ← FW(G)
3 S ← ∅
4 Q ← V[G]
5 WHILE Q ≠ ∅
6   DO u ← RT-EXTRACT-FUZZY-MIN (Q)
7     S ← S ∪ {u}
8   FOR each RN vertex v ∈ Adj[u]
9     DO RT-FUZZY-RELAX (u, v, W)
    
```

Fig. 5 RT-graph G



4.4 Example Explaining RT-FSP Algorithm

Consider the directed RT-Graph G (shown in Fig. 5 below) where the fuzzy weights are shown against each link.

We want to solve the single-source fuzzy shortest paths problem taking the vertex A as the source vertex and the vertex D as the destination vertex, where the LSCs of all the nodes of the RT-Graph G are $I_A, I_B, I_C,$ and I_D given by

$$\begin{aligned}
 I_A &= \{i_{AB}, i_{AC}\} \text{ where } i_{AB} = 1, i_{AC} = 0; \\
 I_B &= \{i_{BC}, i_{BD}\} \text{ where } i_{BC} = 1, i_{BD} = 1; \text{ and} \\
 I_C &= \{i_{CB}, i_{CD}\} \text{ where } i_{CB} = 1, i_{CD} = 1;
 \end{aligned}$$

Clearly the RT-FSPA algorithm computes to yield the following result:

1. $w_{AB} = \widetilde{10}, w_{AC} = \widetilde{3}, w_{BC} = \widetilde{1}, w_{CB} = \widetilde{4}, w_{CD} = \widetilde{6}$ and $w_{BD} = \widetilde{2}$;
2. $S = \{A, B, D\}$, i.e., the RT fuzzy shortest path from the source vertex A is:

$$A \rightarrow B \rightarrow D.$$

3. d -values, i.e., RT fuzzy shortest distance estimate values of each vertex from the starting vertex A is

$$d[A] = 0, d[C] = \text{not estimated}, d[B] = \widetilde{15}, d[D] = \widetilde{17}.$$

5 Conclusion

There are many real-life problems in the networks of transportation, communication, circuit systems, etc., which are initially modeled into graphs and hence solved. But in reality, it is very frequent that due to some reason, few links may be temporarily unavailable to the communication system. Consequently, the complete topology of the graph of a network may not be available to the system but a

subgraph of it. Besides that in many of these directed graphs, the weights of the arcs are not always crisp but fuzzy numbers. In this paper, we make a mathematical modeling of such real-time status of a network by introducing a generalized structure called by “RT-graph” or “RT-graph.” The real-time data/information is updated at each node of the RT-graph at every quantum time q . We then develop a new method to solve the fuzzy shortest path problem (FSPP) from a source vertex to a destination vertex in a directed RT-graph. The importance of our method lies in its potential to give solution in real-time environment, unlike any of the existing algorithms of FSPP. Obviously, our algorithm may produce different results at different real instants of time, but the results will surely significant for that real time.

References

1. Abbasbandy, S.: Ranking of fuzzy numbers, some recent and new formulas. *IFSA-EUSFLAT* **2009**, 642–646 (2009)
2. Allahviranloo, T., Abbasbandy, S., Saneifard, R.: A method for ranking of fuzzy numbers using new weighted distance. *Math. Comput. Appl.* **16**(2), 359–369 (2011)
3. Sujatha, L., Elizabeth, : Fuzzy shortest path problem based on similarity degree. *Appl. Math. Sci.* **5**(66), 3263–3276 (2011)
4. Biswas, S.S., Alam, B., Doja, M.N.: A theoretical characterization of the data structure ‘multigraphs’. *J. Contemp. Appl. Math.* **2**(2), 88–106 (2012)
5. Biswas, S.S., Alam, B., Doja, M.N.: A GRT-multigraphs for communication networks : a fuzzy theoretical model. In: *International Symposium on System Engineering and Computer Simulation (SECS-2013)*. *Advanced in Computer Science and its Applications*, pp. 633–641. Danang, Vietnam, 18–21 Dec 2013. Print ISBN: 978-3-642-41673-6
6. Dat, L.Q., Yu, V.F., Chou, S.-Y.: An improved ranking method for fuzzy numbers using left and right indices. In: *2nd International Conference on Computer Design and Engineering, IPCSIT*, vol. 49, pp 89–94 (2012). doi:[10.7763/IPCSIT.2012.V49.17](https://doi.org/10.7763/IPCSIT.2012.V49.17)
7. Biswas, R.: Fuzzy numbers redefined. *Information* **15**(4), 1369–1380 (2012)
8. Parandin, N, Araghi, M.A.F.: Ranking of fuzzy numbers by distance method. *J. Appl. Math.* **5**(19), 47–55 (2008) (Islamic Azad University of Lahijan)
9. Zadeh, L.A.: Fuzzy sets. *Inform. Control* **8**, 338–353 (1965)

PWLCM-Based Random Search for Strong Substitution-Box Design

Musheer Ahmad, Danish Raza Rizvi and Zishan Ahmad

Abstract Substitution-boxes are the only source of nonlinearity in various symmetric encryption systems and responsible for inducing confusion of plaintext data. The robustness of these systems exclusively depends on the potentiality of S-boxes. The design methods having fast and simple computations which can yield effective S-boxes are preferred. In this paper, a new chaos-based random search is applied to construct cryptographically potent 8×8 S-box. The method explores the features of piecewise linear chaotic map for candidate generation and random search. The optimized S-box obtained is tested against standard statistical tests like bijectivity, equiprobable I/O XOR distribution, nonlinearity, and strict avalanche criteria revealing its superior performance. The proffered substitution-box is further compared with some contemporary chaotic substitution-boxes. The results confirm a consistent design, suitable for building strong block encryption systems.

Keywords Random search · Piecewise linear chaotic map · Substitution-box · Nonlinearity · Secure communication

1 Introduction

The seminal work of Shannon, in 1949, became the foundational treatment of modern cryptography [1]. He identified two fundamental properties of *confusion* and *diffusion* for the design of cryptographically secure and strong encryption systems.

M. Ahmad (✉) · D.R. Rizvi · Z. Ahmad
Department of Computer Engineering, Faculty of Engineering and Technology,
Jamia Millia Islamia, New Delhi 110025, India
e-mail: musheer.cse@gmail.com

D.R. Rizvi
e-mail: danish.jmi@gmail.com

Z. Ahmad
e-mail: zishanahmad@outlook.com

The property of *confusion* refers to making the relationship between the ciphertext and the secret key as complex and involved as possible. On the other hand, the *diffusion* refers to dissipating the statistical structure of plaintext over bulk of ciphertext. The property of confusion is accomplished through substitution-box (S-box) while the diffusion is accomplished through a permutation box (P-box). A way to achieve both confusion and diffusion in encryption system is to use substitution-permutation networks (SPN). A substitution-permutation network takes a block of plaintext and the keys as inputs and applies several alternating layers of S-boxes and P-boxes to produce the ciphertext block [2]. Some ciphers using the SPNs are 3-Way, SAFER, SHARK, and Square. An efficient S-box with thorough cryptographic features is essential for the development of strong encryption systems. S-boxes are the only source of nonlinearity in the security system. Nonlinearity is required in security systems since it reduces the input and output correlation as well as the correlation of their linear combination [3]. Often implemented as a lookup table, an $m \times n$ S-box takes m input bits and transforms them into n output bits. Mathematically, an S-box of size $m \times n$ is a nonlinear mapping defined as $S: \{0, 1\}^m \rightarrow \{0, 1\}^n$ [4].

For a closed invariant set, chaotic systems are dynamical systems which exhibit dependence on the initial conditions. The features of chaotic systems have attracted the attention of researchers all over the world and are utilized to build strong chaotic security systems for the protection of data. Chaotic systems, unlike traditional cryptosystems, are considered adept for providing secure communication having low computational overheads [5]. Several methods have been proposed as a result of the ongoing research attempting to design strong chaotic S-boxes. In this paper, a novel scheme to synthesize chaos-based S-box based on chaotic map using random search is presented. The generated S-box is analyzed and its comparative performance assessment shows that the substitution-box has strong cryptographic potency.

The rest of the paper has been segmented as follows: Sect. 2 consists of the description of the used chaotic system with the proposed designing technique of the S-box. The quantification, comparison, and analysis of the performance of the proposed S-box have been described in Sect. 3. Section 4 includes the conclusion of the work.

2 The Method

The basic concepts of the design and construction methodology of the proposed S-box are mentioned in the following subsections.

2.1 Piecewise Linear Chaotic Map

Considered among the most studied chaotic systems, the piecewise linear chaotic map is a one-dimensional dynamical system with its system equation described as follows: [6]:

$$x(n + 1) = \begin{cases} \frac{x(n)}{p} & 0 < x(n) \leq p \\ \frac{1-x(n)}{1-p} & p < x(n) < 1 \end{cases} \tag{1}$$

In the above formula, n is the number of iterations, x is state of the system, $x(n) \in (0,1)$ for all n and $p \in (0, 1)$ is system parameter. The lyapunov exponent of a dynamical system specifies the rate of separation of minutely close trajectories. For discrete dynamical systems $x(n+1) = g(x(n))$, with initial value $x(0)$, the lyapunov exponent is defined in Eq. (2). If at least one lyapunov exponent of a system is positive, then the system is considered chaotic. PWLCM has positive lyapunov exponent for $p \in (0, 1)$ and its largest value is reckoned at $p = 0.5$. The system trajectory of PWLCM $x(n)$ visits the entire interval $(0, 1)$ for every value of p . Unlike the logistic map, PWLCM has no nonchaotic regions [7] and has been studied to exhibit better features than the logistic map [8]. The initial values assigned to $x(0)$ and p for execution of system act as secret key.

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} \ln|g'(x(k))| \tag{2}$$

2.2 Proposed Method

The procedure of the proposed chaos-based random search method is as follows:

- A.1.** Initialize p, x_0, cnt, GB_SBox and set $\Delta = 256, k = 0$.
- A.2.** Take S_1, S_2 and S_3 as empty sets and $k = k + 1$.
- A.3.** Iterate PWLCM map (1) for cnt times and discard the values except the last.
- A.4.** Further iterate the map (1) once and sample the chaotic variable x .
- A.5.** Extract a random number $m \in [0, 255]$ from x as:

$$m = \{floor(x * 10^{14})\} mod(256)$$

```

IF (m is not in S1)
    THEN add m to S1
ELSE IF (m is not in S2)
    THEN add m to S2
ELSE IF (m is not in S3)
    THEN add m to S3

IF (|S1|=Δ & |S2|=Δ & |S3|=Δ)
    THEN goto step A.6
ELSE
    THEN goto step A.4
    
```

- A.6.** Translate S_1, S_2 and S_3 to 16×16 S-Box tables.
- A.7.** Evaluate local best among S_1, S_2 and S_3 and marked it as LB_SBox .
- A.8.** IF local best S-Box is better than global best GB_SBox , THEN update GB_SBox .

3 Performance Evaluation

It is essential for the S-box to withstand attacks such as linear and differential cryptanalysis. The cryptographic strengths of substitution-boxes need to be gauged before their usage. Among cryptologists, the widely accepted and significant S-box strength measures to thwart the above attacks are *bijectionity*, *nonlinearity*, *strict avalanche criteria*, and *equiprobable I/O XOR distributions* [9–11]. Therefore, these parameters are adopted to quantify the potency of proposed S-box so as to find its applicability for block encryption systems. The performance is also compared with few of the recent chaos-based S-boxes. The initial values taken for simulation are: $x(0) = 0.56789$, $p = 0.567$, $N_0 = 100$, and $cnt = 1000$. The chaotic S-box constructed using proposed method is provided in Table 1.

3.1 Bijectionity

A Boolean function f_i is considered bijective if it holds the relation [12]: $wt(\sum_{i=1}^n a_i f_i) = 2^{n-1}$, where $a_i \in \{0, 1\}$, $(a_1, a_2, \dots, a_n) \neq (0, 0, \dots, 0)$ and $wt(\cdot)$ is hamming weight. In other words, an $n \times n$ S-box is said to be bijective if it has all different output values from the interval $[0, 2^n - 1]$. The S-box obtained with the proposed method has been verified to be bijective.

Table 1 Proposed substitution-box

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	89	20	164	236	81	161	56	39	66	240	40	125	158	145	60	113
1	67	138	153	9	108	10	222	49	24	105	13	44	84	63	174	68
2	36	229	47	241	218	220	216	11	232	25	102	192	109	111	3	233
3	71	242	237	34	230	188	219	85	12	197	17	176	18	4	208	123
4	61	73	206	149	204	69	195	135	239	5	65	175	226	140	181	31
5	26	14	231	107	156	120	249	116	162	58	163	29	157	184	30	193
6	150	74	83	186	76	205	52	22	1	144	143	225	32	170	87	196
7	59	2	104	106	165	46	86	179	133	88	173	77	154	43	147	6
8	171	114	51	199	235	224	79	136	214	207	185	183	213	110	142	62
9	244	227	7	35	167	155	70	82	137	172	100	198	201	101	180	215
10	210	38	130	177	95	251	33	254	8	190	191	78	97	112	187	132
11	166	139	42	45	151	252	50	228	159	37	141	48	202	200	134	96
12	169	54	248	178	121	124	98	99	41	246	146	238	189	128	115	255
13	217	223	15	92	160	250	75	221	28	16	168	21	19	122	72	182
14	131	253	148	91	129	127	23	118	80	203	152	212	247	90	53	27
15	117	57	243	245	194	126	119	94	103	0	234	209	93	64	55	211

3.2 Nonlinearity

Nonlinearity of a Boolean function f is defined as its minimum distance to all affine functions. The set of affine Boolean functions is the set of linear Boolean functions and their complements [13]. It is necessary for a strong S-box to have high nonlinearity scores. A Boolean function with low nonlinearity is prone to linear approximation attacks. The nonlinearity NL_f of a Boolean function $f(x)$ is determined as:

$$NL_f = 2^{n-1}(1 - 2^{-n} \max |S_{(f)}(w)|)$$

$$S_{(f)}(w) = \sum_{w \in GF(2^n)} (-1)^{f(x) \oplus x.w}$$

$S_{(f)}(w)$ is the Walsh spectrum of $f(x)$ and $x.w$ denotes the dot product of x and w . For the proposed S-box, the nonlinearity scores for the eight Boolean functions are 104, 108, 110, 108, 106, 102, 108, and 108 with an average of 106.75. After a comparison of the nonlinearity scores of the proposed S-box with those of the existing ones, the obtained results have been shown in Table 2. The proposed S-box provides higher min, max, and mean of nonlinearity scores than their counterparts. The average nonlinearity of the proposed S-box is greater than 105 meaning that the proposed S-box exhibits inherent high nonlinearity.

3.3 Strict Avalanche Criteria

Introduced by Webster and Tavares, in the year 1985, the Strict Avalanche Criteria is a generality of the avalanche effect and completeness. For a Boolean function satisfying the strict avalanche criteria, a change in a single bit might result in the change of half of the output bits. Following the procedure to check whether an

Table 2 Nonlinearity scores of 8×8 S-boxes

S-box	Nonlinearities										
	1	2	3	4	5	6	7	8	min	max	Mean
Proposed	104	108	110	108	106	102	108	108	102	110	106.75
Jakimoski and Kocarev [12]	98	100	100	104	104	106	106	108	98	108	103.25
Tang et al. [14]	103	102	104	101	108	106	102	105	101	108	103.87
Chen et al. [15]	100	102	103	104	106	106	106	108	100	108	104.38
Wang et al. [16]	104	106	106	102	102	104	104	102	102	106	103.75
Özkaynak and Özer [17]	104	100	106	102	104	102	104	104	100	104	103.25
Khan et al. [18]	108	102	100	104	104	102	98	106	98	108	103
Khan and Shah [19]	100	108	106	104	102	102	106	108	100	108	104.5
Gondal et al. [20]	98	100	106	104	106	100	106	104	98	106	103

Table 3 Dependency matrix of proposed S-box

1	2	3	4	5	6	7	8
0.4843	0.5625	0.4531	0.5312	0.5156	0.5781	0.5000	0.4843
0.4531	0.5312	0.5625	0.5312	0.4843	0.5625	0.4531	0.4375
0.5625	0.4062	0.5156	0.5625	0.5000	0.5000	0.5156	0.5468
0.5937	0.5468	0.4687	0.4375	0.5000	0.4843	0.5312	0.5781
0.5156	0.5000	0.5625	0.5468	0.4531	0.5000	0.4531	0.5312
0.5312	0.4843	0.4843	0.5000	0.5000	0.4687	0.5156	0.5312
0.5468	0.4531	0.4843	0.4843	0.4531	0.4218	0.4687	0.4531
0.4843	0.4843	0.5781	0.5000	0.5468	0.5156	0.4687	0.5000

Table 4 SAC and maximum differential probability of S-boxes

S-box	SAC	max DP
Proposed	0.5046	10/256
Jakimoski and Kocarev [12]	0.4972	12/256
Tang et al. [14]	0.5058	14/256
Chen et al. [15]	0.4999	12/256
Wang et al. [16]	0.4964	10/256
Özkaynak and Özer [17]	0.5048	10/256
Khan et al. [18]	0.5012	12/256
Khan and Shah [19]	0.4978	12/256
Gondal et al. [20]	NR	12/256

S-box satisfies the strict avalanche criteria [9], a dependency matrix has been calculated as shown in Table 3. The strict avalanche criteria value of the proposed S-box has been calculated as 0.5046. This value is much close to the ideal value of 0.5 and reveals its excellent performance. A comparison has been drawn in Table 4. As is evident from the table, the proposed S-box provides comparable parameter values with respect to SAC.

3.4 Equiprobable I/O XOR Distribution

Introduced by Biham and Shamir, differential cryptanalysis exploits the disparity of the distribution of the input and output to attack S-boxes [10]. Resistance toward differential cryptanalysis can be accomplished if the XOR value of every output is equally probable as the XOR value of every input. The S-box is resistant against differential cryptanalysis if it is closed in I/O probability distribution. The largest value of DP should be minimal. For a Boolean function $f(x)$, the differential probability is defined as:

$$DP_f = \max_{\Delta x \neq 0, \Delta y} \left(\frac{\#\{x \in X | f(x) \oplus f(x \oplus \Delta x) = \Delta y\}}{2^n} \right)$$

In the above formula, X is the set of all the input values that are possible. The number of its elements are 2^n (here $n = 8$). Table 4 depicts the largest values of DP obtained for the proposed S-box. The largest DP is 10/256 which is also the largest value in Wang's and Özkaynak's S-boxes. The value of differential probability is better than the Gondal's, Jakimoski's, Chen's, Khan's S-boxes, where the maximum value is 12/256, and Tang's value is 14/256. This confirms that in terms of differential cryptanalysis the proposed S-box is stronger than Tang's, Chen's, Khan's, Gondal's, and Jakimoski's S-boxes and comparable to others.

4 Conclusion

In this paper, a new chaos-based random search method is presented to synthesis a cryptographically cogent 8×8 substitution-box. The piecewise linear chaotic map, which has better features than the well-studied logistic map, is selected for initial S-box candidates generation and random search. The proposed method is simple, fast, and credible to construct efficient substitution-boxes. The optimized S-box is tested against some standard statistical parameters to resist linear and differential attacks. It has been verified that proposed S-box has excellent performance strength. Hence, the experimental analyses and comparison with some recent chaotic S-boxes show that the proposed method is testimonial in synthesizing impressive S-boxes.

References

1. Shannon, C.E.: Communication theory of secrecy systems. *Bell Syst. Tech. J.* **28**, 656–715 (1949)
2. Menezes, A.J., Oorschot, P.C.V., Vanstone S.A.: *Handbook of Applied Cryptography*, CRC Press (1997)
3. Ahmad, M., Khan, P.M., Ansari, M.Z.: A simple and efficient key-dependent S-box design using fisher-yates shuffle technique. In: Martínez, P.G., Thampi, S.M., Ko, R., Shu, L. (eds.) *SNDS 2014. CCIS*, vol. 420, pp. 540–550. Springer, Heidelberg (2014)
4. Ahmad, M., Chugh, H., Goel, A., Singla, P.: A chaos based method for efficient cryptographic S-box design. In: Thampi, S.M., Atrey, P.K., Fan, C.-I., Pérez, G.M. (eds.) *SSCC 2013. CCIS*, vol. 377, pp. 130–137. Springer, Heidelberg (2013)
5. Alvarez, G., Li, S.: Some basic cryptographic requirements for chaos-based cryptosystems. *Int. J. Bifurcat. Chaos* **16**(8), 2129 (2006)
6. Li, S., Chen, G., Mou, X.: On the dynamical degradation of digital piecewise linear chaotic maps. *Int. J. Bifurcat. Chaos* **15**(10), 3119–3151 (2005)
7. Kanso, A., Yahyaoui, H., Almulla, M.: Keyed hash function based on a chaotic map. *Inf. Sci.* **186**(1), 249–264 (2012)

8. Hermassi, H., Rhouma, R., Belghith, S.: Improvement of an image encryption algorithm based on hyper-chaos. *Telecommun. Syst.* **52**(2), 539–549 (2013)
9. Webster, A.F., Tavares, S.E.: On the design of S-boxes. *Adv. Cryptology Lect. Notes Comput. Sci.* **218**, 523–534 (1986)
10. Biham, E., Shamir, A.: Differential cryptanalysis of DES-like cryptosystems. *J. Cryptology* **4**(1), 3–72 (1991)
11. Youssef, A.M., Tavares, S.E., Gong, G.: On some probabilistic approximations for AES-like S-boxes. *Discrete Math.* **306**(16), 2016–2020 (2006)
12. Jakimoski, G., Kocarev, L.: Chaos and cryptography: block encryption ciphers based on chaotic maps. *IEEE Trans. Circ. Syst.* **48**(2), 163–169 (2001)
13. Cusick, T.W., Stanica, P.: *Cryptographic Boolean Functions and Applications*. Elsevier, Amsterdam (2009)
14. Tang, G., Liao, X., Chen, Y.: A novel method for designing S-boxes based on chaotic maps. *Chaos, Solitons Fractals* **23**(2), 413–419 (2005)
15. Chen, G., Chen, Y., Liao, X.: An extended method for obtaining S-boxes based on three-dimensional chaotic Baker maps. *Chaos, Solitons Fractals* **31**(3), 571–577 (2007)
16. Wang, Y., Wong, K.W., Liao, X., Xiang, T.: A block cipher with dynamic S-boxes based on tent map. *Commun. Nonlinear Sci. Numer. Simul.* **14**(7), 3089–3099 (2009)
17. Özkaynak, F., Özer, A.B.: A method for designing strong S-boxes based on chaotic Lorenz system. *Phys. Lett. A* **374**(36), 3733–3738 (2010)
18. Khan, M., Shah, T., Mahmood, H., Gondal, M.A.: An efficient method for the construction of block cipher with multi-chaotic systems. *Nonlinear Dyn.* **71**(3), 489–492 (2013)
19. Khan, M., Shah, T.: An efficient construction of substitution box with fractional chaotic system. *SIViP* (2013). doi:[10.1007/s11760-013-0577-4](https://doi.org/10.1007/s11760-013-0577-4)
20. Gondal, M.A., Raheem, A., Hussain, I.: A scheme for obtaining secure S-boxes based on chaotic Baker's map 3D. *Research* **5**(3), 1–8 (2014)
21. Yin, R., Yuan, J., Wang, J., Shan, X., Wang, X.: Designing key-dependent chaotic S-box with large key space. *Chaos, Solitons Fractals* **42**(4), 2582–2589 (2009)
22. Szaban, M., Serebinski, F.: Designing cryptographically strong S-boxes with the use of cellular automata. *Ann. UMCS Informatica Lublin-Polonia Sectio AI* **8**(2), 27–41 (2008)

Fuzzy C-Means and Fuzzy TLBO for Fuzzy Clustering

P. Gopala Krishna and D. Lalitha Bhaskari

Abstract The choice of initial center plays a great role in achieving optimal clustering results in all partitional clustering approaches. Fuzzy C-means is a widely used approach but it also gets trapped in local optima values due to sensitiveness to initial cluster centers. To alleviate this issue, a new approach of using an evolutionary technique known as Teaching–Learning-Based Optimization (TLBO) is used hybridized with fuzzy approach. The proposed approach is able to deal with the sensitiveness of cluster centers. Results presented are very encouraging.

Keywords FCM · Fuzzy clustering · Real-life data sets

1 Introduction

The objective of clustering is to group similar objects together. It uses some similarity measures like Euclidian distance measure, etc. for computing the closeness. Fuzzy C-means and Hard C-means are two popular clustering algorithms. The success of these algorithms is heavily dependent on the initial cluster centers. Different initial cluster centers provide different clustering results. The guarantee of the optimal clustering results cannot be ensured. This issue remains to be a big challenge in clustering research communities. But this type of algorithms does not achieve the optimal value rather it falls in local minimum [1]. The local search fails to go for a wider search since it always directs toward a single path and thereby reproducing the same cluster center. The idea is to run the algorithms with different cluster centers simultaneously to seek for better results.

P. Gopala Krishna (✉)
Department of IT, Griet, Hyderabad, India
e-mail: p_gopala_krishna@yahoo.co.in

D. Lalitha Bhaskari
Department of CS&SE, AUCE(a), Andhra University, Visakhapatnam, India
e-mail: lalithabhaskari@yahoo.co.in

In [2–8] many population-based techniques are proposed which have wide variety of applications. All these techniques are good in handling issue of local optima and at the same time able to search wider search space. The main plus point in these algorithms are their abilities to cope with local optima and explore the wider search space for better solution simultaneously [9]. Teaching–learning-based optimization (TLBO) [10] is another such algorithm recently developed. The concept is adapted from the learning experience of a class room. The influence of a teacher on the learner is the key concept of TLBO. This is also a population-based approach like any other nature-inspired algorithms. The learners are the population or in other words the candidate solutions. There are many applications of TLBO for various optimization problems [11–17]. The main focus point of TLBO is that it has no parameters to tune while simulating for the problem. Unlike many other population-based approaches the results of TLBO do not dependent on fine tuning of parameters.

In this paper, a variant of TLBO for addressing clustering problem using FCM algorithm has been introduced. The rest of the paper is divided into six sections. Section 2 gives an overview of Fuzzy C-means followed by a short description of TLBO in Sect. 3. Section 4 presents our FCM–TLBO approach and results are discussed in Sect. 5 and conclusion in Sect. 6.

2 Fuzzy C-Means Algorithm

In this approach all objects (n objects) having d dimensions are portioned into c number of fuzzy clusters where c is between 1 and n . Z is the set of C cluster centers. μ defines the fuzzy matrix with n rows and c columns. μ_{ij} is the membership function of the i th object with the j th cluster. The properties of μ are as follows:

$$\mu_{ij} \in [0, 1] \quad \forall i = 1, 2, \dots, n; \quad \forall j = 1, 2, \dots, c \quad (1)$$

$$\sum_{j=1}^c \mu_{ij} = 1 \quad \forall i = 1, 2, \dots, n \quad (2)$$

$$0 < \sum_{i=1}^n \mu_{ij} < n \quad \forall j = 1, 2, \dots, c \quad (3)$$

The aim is to minimize Eq. (4) as given below:

$$J_m = \sum_{j=1}^c \sum_{i=1}^n \mu_{ij}^m d_{ij} \quad (4)$$

where

$$d_{ij} = \|o_i - z_j\| \tag{5}$$

in which, m ($m > 1$) is a scalar exponent term which manages the fuzziness of the derived clusters and d_{ij} is the Euclidian distance from object o_i to the cluster center z_j . The z_j , centroid of the j th cluster, is obtained using Eq. (6).

$$z_j = \frac{\sum_{i=1}^n \mu_{ij}^m o_i}{\sum_{i=1}^n \mu_{ij}^m} \tag{6}$$

The FCM algorithm is given below:

Algorithm for Fuzzy C-means:

1. Choose m ($m > 1$); initialize μ_{ij} , $i = 1, 2, \dots, n; j = 1, 2, \dots, c$
2. Find the cluster centers $z_j, j = 1, 2, \dots, c$, according to Eq. (6).
3. Find Euclidian distance $d_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, c$.
4. Recalculate the membership function $\mu_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, c$ according to Eq. (7).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \tag{7}$$

5. If not converged, go to step 2.

The algorithm can be stooped based on several stooping criteria. It can be terminated if the center values do not significantly change or Eq. (4) cannot be minimized further. However, it is found that this algorithm falls in local optima due to sensitiveness on choice of centers.

3 Teaching–Learning-Based Optimization

TLBO is based on the phenomena influenced in a class room scenario. The impact of a teacher on a learner is the basis of the development of TLBO. The learners constitute the population and the communication, comparisons, etc. of the learners lead to explore and exploit the optimal solution of the problem. Different features of an optimization problem are analogous to different subjects taught in the class. The scores (marks) obtained in each such subjects or total marks obtained is considered as the “fitness”. The teacher is considered as the best solution since the teacher is termed as the best learned person in society. TLBO working is divided into two

phases, i.e., Teacher Phase and Learner Phase. Both parts are briefly described below. The entire process of TLBO is briefly presented below with pseudo algorithm.

A. Initialization

Following are the notations used for describing the TLBO:

- N number of learners in a class, i.e., “class size” which is to be defined by user depending on the problem
- D number of courses offered to the learner which is nothing but the dimensions/features in the data set under investigation
- MAXIT maximum number of iterations to be decided by the user experience for obtaining optimal solution

Equation (8) is the initialization of the candidate solutions:

$$x_{(i,j)}^0 = x_j^{\min} + \text{rand} \times (x_j^{\max} - x_j^{\min}) \tag{8}$$

where rand is the random number between 0 and 1, x_j^{\min} and x_j^{\max} represent the minimum and maximum values for j th parameter.

The parameters of i th learner for the generation g are given by

$$X_{(i)}^g = [x_{(i,1)}^g, x_{(i,2)}^g, x_{(i,3)}^g, \dots, x_{(i,j)}^g, \dots, x_{(i,D)}^g] \tag{9}$$

B. Teacher Phase

The mean parameter M^g of each subject of the learners in the class at generation g is given as

$$M^g = [m_1^g, m_2^g, \dots, m_j^g, \dots, m_D^g]$$

The best learner is termed as the teacher (X_{Teacher}^g) in TLBO. The best learner is computed based on the learner who achieves best objective function value. This phase of the algorithm is responsible to shift the mean learning experience of the class toward the learning experience of the teacher. Equation (10) is used to compute the new learning of a learner in a class.

$$X_{\text{new}(i)}^g = X_{(i)}^g + \text{rand} \times (X_{\text{Teacher}}^g - T_F M^g) \tag{10}$$

where T_F is a teaching factor which can be either 1 or 2.

If $X_{\text{new}(i)}^g$ is found to be a superior learner than $X_{(i)}^g$ in generation g , then it replaces inferior learner $X_{(i)}^g$ in the matrix.

C. *Learner Phase*

This phase of TLBO is analogous to the scenario of interactions among the learners in a class to improve individual learning experiences. A learner in a class may randomly interact with another learner and get influenced or influence other with knowledge. In Eq. (11) the interaction effect is defined.

For a given learner $X_{(i)}^g$, another learner $X_{(r)}^g$ is randomly selected ($i \neq r$). The i th parameter of the matrix X_{new} in the learner phase is given as

$$X_{new_{(i)}}^g = \begin{cases} X_{(i)}^g + \text{rand} \times (X_{(i)}^g - X_{(r)}^g) & \text{if } f(X_{(i)}^g) < f(X_{(r)}^g) \\ X_{(i)}^g + \text{rand} \times (X_{(r)}^g - X_{(i)}^g) & \text{otherwise} \end{cases} \quad (11)$$

D. *Algorithm Termination*

The algorithm is terminated after *MAXIT* iterations are completed.

Details of TLBO can be referred in [1].

4 Fuzzy Teaching–Learning-Based Optimization for Fuzzy Clustering

Here X is the position of the learner which shows the fuzzy relation from set of data objects, $O = \{o_1, o_2, o_3, \dots, o_n\}$, to set of cluster centers, $Z = \{z_1, z_2, z_3, \dots, z_c\}$.

$$X = \begin{bmatrix} \mu_{11} & \dots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \dots & \mu_{nc} \end{bmatrix} \quad (12)$$

In which μ_{ij} is the membership function as per the constraints stated in (1) and (2). We get Eq. (13)

$$X_{new} = X_i + \text{rand} \times (X_{teacher} - T_F M) \quad (13)$$

where M is the mean of parameter of each subject of the learners in the class in teacher phase and Eq. (14)

$$X_{new} = \begin{cases} X_i + \text{rand} \times (X_i - X_r) & \text{if } f(X_i) < f(X_r) \\ X_i + \text{rand} \times (X_r - X_i) & \text{otherwise} \end{cases} \quad (14)$$

in the learner phase for updating the positions of each learner i .

If the constraints in (1) and (2) are not satisfied then the position matrix is normalized. It can be done in making all the negative elements in matrix to become

zero and if all elements in the matrix are found to be zeros then we can reevaluate using a random value between [0, 1].

$$X_{\text{normal}} = \begin{bmatrix} \mu_{11}/\sum_{j=1}^c \mu_{1j} & \cdots & \mu_{1c}/\sum_{j=1}^c \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{n1}/\sum_{j=1}^c \mu_{nj} & \cdots & \mu_{nc}/\sum_{j=1}^c \mu_{nj} \end{bmatrix} \quad (15)$$

Fuzzy TLBO is similar to any evolutionary algorithm and hence an objective function or fitness function is defined. Eq. (16) is the fitness function for same.

$$f(X) = J_m \quad (16)$$

where J_m is the objective function Eq. (4) which is to be minimized.

Algorithm for Hybrid FCM–TLBO for fuzzy clustering

1. Set the parameters of fuzzy TLBO and FCM and population size N and m .
2. Create a group with N learners (X is $n \times c$ matrix).
3. Fuzzy TLBO algorithm:
 - 3.1 Find the cluster centers for each learner using Eq. (6).
 - Find the fitness value of each learner using Eq. (16)
 - Find out for X_{teacher} for the group of learners
 - Update the position of group of learner using Eq. (13)
 - Update the position of group of learner using Eq. (14)
 - Repeat the loop if fuzzy TLBO not terminating.
4. FCM algorithm
 - Find center using Eq. (6).
 - Find $d_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, c$; for each learner using Eq. (5).
 - Recalculate $\mu_{ij}, i = 1, 2, \dots, n; j = 1, 2, \dots, c$; for each learner using Eq. (7).
 - Evaluate fitness for each learner.
 - Evaluate best fitness for the swarm.
 - If terminating condition not achieved, go to step 5.
5. If FCM–TLBO terminating condition is not met, go to step 4.

5 Experimental Results

5.1 Parameter Settings

The TLBO algorithm is free from algorithmic specific parameter. So there is no need to adjust the parameters in order to optimize the performance of the FCM–TLBO algorithm. The terminating condition in algorithm for Fuzzy C-means is

Table 1 FCM and FCM–TLBO algorithms on six real data sets

Data set name	FCM			FCM–TLBO		
	Mean	Worst	Best	Mean	Worst	Best
Iris	69.3611	71.1867	65.4311	63.6909	63.9106	63.5042
Glass	71.9619	73.0170	70.7834	70.4460	70.5036	70.3511
Cancer	2255.0	2288.0	2238.0	2238.0	2238.0	2238.0
Wine	11975.0	12198.0	11533.0	10891	10908	10863
Vowel	70771.0	73648.0	68396.0	64927.0	65080.0	64784.0
Ecoli	16.2662	16.4455	16.1181	16.0775	16.0986	16.0609

attained when the generated solution is not improved further and that of TLBO is set to maximum 20 iterations with learners group size of 50. In all of algorithms the m , the weighting exponent, is set to 2.

5.2 Experimental Results

For evaluating the FCM and FCM–TLBO algorithms, six well-known real-world data sets have been considered: Fisher’s iris data set, Glass, Wisconsin breast cancer data set, Wine, Vowel data set, and Ecoli data set (Collected from UCI machine repository).

FCM was run for 50 independent runs and FCM–TLBO is for 40. The obtained results are summarized in Table 1. The mean, best, and worst values are reported for better insight. The values reported are the fitness values as per Eq. (4). It can be clearly verified from the table that FCM–TLBO is able to obtain superior results compared to FCM in all data sets.

6 Conclusion

This paper has suggested a new approach of hybridizing Fuzzy C-means with TLBO approach for clustering. The hybridization is found to be very useful in dealing with the initial cluster centers seeding issue. The introduction of an evolutionary technique known as TLBO is able to generate randomly many initial centers for clusters and in turn is able to obtain optimal clustering results due to its inherit nature of cooperation and communication among individuals. Six data sets from UCI machine repository are simulated and results are compared with FCM and the suggested FCM–TLBO approach. The results are seen to be very encouraging for FCM–TLBO.

References

1. Hathaway, R.J., Bezdek, J.C.: Local convergence of the fuzzy *c*-means algorithms. *Pattern Recogn.* **19**(6), 477–480 (1986)
2. Selim, S.Z., Alsultan, K.: A simulated annealing algorithm for the clustering problem. *Pattern Recogn.* **24**(10), 1003–1008 (1991)
3. Al-Sultan, K.S.: A tabu search approach to the clustering problem. *Pattern Recogn.* **28**(9), 1443–1451 (1995)
4. Hall, L.O., Ozyurt, I.B., Bezdek, J.C.: Clustering with a genetically optimized approach. *IEEE Trans. Evol. Comput.* **3**(2), 103–112 (1999)
5. Bandyopadhyay, S., Maulik, U.: An evolutionary technique based on *k*-means algorithm for optimal clustering in *m*. *Inf. Sci.* **146**(1–4), 221–237 (2002)
6. Lili, L., Xiyu, L., Mingming, X.: A novel fuzzy clustering based on particle swarm optimization. In: *First IEEE International Symposium on Information Technologies and Applications in Education, ISITAE*, pp. 88–90 (2007)
7. Kanade, P.M., Hall, L.O.: Fuzzy ants and clustering. *IEEE Trans. Syst. Manage. Cybern. Part A* **37**(5), 758–769 (2007)
8. Maulik, U., Saha, I.: Modified differential evolution based fuzzy clustering for pixel classification in remote sensing imagery. *Pattern Recogn.* **42**(9), 2135–2149 (2009)
9. Paterlini, S., Krink, T.: Differential evolution and particle swarm optimisation in partitioned clustering. *Comput. Stat. Data Anal.* **50**(5), 1220–1247 (2006)
10. 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**, 303–315 (2011)
11. Rao, R.V., Savsani, V.J.: *Mechanical design optimization using advanced optimization techniques*. Springer, London (2012)
12. Satapathy, S.C., Naik, A.: Data clustering based on teaching-learning-based optimization. In: *Swarm, Evolutionary, and Memetic Computing. Lecture Notes in Computer Science*, vol. 7077, pp. 148–156. Springer, Berlin (2011)
13. Satapathy, S.C., Naik, A., Parvathi, K.: High dimensional real parameter optimization with teaching learning based optimization. *Int. J. Ind. Eng. Comput.* (2012). doi:[10.5267/j.ijiec.2012.06.001](https://doi.org/10.5267/j.ijiec.2012.06.001)
14. Satapathy, S.C., Naik, A., Parvathi, K.: Teaching learning based optimization for neural networks learning enhancement. In: *LNCS*, vol. 7677, pp. 761–769. Springer, Berlin (2012)
15. Satapathy, S.C., Naik, A., Parvathi, K.: 0–1 Integer Programming For Generation maintenance Scheduling in Power Systems based on Teaching Learning Based Optimization (TLBO), *CCIS* 306, pp. 53–63. Springer, Berlin (2012)
16. Satapathy, S.C., Naik, A., Parvathi, K.: Improvement of initial cluster center of *c*-means using Teaching learning based optimization. *Elsevier, Procedia Technology* **6**(2012), 428–435 (2012)
17. Naik, A., Parvathi, K., Satapathy, S.C., Nayak, R., Panda, B.S.: QoS Multicast Routing Using Teaching Learning Based Optimization, pp. 49–55. Springer, Berlin (2012)

A Mean-Based Thresholding Approach for Broken Character Segmentation from Printed Gujarati Documents

Riddhi J. Shah and Tushar V. Ratanpara

Abstract The major problem faced by a Gujarati optical character recognition (OCR) can be attributed to the presence of broken character in machine printed Gujarati document image. This character could cause the error in character segmentation process. Broken characters are generated due to noise scanning, older documents with low-quality printing, and thresholding error. It is necessary to identify and segment it properly. So this paper presents mean-based thresholding technique for broken character segmentation from printed Gujarati documents. Line segmentation is used to extract lines from Gujarati document image. Individual characters are extracted using vertical projection profile method. Then, broken characters are identified using mean-based thresholding (MBT) algorithm. Heuristic information is used to merge the identified broken characters. The main purpose of this paper is to merge vertical and naturally broken Gujarati characters as a single glyph from the document image. Experimental results are carried out using various types of Gujarati documents (A, B, C, and D). 79.93 % accuracy is achieved from experimental results.

1 Introduction

Optical character recognition (OCR) for Gujarati script is an important research technology which is used to read scanned document images and convert them into editable text [1, 2]. Gujarati script has large character set with various vowels,

R.J. Shah (✉) · T.V. Ratanpara
Department of Computer Engineering, Dharmsinh Desai University,
Nadiad, Gujarat, India
e-mail: riddhijshah91@gmail.com

T.V. Ratanpara
e-mail: tushar.ratanpara@gmail.com

consonants, and conjuncts. Gujarati OCR faces many challenges that reduce its efficiency. In some kind of books, there are characters having irregular thickness. Hence, problem occurs due to broken and joint characters. One of the most important problems that decrease the accuracy of Gujarati character segmentation method is broken character [3] in printed document image. Humans can read broken characters easily but it is far complicated for OCR system to identify them. By definition, single character breaks apart is called broken character or character that composed several connected components [4]. Broken character poses many problems in Gujarati OCR system. Character segmentation phase of Gujarati OCR will not be able to identify the broken character as a single glyph.

In this paper, a mean-based thresholding (MBT) method is presented for broken character segmentation as a single glyph using heuristic information from Gujarati printed documents. Typically, broken characters in Gujarati sentence can be portioned into two groups: horizontal broken characters and vertically broken characters. Vertical broken characters are those characters which segment more than two parts using character segmentation method. There are four types of document images for Gujarati OCR which are discussed in [5].

2 Related Work

Otsu's histogram-based global thresholding approach is good for image binarization [6, 7]. Line segmentation is performed using HPP [8] and character segmentation using VPP [9, 10]. Peerawit et al. [4] categorized a method for broken character segmentation in Thai printed document. The main advantage of the proposed method is the ability to segment broken characters. It divides broken characters into a large number of connected components. Character segmentation is handled for both complete and broken characters and segmentation accuracy is 91.09 %. Droettboom [11] discussed a technique for broken character which is based on graph combinatory. It presents that connected component is a set of black pixels and character is broken when it is made up of too many connected components (CCs). Broken characters are merged by using distance of the CCs. K-NN classifier is used for it. BCC correctly finds 71 % broken characters. Sumetphong [12] presented a novel technique for recognizing broken Thai characters using set partitioning problem (SPP) and heuristic integer iterative programming (HIIP) [13]. 93 % recognition accuracy is achieved. S. Tangwongsan proposed heuristic enumerative method to find broken pieces in Thai and English documents [14, 15]. Kumar et al. [16] implemented a novel approach to recognize English capital and small letter broken character based on gradient pattern and its spatial relationship. Sigelle and Sulem [17] have recognized American characters from printed old books using dynamic bayesian network (DBN). Adaptive technique is proposed by P. Stubberud and J. Kannani which restores touching and broken characters [18].

3 Proposed Approach

An abstract model of proposed system is shown in Fig. 1. Printed Gujarati document image is taken as an input image. Binarization method is applied on input image which is described in Sect. 3.1.

3.1 Binarization Using Otsu’s Method

Otsu’s method [6, 19] is used for image binarization to convert the grayscale image into binary image. The grayscale image of such documents generally exhibits a large amount of broken characters. The need of binarization is to increase the level of brokenness in the binary image. Figure 2 shows the binarized image of Gujarati document type D.

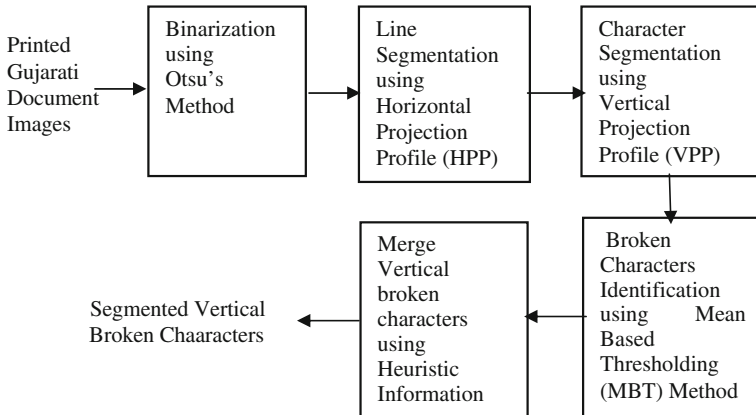


Fig. 1 Abstract model of proposed system

Fig. 2 Binarized image

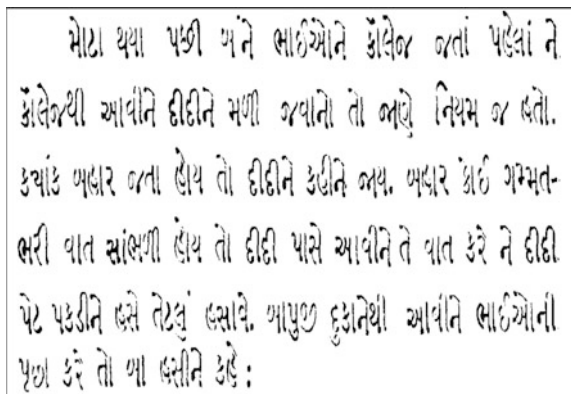
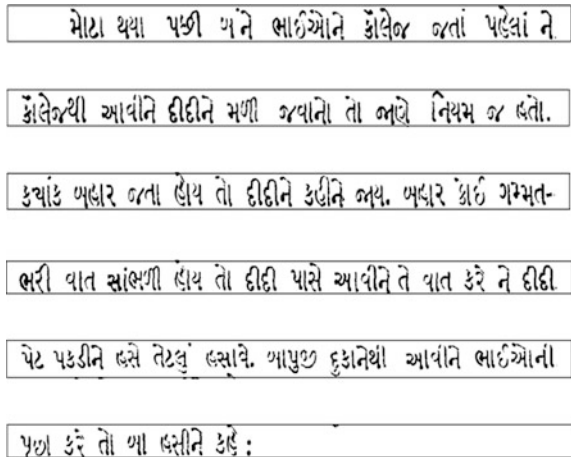


Fig. 3 Line segmentation



3.2 Line Segmentation Using Horizontal Projection Profile (HPP)

Line segmentation is used to extract text lines from a binarized input document image. Horizontal projection profile (HPP) method [8, 10] is used for line segmentation. Histogram is plotted based on the intensity of each pixel present in horizontal direction. The lines and gaps are identified separately based on threshold value. Figure 3 shows output of line segmentation.

3.3 Character Segmentation Using Vertical Projection Profile (VPP)

Character segmentation is used to segment each individual character from the extracted lines using vertical projection profile (VPP) method [10, 20]. A histogram is plotted based on the intensity of character pixel present in vertical direction. The characters and intermediate gaps are identified separately based on threshold value. One single line is taken for character segmentation. There are 10 total characters in a line as shown in Fig. 4. VPP method segments only five full characters. So it is

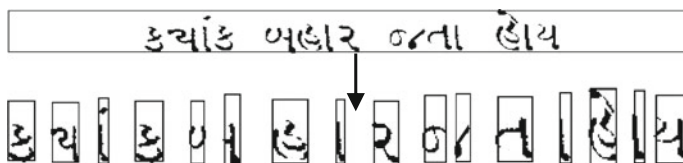
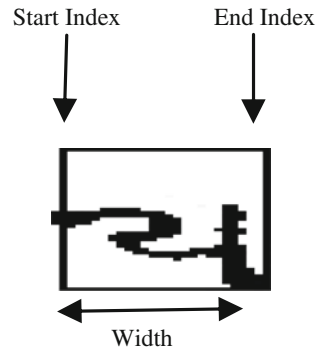


Fig. 4 Character segmentation

Fig. 5 Width of normal character



necessary to identify the all characters as a single glyph. Section 3.4 describes a method to identify the broken characters from the segmented characters using VPP.

3.4 Broken Characters Identification Using Mean-Based Thresholding Method

Main task is to identify the broken characters. Characters are classified into two categories: Full characters and broken characters. Mean-Based Thresholding (MBT) method is used to identify vertical broken characters from segmented characters in printed Gujarati documents. Characters are categorized by character’s width (size). Compute the width of each segmented character in a line. Width of character is found using Eq. (1). Normal character width is shown in Fig. 5. Figure 6 shows width of broken character.

$$\text{Width} = \text{Endindex of Char} - \text{Startindex of Char} \tag{1}$$

Figure 7 indicates that the width of broken characters is less than the width of normal characters. Figure 7 shows the width of all characters in a line.

The mean value of characters width is computed to estimate the normal character width from a line. The mean value is considered as a normal character. Mean value is 16 which is computed from the extracted characters in Fig. 7. Consider the mean value as a threshold value (Th). Broken characters are those whose width is $X\%$ smaller than threshold value which is calculated using Eq. (2).

$$\text{Character} = \begin{cases} \text{Full} & \text{width} \geq 95\% \text{ of mean value} \\ \text{Broken or half} & \text{otherwise} \end{cases} \tag{2}$$

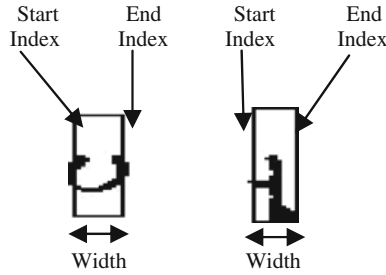


Fig. 6 Width of broken characters

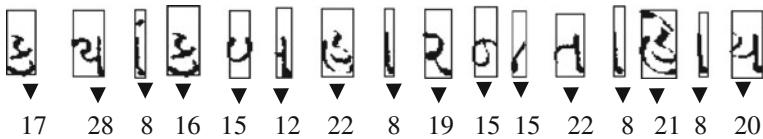


Fig. 7 Width of all characters in a line

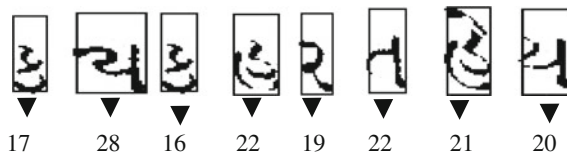


Fig. 8 Identified full characters

The MBT method identified eight full characters in a line from the segmented characters as shown in Fig. 8.

Figure 9 gives the identified eight vertical broken and half characters from segmented characters in a line using MBT method.

After identification, the vertical broken/half characters are required to merge as a single glyph which increases the accuracy of Gujarati OCR system. Modifier (kano) is considered as a half character due to its width. So it is also required to merge modifier with a character.

3.5 Merge Vertical Broken Characters Using Heuristic Information

In this section, vertical broken characters are merged using heuristic information to segment a broken character as a single glyph. The space between two characters is computed to merge two consecutive broken characters. Therefore, space is

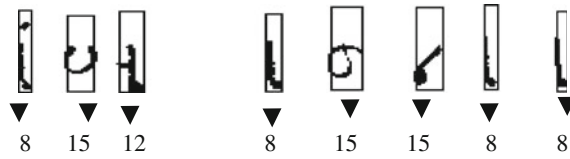


Fig. 9 Identified vertical broken/half characters

considered as a heuristic information. Space between two characters is computed using Eq. (3).

$$\text{Space} = \text{NS} - \text{CE} \tag{3}$$

where, NS is next character’s start index value and CE is previous character’s end index value. Left space of a broken character is found using Eq. 4.

$$\text{Left Space} = \text{BS} - \text{PE} \tag{4}$$

where, BS is broken character’s start index value and PE is previous character’s end index value. Right space of a broken character is found using Eq. 5.

$$\text{Right Space} = \text{BNS} - \text{BE} \tag{5}$$

where, NS is broken character start index value and BE is broken character end index value.

Compute the space between all characters from segmented characters in a line using Eq. 3 which is shown in Fig. 10.

If there are vertical broken characters in a line, the space between two consecutive broken or half characters must be smaller than the space between normal characters as shown in Fig. 10. Compute the minimum space and maximum space between all characters in a line. Aspect ratio is calculated for minimum space and maximum space using Eq. 6.

$$\text{Aspect Ratio(AR)} = \frac{\text{Minimum space}}{\text{Maximum space}} \% \tag{6}$$

Figure 11 represents segmented characters using heuristic information. Vertical broken characters are merged using following equations. Equation 7 indicates that broken or half character merged with previous character when space left of broken character is smaller than the aspect ratio.

$$\text{Left Space} < = \text{Aspect ratio value} \tag{7}$$

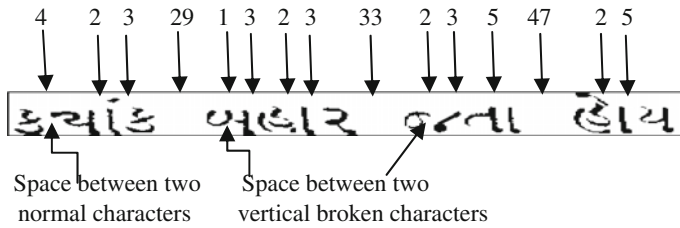


Fig. 10 Space between characters using heuristic information

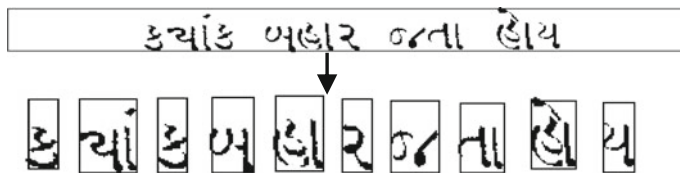


Fig. 11 Vertical broken character segmentation using heuristic information

Broken or half character merged with next character whenever the space right of broken character is smaller than the aspect ratio which is shown in Eq. 8.

$$\text{Right Space} < = \text{Aspect ratio value} \tag{8}$$

4 Experiment Result

Proposed approach has been tested on 20,000 characters from different types of Gujarati document images scanned at 300 dpi. Document image type C and D contain more broken characters so our main emphasis is on these two types of document images for broken character segmentation. Table 1 shows the result of total identified and merged broken characters.

MBT method is based on mean. Other statistical parameters are also used to merge the broken characters. Figure 12 shows the result of identification and segmentation of vertical broken characters using different statistical parameters such like median, standard deviation, and variance. Conventional approach of character segmentation using VPP method obtained 74.89 % accuracy. An experimental result indicates that accuracy of our approach for broken character segmentation is 79.93 %.

Table 1 Total identified and merged broken characters

Doc Type	Total characters	Vertical broken character	Identified characters using MBT	Merged characters using heuristic information	Total segmented characters using heuristic information	Precision	Recall
A Doc 1	151	12	12	12	149	0.98	0.92
B Doc 2	319	18	16	16	317	0.99	0.94
C Doc 3	50	14	14	14	43	0.86	0.75
C Doc 4	165	26	26	26	138	0.84	0.84
C Doc 5	1264	196	168	168	864	0.68	0.82
C Doc 6	1319	372	306	306	981	0.74	0.73
C Doc 7	591	66	66	66	475	0.80	0.88
D Doc 8	180	24	24	24	150	0.83	0.86
D Doc 9	588	146	140	140	504	0.86	0.77
D Doc 10	534	99	93	93	426	0.80	0.81
D Doc 11	633	160	148	148	527	0.83	0.77
D Doc 12	593	93	93	93	498	0.84	0.84
D Doc 13	592	141	129	129	506	0.85	0.78
Average accuracy				90.34 %	79.93 %		

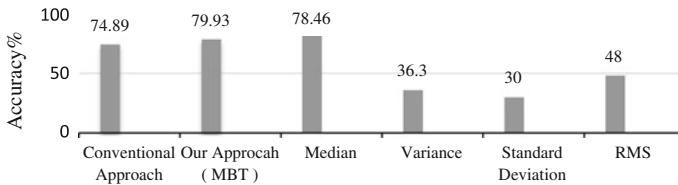


Fig. 12 Broken character segmentation using various statistical parameters

5 Conclusion

Proposed approach is tested on various types of classes (Types A, B, C, and D) of document images. In this proposed approach, a method is implemented which helps to increase the accuracy of character segmentation method even the image is noisy. The advantage of the proposed method is the ability to classify the broken characters and full characters from the Gujarati document images. Accuracy of the proposed approach is better than the conventional approach which is shown in the experimental results. Although the proposed method is effective for broken character segmentation, it will not work well when input document image contains graphics. Input image must be skew free.

References

1. Singh, R., Yadav, C.S., Verma, P., Yadav, V.: Optical character recognition (OCR) for printed Devnagari script using artificial neural network. *Int. J. Comput. Appl.* **1**, 91–95 (2010)
2. Antani, S., Agnihotri, L.: Gujarati character recognition. In: *International Conference on Document Analysis and Recognition* (1999)
3. Solanki, P., Bhatt, M.: Printed Gujarati script OCR using hopfield neural network. *Int. J. Comput. Appl.* **69**, 33–37 (2013)
4. Peerawit, P., Yingsaree, W., Kawtrakul, A.: The utilization of closing algorithm and heuristic information for broken character segmentation. *IEEE Conf. Cybern. Intell. Syst.* **2**, 775–779 (2004)
5. Prajapati, H., Rama Mohan, S.: Removal of graphics from text-document and segmentation of Gujarati documents using connected component theory. *IOSR J. Eng.* **2**(5), 1029–1031 (2012)
6. Sankura, B., Sezgin, M.: *Thresholding Review*. Boğaziçi University Electric-Electronic Engineering Department, Bebek
7. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Pearson Education, Singapore
8. Patel, C., Desai, A.: Segmentation of text lines into words for Gujarati handwritten text. In: *International Conference on Signal and Image Processing*, pp. 130–134 (2010)
9. Bishnu, A., Chaudhari, B.B., Segmentation of Bangla handwritten text into characters by recursive contour following. In: *International Conference on Documents Analysis and Recognition* (1999)
10. Thakkar, N., Dangarwala, K.: A survey on offline-methods of character segmentation. In: *International Conference on Advances in Computer Science and Electronics Engineering* (2013)
11. Droetboom, M.: Correcting broken character in the recognition of the historical printed documents. *IEEE Joint Conf. Digit. Libr.* **2**, 775–779 (2003)
12. Tangwongsan, S., Sumetphong, C.: Optical character recognition techniques for restoration Thai historical documents. In: *International Conference on Computer and Electrical Engineering*, pp. 531–535 (2008)
13. Sumetphong, C., Tangwongsan, S.: An optimal approach towards recognizing broken Thai characters in OCR systems. In: *International Conference on Digital Image Computing Techniques and Applications*, pp. 1–5 (2012)
14. Sumetphong, C., Tangwongsan, S.: Effectively recognizing broken characters in Historical documents. *IEEE Int. Conf. Comput. Sci. Autom. Eng.* **3**, 104–108 (2012)
15. Sumetphong, C., Tangwongsan, S.: Recognizing broken characters in Thai historical documents. *Int. Conf. Adv. Comput. Theory Eng.* **1**, 99–103 (2010)
16. Babu, R., Ravinsankar, M., Kumar, M., Raj, A., Wadera, K.: Recognition of machine printed broken characters based on gradient patterns and its spatial relationship. *Int. Conf. Comput. Sci. Inf. Technol.* **1**, 673–676 (2010)
17. Sulem, L., Sigelle, M.: Recognition of broken characters from historical printed books using dynamic Bayesian networks. In: *International Conference on Document Analysis and Recognition*, vol. 1, pp. 173–177 (2007)
18. Stuberud, P., Kanai, J., Kalluri, V.: Adaptive image restoration of text images that contain touching or broken characters. In: *International Conference on Document Analysis and Recognition*, vol. 2, pp.778–781 (1995)
19. Dholakiya, J.: *Mathematical Techniques for Gujarati Document Analysis and Character Recognition*. M. S. University of Baroda, Baroda (2010)
20. Singh, S.: Optical character recognition techniques: a survey. *J. Emerg. Trends Comput. Info. Sci.* **6**, 545–550 (2013)

Entity Resolution-Based Jaccard Similarity Coefficient for Heterogeneous Distributed Databases

Ramesh Dharavath and Abhishek Kumar Singh

Abstract Entity Resolution (ER) is a task for identifying same real world entity. It refers to data object matching or deduplication. It has been a leading research in the field of structure database. Due to its significance, entity resolution continues to be a most important challenge for heterogeneous distributed databases. Several methods have been proposed for the Entity resolution, but they have yielded unsatisfactory results. In this paper, we propose an efficient integrated solution to the entity resolution problem based on Jaccard similarity coefficient. Here we use Markov logic and Jaccard similarity coefficient for providing an efficient solution towards ER problem in heterogeneous distributed databases. The approach that we have implemented gives an overall success rate of about 98 %, thus proving better than the previously implemented algorithms.

Keywords Entity resolution (ER) · Distributed database · Jaccard similarity coefficient · Markov logic

1 Introduction

Entity resolution (ER) is also known as deduplication (matching) that refers to the task of disambiguating manifestations of real world entities in various records or mentions by linking and grouping. It is a crucial step in identifying and

R. Dharavath (✉) · A.K. Singh
Department of Computer Science and Engineering, Indian School of Mines,
Dhanbad 826004, India
e-mail: ramesh.d.in@ieee.org

A.K. Singh
e-mail: abhishek@ismdhanbad.ac.in

linking/grouping different manifestations of the same real world object [1]. This leads inaccuracies and data redundancy in query processing and information pulling out [2]. We reduce these problems with the help of entity resolution. ER enhances by discovering the entities and maps each database to the same entity. ER is necessary while joining the data by suitable entities that share a common name (e.g., URL, database key, nationwide identification number) as may be the case due to differences in record size and storage.

Large clients-based companies (face book, tweeter, amazon, government databases) frequently merge extensive lists of client names and addresses with partially overlapping sets of clients. Another methodology where deduplication [3] is necessary to construct web portal-related activities and integrate data from suitable pages which were created in a distributed scenario. We have analyzed numerous research methodological instincts according to their evaluation of entity resolution approaches. Designing a system that can find when two tuples refer to same entity [4] is the main challenge. In inconsistencies and spite of errors in the data, we consider as our motivating example as the problem of resolving the authors of the publications. By considering the following set of sequence represents some paper references and each oval represents an author reference. We consider the problem of resolving the author publications.

- (1) D. Ram, AK. Singh, C. Palok, “*Enhanced Abstract Data Types in OR Databases*”
- (2) D. Ram, C. Palok, “*Priority Assignment in Real-Time Active Databases*”
- (3) K. Di, AK. Singh, D. Ram, “*Lineage Tracing for General Data Warehouse*”
- (4) D.D. Ram, C. Palok, “*Free Parallel Data Mining*”

For Example, there are three authors in paper 3 “D Ram,” “AK Singh,” and “K. Di.” Overall, there are ten authors that map to the three author entities: “A Kumar Singh,” “D Ramesh HN,” and “Kamo Dibro.” In the following example, all three writers “A Kumar Singh,” “A K Singh,” and “A Kumar S” refer to the similar entity entitled “A K Singh.” If we look at pairs of references individually and try to decide if they refer to same entity that may not be a difficult task for uncommon names like “A K Singh” and “A Kumar Singh.” But for frequently occurring names like “A K Singh” and “Abhishek K Singh,” it is a problem. While they will be duplicates in some cases, in others they will be distinct. We can however make use of additional evidence if we make these decisions collectively. Here, we also discuss about another example related to duplicate and non-duplicate data [5, 6]. In this example, we show why entity resolution is important and how it is a tough task to remove duplicate data in the heterogeneous database. In this example, duplicate citations of a book have been considered [7].

- J. Breiman, L. Friedman, and P. Stone, (1984). *Classification and Regression*. Wadsworth, Belmont, CA.
- John Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Classification and Regression Trees*. Wadsworth and Brooks/Cole, 1984.

2 Related Work

Many different approaches to the ER problem have been proposed [1, 2, 8–19], therefore comparison is needed with different approaches in order to assess the overall suitability of the algorithm. We recommend a strong need for suitable judgments on different schemes. Until now most ER solutions have been evaluated alone using different methodologies [7], configurations [20], and test problems making it difficult to assess the overall quality of each approach. Only few attempts have been made to evaluate different similarity instances [21] and of blocking approaches [14]. Several proposals have been made regarding entity resolution [15, 16] but they are yet to be implemented and applied in real world. A major difficulty that faces us during ER scenarios is that they require different parameters to be set such as the similarity functions for comparing attribute values or similarity thresholds to be exceeded by matching entities [19]. Some approaches utilize algorithms which require some specific parameters related to training data. Usually, the proposed algorithm configures some factors and provides the result for matching [5, 22]. We propose a mathematical formulation of the ER that incorporates non-independent, identically distributed methods. It utilizes Jaccard similarity coefficient in multi-relation dataset in heterogeneous databases. In proposed Markov logic [17, 18], it merges predicate calculus and Jaccard similarity coefficient with threshold values and uses an algorithm for new learning's [12]. The proposed model provides a practical method to integrate many validated factors into an efficient solution for ER system [3, 23, 24].

The remaining part is arranged as the following: in Sect. 3, we briefly review the necessary background on mathematical models like Markov logic, predicate calculus, and Jaccard similarity. Entity resolution solutions with these mathematical models are presented in Sect. 4. We briefly discuss our dataset and performance evaluation and result in Sect. 5. Finally, Sect. 6 concludes our performance and experiments.

3 Projected Terminologies

3.1 Predicate Calculus

A predicate calculus knowledge base (KB) is a set of sentences or formulas in predicate calculus (first-order logic). First-order logic (FOL) models the world in terms of four types of symbols: Predicates, Constants, Functions, and variables. A term is a kind of expression which represents a related object in the suitable domain. It may be a variable, a function /constant applied to a particular set of tuples. For example, $\langle \text{constant} \rangle$; Greatest Common Divisor(a, b) and $\langle \text{name} \rangle$ are terms. If statement is true it returns true otherwise false. We replace all variables by constants. Here, as an example, we apply predicate calculus in Table 1. Let a

Table 1 Paper (title, author, year)

Title	Author	Year
Entity resolution using active learning	Sunita sarawal and Arundha	2007
Entity resolution for uncertain data	N.aayt, r.k barinia	2011

predicate $\text{Hasword}(X, Y)$ defines “*Hasword is true if Y present in X .*” If word = entity, $\text{Hasword}(\langle \text{field1} \rangle, \langle \text{word} \rangle) = \text{true}$. $\text{Hasword}(\langle \text{field2} \rangle, \langle \text{word} \rangle) = \text{false}$.

3.2 Markov Logic

A Knowledge-Base (first-order) is a set of some constraints on a particular set of group of strings. If a particular string of group is violated then it leads to null representation. The basic principle used in Markov model is to reframe these factors [17]. When a string is violated then it may be computed by applying Markov principles.

3.3 Jaccard Similarity Coefficients

The Jaccard similarity coefficient is also known as Jaccard index. It is a kind of functionality used for comparing the similarity of sets. The Jaccard coefficient measures similarity between finite sample sets and is defined as “*size of the intersection divided by the size of the union of the sample sets*”:

Let X and Y are two finite sets

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad (3.1)$$

$J(X, X) = 1$ and $J(X, Y) = 0$ if $X \cap Y = 0$. Equation for Jaccard coefficient is also represented as

$$T = \frac{N_z}{N_x + N_y - N_z}, \quad (3.2)$$

where N_x —number of elements in set X , N_y —number of elements in set Y , and N_z —number of elements in intersecting set. For example, to calculate the similarity between night and nacht, we would find the set of bigrams in each word: {ni, ig, gh, ht} and {na, ac, ch, ht}. Each set has four elements, and the intersection of these two sets has only one element: ht. Inserting these numbers into the formula, we calculate $s = 1/4 = 0.25$.

4 Entity Resolution

4.1 Equality in Markov Logic

Most of the systems for inference, in first-order logic, make assumption as unique names. Different constants refer to different objects in the domain. This assumption can be removed by introducing the equality predicate (equals (p, q) or $p = q$ for short) and its axioms [13].

- $\forall p1, p2, q1, q2 \ p1 = p2 \wedge q1 = q2 \Rightarrow (R(p1, q1) \Leftrightarrow R(p2, q2))$. Same type of forms are required for to form similar type of entries.

Reverse predicate equivalence:

- $R: \forall p1, p2, q1, q2 \ R(p1, q1) \wedge R(p2, q2) \Rightarrow (p1 = p2 \Leftrightarrow q1 = q2)$, for same type of functions which are presented in the string of worlds. It can be verified by the below findings:
- $\forall p1, p2, q1, q2 \ R(p1, q1) \wedge R(p2, q2) \wedge p1 = p2 \Rightarrow q1 = q2$.
- $\forall p1, p2, q1, q2 \ R(p1, q1) \wedge R(p2, q2) \wedge q1 = q2 \Rightarrow p1 = p2$.

As we have used first-order logic for our statements, these clauses were found as false. The reason is that the different related strings of the same predicate cannot represent same type of tuples. But in markov logic these clauses are correct.

4.2 Problem Formulation

In this section, we discuss our proposed methodology in detail. We use deduplicated database that represents binary relation. For example, if a database contains some related tuples such as, Conference (<paper title>, <author name>, and <place>), it can be replaced by the following schemas, HasVenue (<conference>, <place>), HasAuthor (<conference>, <author name>) and HasTitle (<conference>, <paper title>). The presented entity (e.g. <place>, <author name>, <paper title>) can be identified represented by the strings which can be appeared as arguments forms of a specified schema. As an example, various schema models can be represented as IC3T-2015, Second IC3T, and springer IC3T'15. We assume that predicates which represent relations in real world are *typed*. Suppose that, the predicate is HasTitle (paper, title). First argument in HasTitle predicate is of type paper and second is of type title. Our task is to make sure that, each pair of constants of the same type($p1, p2$), represent the same real world entity in both cases.

We proposed the model for ER problem which is in the form of markov logic and Jaccard similarity coefficient $\frac{|\text{Field} \cap \text{Word}|}{|\text{Field} \cup \text{Word}|}$. We use hand-written code to find the similarity for each predicate based on Jaccard similarity coefficient. And compare the weight with our threshold value. If the presented value is more or less greater to threshold value than string is matched otherwise not-matched. We apply this result in markov equality.

4.3 A Comparison Scenario

We visualize database fields as a string which is the combination of words. Our task is to deduplicate the database fields. The predicates are defined as follows:

- $HasWord(field, word)$
- $HasField(record, field)$
- $SameField(field, field)$
- $SameRecord(record, record)$

Rule 1: $HasWord(f_1, w) \wedge HasWord(f_2, w) \Rightarrow SameField(f_1, f_2)$

Alternatively, fields having a common word are more or less same. If we apply this field and word in Jaccard similarity coefficient formula as $J(field, word) = \frac{|Field \cap Word|}{|Field \cup Word|}$ and if Jaccard similarity coefficient function more or less equal to the threshold function. If same word is present in the field₁ and field₂ then both fields are same. Let $w = \text{"PetaBytes,"}$ we apply our mathematical model for finding similarity with this string. In Table 2, "PetaBytes" is presented in row no.3. So Jaccard similarity returns fraction value which is nearer to 1(true). If out threshold value is 65 %, then there is a match, and success rate is up to 98–100 %. In Table 3, there is no such row which contains the string "PetaBytes." So, similarity function returns false.

Now, with the help of markov equality clause, we define the following rules, which are helpful for finding same records in two different databases.

Table 2 DBLP Database with paper (Title, Author, Venue, Year)

Row No.	Title	Author	Venue	Year
1	SPRINT: A scalable parallel classifier for data mining	John C. Shafer, Rakesh Agrawal, Manish Mehta	Very large data bases	1996
2	Database patchwork on the internet	Reinhard Braumandl, Alfons Kemper, Donald Kossmann	Int. Conference on management of data	1999
3	Petabyte databases	Dirk Düllmann	Very large data bases	1999
4	Relational databases for querying XML documents: Limitations and opportunities	Jayavel Shanmuga sundaram, Kristin Tufte, Chun Zhang, Gang He, David J. DeWitt, Jeffrey	Very large data bases	1999

Table 3 ACM database with paper (Title, Author, Venue, Year)

Row No.	Title	Author	Venue	Year
1	SPRINT: A Scalable Parallel Classifier for Data Mining	John C. Shafer, Rakesh Agrawal, Manish Mehta	VLDB	1996
2	Database Patchwork on the Internet	Reinhard Braumandl, Alfons Kemper, Donald Kossmann	SIGMOD Conference	1999
3	Relational Databases for Querying XML Documents: Limitations and Opportunities	Jeffrey F. Naughton, Kristin Tufte, Chun Zhang, Gang He, David J. DeWitt, Jayavel Shanmugasundaram	VLDB	1999

Rule 2: $HasWord(f_1, w) \wedge HasWord(f_2, w) \wedge HasField(r_1, f_1) \wedge HasField(r_2, f_2) \Rightarrow SameRecord(r_1, r_2)$.

Here, we use markov equality for defining same record predicate. In other Word, if all fields match then record1 it is equal to record2. According to the above discussion, we can implement two more rules which help us for removing duplicate results.

Rule 3: $HasField(r_1, f_1) \wedge HasField(r_2, f_2) \wedge SameField(f_1, f_2) \iff SameRecord(r_1, r_2)$ {Implement collective classification}.

Rule 4: $SameRecord(r_1, r_2) \wedge SameRecord(r_2, r_3) \Rightarrow SameRecord(r_1, r_3)$ {Implement transitivity}

5 Simulation Results

In this section, the efficiency of the proposed approach has been presented. In order to opt better results synthetic and real datasets have been considered.

5.1 Repository Datasets

We have utilized three freely available databases in our experiments: Cora, Abt-Buy, and Amazon-Google Products. Table 4 depicts about datasets used in experiments, number of attributes, and number of tuples. The Cora dataset includes bibliographical information about scientific papers in 12 different attributes from which we only use author, title, and venue attributes. The aforementioned datasets have been represented with 1295 related tuples from to computer science research papers.

Table 4 Dataset information

Name	Attributes	No. of tuples
Cora	Title, Author, Venue, Year	1295
Amazon, Google Products	Manufacture, price, product name, description	1300
Abt-Buy	Manufacturer, price, product name, description	1097

Abt—Buy Dataset contains 1097 matching product titles. This dataset comprises of electronic products and usually has unique identifiers in each title. Amazon—Google Dataset has 1300 matching product titles and unlike the Abt—Buy dataset, the titles do not always have discriminatory tokens.

5.2 Models

The following models have been compared and generated suitable experimental results.

- a. *AB*: It represents a simple Bayes model.
- b. *ABC(X)*: This model is near to EW approach described above. This model has four distinct rules and connects each word to the related string. The model contains a clause and *SameEntity(p1, p2)*. The algorithm given by Singla and Domingos [12] is utilized to learn the weights of rules.
- c. *ABC(X + Y+ Z)*: This model comprises of both reverse predicate equivalence and transitive closure rules and thus it includes the enhancements proposed by both McCallum and Wellner [25] and Singla and Domingos [12].
- d. *ABC(X + Y+ Z + M)*: This model is useful for datasets, which contain author attributes. It also employs statistical formulations according to the strings which are formulated in a duplicate manner.

5.3 Result Performance

In our experiments, we use above described dataset and data model. We apply efficient threshold value for evaluating the result. We use java package for developing mathematical model. In our experiment, when similarity match coefficient is greater than threshold value then we declare it similar. With the help of Jaccard similarity we find the match coefficient. With the help of our problem formulation, first, we compare field to filed as per Rule 1. When we compare every field's string with the help of Jaccard similarity coefficient, our function returns some similarity probability value ($0 < p < 1$). When we compare it with threshold, if value is greater than threshold value then match. Now we apply our second formula and continue: if

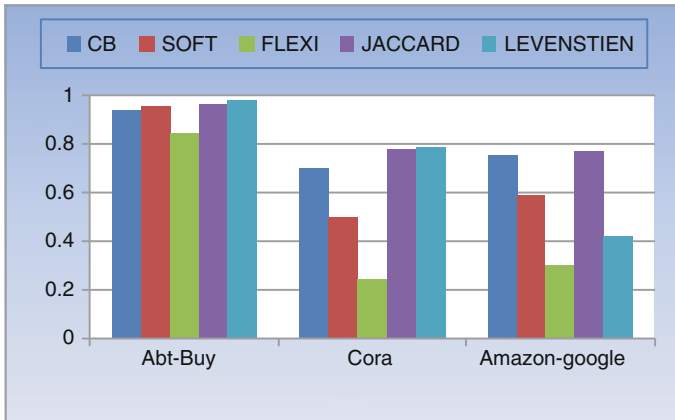


Fig. 1 Success rate over different datasets

all fields satisfy threshold value then we declare tuple is match. As we go on increasing collective inference features to the model, we observe a drastic growth in all test cases. Our experiment shows that the best performing model is $ABC(X + Y + Z)$. Thus, the performance is enhanced upon adding more features to the model.

The proposed approach has been compared with the strategy of edit-distance, SoftTFDF [26] and a SofttfIDF, which is represented as FIDF. We have used second string package for generating this functionality. The SoftFIDF and softFIDF functions use a methodology for finding similarity strings. In Fig. 1, we have shown the success rate of CB, Jaccard, soft TFIDF, and edit-distance function over the cora, abt-Buy, amazon google products. We observe performance a little less than the best similarity function and never is the worst similarity function. In our experiment, we successfully provide a efficient solution for entity resolution. Response time of our function is similar to Soft function.

6 Conclusion

We have presented a compact framework for entity resolution strategy. We demonstrate how a few rules in Markov logic and Jaccard similarity give more efficient results, when compared with other alternative approaches for this problem. Here we have captured the essential features of many different approaches to this problem, in particular identically distributed and non-independent. Experiments on three citation databases evaluate the contributions of Jaccard similarity coefficient and markov logic. Our studies enable us to easily build an entity resolution system.

References

1. Weis, M., Naumann, F., Brosy, F.: A duplicate detection benchmark for XML (and relational) data. In: Proceedings of Workshop on Information Quality for Information Systems (IQIS) (2006)
2. Re, C., Dalvi, N., Suciú, D.: Efficient top-k query evaluation on probabilistic data. In: IEEE 23rd International Conference on Data Engineering, 2007, ICDE 2007, pp. 886–895. IEEE (2007)
3. Panse, F., Van Keulen, M., De Keijzer, A., Ritter, N.: Duplicate detection in probabilistic data. In: 2010 IEEE 26th International Conference on Data Engineering Workshops (ICDEW), pp. 179–182. IEEE (2010)
4. Kopcke, H., Rahm, E.: Frameworks for entity matching: a comparison. *Data Knowl. Eng.* **69**(2), 197–210 (2010)
5. Xiao, C., Wang, W., Lin, X., Yu, J.X., Wang, G.: Efficient similarity joins for near-duplicate detection. *ACM Trans. Database Syst. (TODS)* **36**(3), 15 (2011)
6. Bilenko, M., Mooney, R.J.: Adaptive duplicate detection using learnable string similarity measures. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 39–48. ACM (2003)
7. Bhattacharya, I., Getoor, L.: Iterative record linkage for cleaning and integration. In: Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, pp. 11–18. ACM (2004)
8. Ayat, N., Akbarinia, R., Afsarmanesh, H., Valduriez, P.: Entity resolution for probabilistic data. *Inf. Sci.* **277**, 492–511 (2014)
9. Schewe, K.D., Wang, Q.: A theoretical framework for knowledge-based entity resolution. *Theoret. Comput. Sci.* **549**, 101–126 (2014)
10. Suciú, D., Connolly, A.J., Howe, B.: Embracing uncertainty in large-scale computational astrophysics. In: MUD, pp. 63–77 (2009)
11. Soliman, M.A., Ilyas, I.F., Chen-Chuan Chang, K.: Top-k query processing in uncertain databases. In: IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007, pp. 896–905. IEEE (2007)
12. Singla, P., Domingos, P.: Discriminative training of Markov logic networks. In: AAAI, vol. 5, pp. 868–873 (2005)
13. Hassanzadeh, O., Chiang, F., Lee, H.C., Miller, R.J.: Framework for evaluating clustering algorithms in duplicate detection. *Proc. VLDB Endowment* **2**(1), 1282–1293 (2009)
14. Baxter, R., Christen, P., Churches, T.: A comparison of fast blocking methods for record linkage. In: ACM SIGKDD, vol. 3, pp. 25–27 (2003)
15. Kopcke, H., Thor, A., Rahm, E.: Learning-based approaches for matching web data entities. *IEEE Internet Comput.* **14**(4), 23–31 (2010)
16. Kopcke, H., Rahm, E.: Training selection for tuning entity matching. In: *QDB/MUD*, pp. 3–12 (2008)
17. Singla, P., Domingos, P.: Entity resolution with markov logic. In: Sixth International Conference on Data Mining, 2006. ICDM'06, pp. 572–582. IEEE (2006)
18. Kok, S., Domingos, P.: Learning the structure of Markov logic networks. In: Proceedings of the 22nd International Conference on Machine Learning, pp. 441–448. ACM (2005)
19. Ayat, N., Akbarinia, R., Afsarmanesh, H., Valduriez, P.: Entity resolution for uncertain data. In: BDA'2012: 28e Journées Bases de Données Avancées, p. 20 (2002)
20. Das Sarma, A., Benjelloun, O., Halevy, A., Widom, J.: Working models for uncertain data. In: Proceedings of the 22nd International Conference on Data Engineering, 2006, ICDE'06, pp. 7–7. IEEE (2006)
21. Cohen, W., Ravikumar, P., Fienberg, S.: A comparison of string metrics for matching names and records. In: *Kdd Workshop on Data Cleaning and Object Consolidation*, vol. 3, pp. 73–78 (2003)

22. Yi, K., Li, F., Kollios, G., Srivastava, D.: Efficient processing of top-k queries in uncertain databases with x-relations. *IEEE Trans. Knowl. Data Eng.* **20**(12), 1669–1682 (2008)
23. Yuen, S.M., Tao, Y., Xiao, X., Pei, J., Zhang, D.: Superseding nearest neighbor search on uncertain spatial databases. *IEEE Trans. Knowl. Data Eng.* **22**(7), 1041–1055 (2010)
24. Peng, L., Diao, Y., Liu, A.: Optimizing probabilistic query processing on continuous uncertain data. *Proc. VLDB Endowment* **4**(11), 1169–1180 (2011)
25. McCallum, A., Wellner, B.: Object consolidation by graph partitioning with a conditionally-trained distance metric. In: *KDD Workshop on Data Cleaning, Record Linkage and Object Consolidation* (2003)
26. Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P., Fienberg, S.: Adaptive name matching in information integration. *IEEE Intell. Syst.* **18**(5), 16–23 (2003)

Energy Efficient Algorithms for Hot Spot Problem in Wireless Sensor Networks

Srikanth Jannu and Prasanta K. Jana

Abstract The major bottleneck in the operation of a wireless sensor network (WSNs) is the limited power sources of the sensor nodes. Therefore, effective use of sensors' energy is the most challenging issue for improving network lifetime. In general, the sensor nodes near to the sink have huge traffic load as they relay data from other nodes to reach the sink. Thus, their energy is exhausted quickly and it results in network segmentation. This problem is generally known as hot spot problem. In this paper, we deal with the hot spot problem and present energy efficient clustering and routing algorithms. The simulated results demonstrate that the proposed algorithms perform better compared to the existing algorithms in terms of various performance metrics.

Keywords Wireless sensor networks · Unequal clustering · Routing · Energy consumption

1 Introduction

Potential applications of wireless sensor networks (WSNs) such as military surveillance, environmental monitoring, and health care monitoring have gained the enormous attention of research community [1]. WSNs are collection of hundreds or thousands of tiny sensor devices which are randomly deployed in a target area. The sensor node senses the local information, processes them and sends the information to the base station called sink. The major constraint of the sensor node is that the power of the battery is limited and usually irreplaceable in nature. Therefore, proper

S. Jannu (✉) · P.K. Jana
Department of Computer Science and Engineering,
Indian School of Mines, Dhanbad 826 004, India
e-mail: j.srikanth@live.com

P.K. Jana
e-mail: prasantajana@yahoo.com

utilization of battery power of the sensor nodes has become the challenging task to improve the lifetime of the network. Various energy saving algorithms have addressed on this problem [2]. However, clustering is one of the most prominent areas that have been studied extensively for WSNs.

In a cluster-based WSN, the sensor nodes are divided into groups, called clusters and each cluster has a leader, called cluster head (CH). The responsibilities of CHs are collecting the sensed information from the member sensor nodes of their clusters, aggregating the information and sending them to the sink. Clustering has the following advantages. (1) It reduces energy consumption of the sensor nodes. (2) It conserves communication bandwidth. (3) It also improves the network scalability.

A number of clustering and routing algorithms have been developed which can be found in the survey papers [3, 4]. However, cluster-based multi-hop routing for WSN has been considered as one of the efficient techniques to prolong the lifetime of the network [5–7]. In multi-hop communication, a CH sends its data packet through other CHs to reach the sink. One of the major drawbacks in this technique is that CHs closer to the sink are over burdened with heavy traffic load which results in imbalance of energy utilization causing the CHs which are near to the sink to deplete their energy quickly. Such problem is usually known as hot spot or energy hole problem. Many algorithms have been proposed in the literature to mitigate such problem. However, clustering with unequal size has been the main focused research among them to handle such problems.

In this paper, we address the problem of balancing of the energy consumption among sensor nodes and propose unequal clustering and routing algorithms towards the solution of the hot spot problem. The experimental results are compared with existing algorithms to show the efficiency of the algorithms in terms of number of dead sensor nodes and the residual energy of the network.

Several algorithms have been proposed in the literature, with the objective of efficient utilization of energy by balancing energy consumption of a sensor node. Quite a few unequal clustering and routing algorithms have been proposed [8–11]. In [12], the authors have proposed a distributed cost-based energy balanced clustering and routing algorithm called CEBCRA where CHs are selected and clusters are formed depending on some cost value of the sensor nodes. In the routing phase of this algorithm, the neighbor CH is selected by measuring the cost of each path toward the sink. However, it suffers from the connectivity problem of CHs. Kuila et al. [13] have proposed a cost-based distributed energy balanced clustering and routing algorithm for CH selection and cluster formation. But, the algorithm suffers from the connectivity problem of the selected CHs. An energy-driven unequal clustering algorithm (EDUC) [14] has been proposed. Here, the clusters farther away from the sink have smaller sizes than those closer to the sink which increases the energy hole.

The rest of the paper is organized as follows. Energy model is given in Sect. 2. Section 3 explains the network model. The proposed algorithms are described in Sect. 4. The simulation results are given in Sect. 5 and finally conclusion is specified in Sect. 6.

2 Energy Model

The radio model used in this paper is same as that has been presented in [15]. In the simulation, we use both the free space (fs) and the multipath fading (mp) channel models for energy dissipation. However, the energy dissipation depends on the distance between the transmitter and the receiver [15]. For transmitting a k -bit message to a distance d , the energy expended by the radio is given by the following equation [15]:

$$E_{TX} = k \times E_{\text{elect}} + E_{\text{amp}}(k, d) = \begin{cases} k \times E_{\text{elect}} + k \times e_{\text{fs}} \times d^2, & d < d_0 \\ k \times E_{\text{elect}} + k \times e_{\text{mp}} \times d^4, & d \geq d_0 \end{cases} \quad (2.1)$$

where E_{TX} is the dissipated energy per bit at transmitter, E_{amp} is the amplification factor, E_{elect} is the energy required by the electronic circuit, e_{fs} is the free space coefficient, e_{mp} is the multi path coefficient and d_0 is the distance threshold value.

Similarly, the energy dissipation per bit at receiver is given as follows [15]:

$$E_{RX}(k, d) = k \times E_{\text{elect}}. \quad (2.2)$$

3 Network Model and Terminologies

In this paper, we assume homogeneous WSN i.e., the sensor nodes of the network have the same potential abilities such as processing and communication. The sink is positioned in the middle of the network and also these sensor nodes and the sink are immobile after deployment. We assume that the sensor nodes communicate through symmetric wireless links. A wireless link is established between two nodes if they are within communication range of each other. As the link is symmetric, a node can calculate the distance approximately to another node based on its received signal strength [15]. We use the following notations and terminologies.

- (1) A set of sensor nodes denoted by $S = \{S_1, S_2, \dots, S_n\}$.
- (2) The set of CHs is described by $\xi = \{\text{CH}_1, \text{CH}_2, \dots, \text{CH}_p, \text{Sink}\}$ where $n > p$.
- (3) $\text{dis}(S_i, S_j)$ denotes the distance between S_i and S_j .
- (4) $E_{\text{Residual}}(S_i)$ denotes the remaining energy of S_i .
- (5) $\text{Neighbor}(\text{CH}_i)$: The set of cluster heads, which are within communication range of CH_i . In other words, $\text{Neighbor}(\text{CH}_i) = \{\text{CH}_j | \forall \text{CH}_j \in \xi \text{ and } \text{dis}(\text{CH}_i, \text{CH}_j) < \text{Maximum communication range}\}$.
- (6) *Threshold*: Threshold is a lower bound of the residual energy of a sensor node.

We also assume that the sensor nodes communicate to the sink directly if they are in 1-hop distance to it. Furthermore, those sensor nodes can act as relay nodes for the next level CHs.

4 Proposed Algorithms

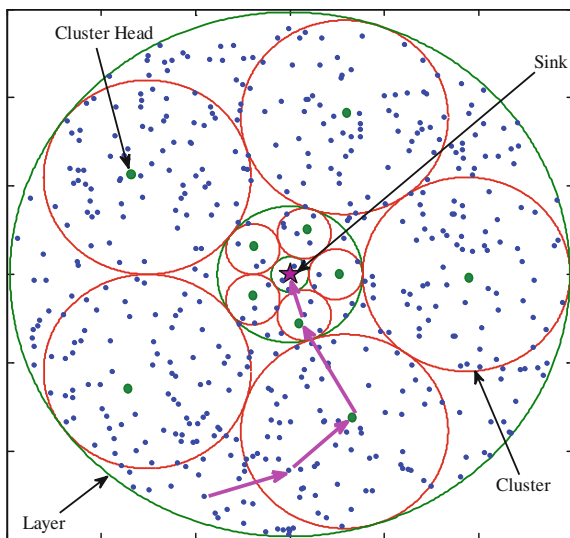
All sensor nodes of the network are divided into unequal size clusters as explained in the initialization phase. With intent, the cluster size is small which is closer to the sink and the cluster size increases as the distance from the sink increases. In every cluster, a normal sensor node is selected as the CH which the sensor node has the minimal distance from other sensor nodes within the cluster and has residual energy more than the threshold value. In that cluster, if residual energy of the CH reaches to the threshold value, another sensor node will be elected as CH as mentioned earlier. The reader may refer Fig. 1 which is obtained from the simulation run of the CH selection phase. Our algorithm works in rounds. Every round is divided into following phases: initialization phase, set-up phase, and routing phase which are described subsequently. The initialization phase deals with the construction of the clusters, the set-up phase with the CH selection and the routing phase with the multi-hop routing for sending the data from CHs to the sink. To present the proposed algorithms, we define several useful terminologies as follows.

Definition 1 (*Total Distance*) Total distance is the sum of distances from one sensor node i to all the sensor nodes within the cluster, which is calculated as follows.

$$D_g(i) = \sum_{j=1}^n d_g(i, j), \quad (4.3)$$

where $d_g(i, j)$ is the distance between sensor node i to sensor node j in the cluster g .

Fig. 1 Organization of the proposed method



Definition 2 (*Minimal Distance*) This is the minimum total distance of cluster g defined as follows:

$$M_g(i) = \min_{1 \leq i \leq n} \{D_g(i)\}. \quad (4.4)$$

Definition 3 (*Sink Distance*) This is the distance between a CH to the sink and defined as follows.

$$\text{dis}(\text{CH}_g, \text{sink}) = \sqrt{(x_g - x_{\text{sink}})^2 + (y_g - y_{\text{sink}})^2}, \quad (4.5)$$

where x_g and y_g are the coordinates of CH of the grid g ; x_{sink} and y_{sink} are the coordinates of the sink.

4.1 Initialization Phase

In this phase, at first we locate the sink at the middle of the sensing area. Then, we divide the area into clusters with various sizes and layers. Each layer contains n equal number of clusters with same size as shown in Fig. 1. We obtain the radius of the cluster of the 1st layer as follows:

$$r_1 = \frac{R \times \sin\left(\frac{\Pi}{n}\right)}{1 - \sin\left(\frac{\Pi}{n}\right)}.$$

Here, R is the communication range where the sensor nodes can communicate to the sink in a single hop distance. Generally, the radius r_m for the m th layer cluster is given by the following equation:

$$r_m = \frac{(2 \times (r_1 + r_2 + \dots + r_{m-1}) + R) \times \sin\left(\frac{\Pi}{n}\right)}{1 - \sin\left(\frac{\Pi}{n}\right)}. \quad (4.6)$$

Note the radius of the sensing area is $2 \times (r_1 + r_2 + \dots + r_m) + R$.

5 Set-up Phase

In this phase, we select efficient CHs for all the clusters. The sensor nodes which are in 1-hop distance to the sink need not to form clusters and also they can communicate directly to the sink. Furthermore, these sensor nodes act as relay nodes for next layer CHs. For the next layers, we calculate the total distances of all the sensor nodes in a cluster. In that cluster, a node which has the minimum total distance and residual energy greater than the threshold value is designated as CH. Some sensor nodes which

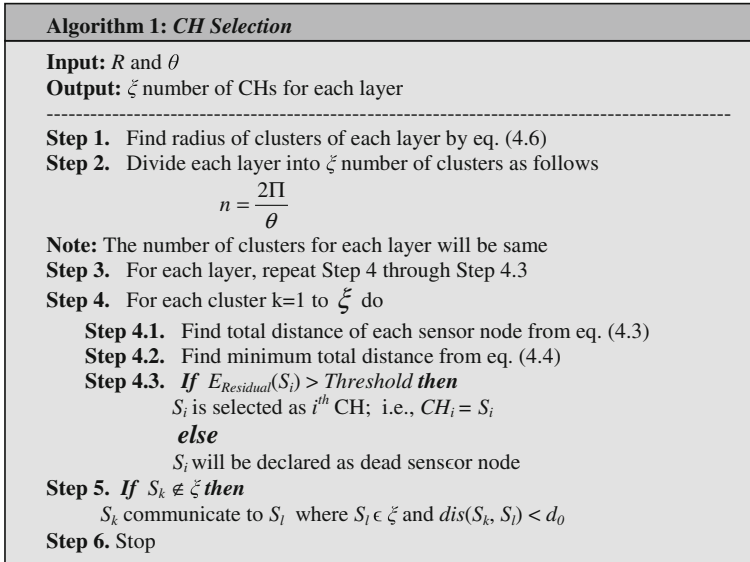


Fig. 2 CH selection algorithm

do not belong to any cluster can communicate to the nearest sensor nodes which belong to a cluster. The detailed CH selection algorithm is presented in Fig. 2.

5.1 Routing Phase

The proposed multi-hop routing works as follows. In a particular cluster i , a CH C_i finds all the CHs of its neighbor clusters and then forwards data to a CH C_j which has the least distance from the sink. C_j follows the same procedure for forwarding the data to the next CH of one of its neighbor clusters. This process is continued until the data is finally reached to a top layer sensor node. Moreover, that sensor node must have significant residual energy. However, all the sensor nodes with one hop distance from the sink can directly send their data to the sink. The detailed multi-hop routing algorithm is presented in Fig. 3.

6 Simulation Results

The proposed algorithms were experimented extensively using MATLAB (version 7.5) and C programming language on an Intel Core i7-2600 processor, 3.40 GHz CPU and 2 GB RAM running on the platform Microsoft Windows 7 professional. For the experiments, we assumed the scenario as follows. We considered a circular

Algorithm 2: Multi-Hop Routing

Input: Distance to its neighbor and sink.
Output: Selection of a CH which is near to sink *NextHop*, such that the required energy to transmit the data packet to sink is minimized.

Step 1: $NextHop = Sink$ /* initially, next-hop CH is Sink. */
Step 2: $SinkNeighbour(C_i) = Sink_CH \cap Neighbor(C_i)$
Step 3: **if** ($|SinkNeighbour(C_i)| == 0$) **then**
 $NextHop(C_i) = C_j$,
 where, $dis(C_i, C_j) = \text{Min} \{dis(C_i, C_l), \forall C_l \in Neighbor(C_i)\}$
 else
 $NextHop(C_i) = C_j$.
 where,
 $dis(C_i, C_j) = \text{Min} \{dis(C_i, C_l), \forall C_l \in SinkNeighbour(C_i)\}$
Step 4: The C_i sends the data to the sink using C_j as next-hop CH towards sink.
Step 5: Stop.

Fig. 3 Multi-hop routing algorithm

area of radius 60 m in which 1000 sensor nodes are randomly deployed. Each sensor node was assumed to have an initial energy of 0.5 J. The typical parameters and rounds were set same as EDUC [14]. For the sake of comparison, we also ran various existing algorithms including CEBCRA [12] and EDUC [14].

6.1 Number of Dead Sensor Nodes

We compare the proposed algorithms with CEBCRA and EDUC algorithms in terms of number of dead sensor nodes per round as shown in Fig. 4. Here, proposed

Fig. 4 Comparison in terms of number of dead sensor nodes

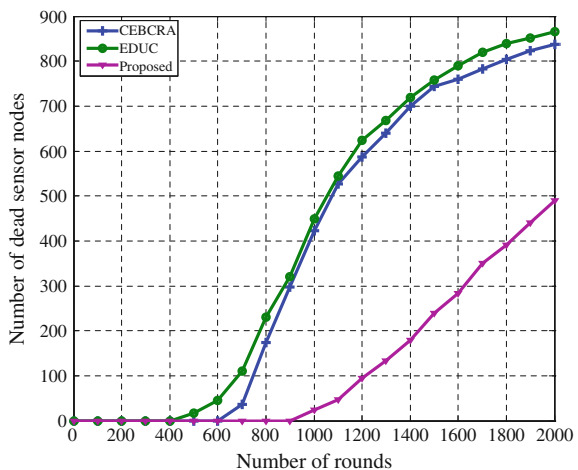
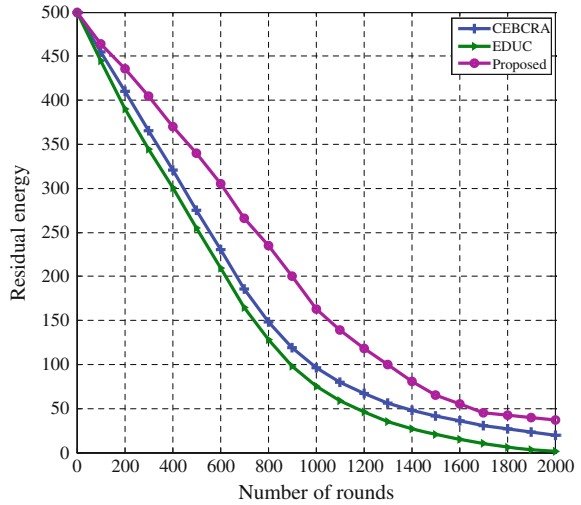


Fig. 5 Comparison in terms of residual energy



algorithm performs better than CEBBCRA and far better than EDUC. The rationale behind is that CEBBCRA forms the cluster based on the maximum residual energy of the CH which leads to the maximum load of the CH. In EDUC, the clusters farther away from the sink have smaller sizes than those closer to the sink which increases the number of dead sensor nodes. However, proposed algorithm selects the CH based on the minimal distance, residual energy of the sensor node and forms the cluster such a way that the size of the cluster which is near to the sink is smaller than the cluster which is farther away from the sink. As a result, proposed algorithm performance better in terms of number of dead sensor nodes of the network.

6.2 Residual Energy

We also ran the algorithms and compare the results with CEBBCRA and EDUC in terms of energy consumption of the network per round as shown in Fig. 5. Here also we get better results for our proposed work.

7 Conclusion

In this paper, we have presented energy efficient algorithms for unequal clustering and routing for wireless sensor networks with the aim of energy saving of the sensor nodes. The proposed algorithms also addressed the hotspot problem. We have tested our algorithms extensively and simulation results have shown that the

proposed algorithms perform better than the existing algorithms, CEBCRA, and EDUC with respect to energy consumption of the network and the number of dead sensor nodes.

References

1. Akyildiz, I.F. et al.: A survey on sensor networks. *IEEE Commun. Magaz.* **40**(8), 102–114 (2002)
2. Calhoun, B.H. et al.: Design considerations for ultra-low energy wireless microsensor nodes. *IEEE Trans. Comput.* **54**(6), 727–740 (2005)
3. Bari, A., Jaekel, A., Jiang, J., Xu, Y.: Design of fault tolerant wireless sensor networks satisfying survivability and lifetime requirements. *Comput. Commun.* **35**(3), 320–333 (2012)
4. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for ireless sensor networks. *Comput. Commun.* **30**, 2826–2841 (2007)
5. Fariborzi, H., Moghavvemi, M.: EAMTR: energy aware multi-tree routing for wireless sensor networks. *IET Commun.* **3** (5), 733–739 (2009)
6. Gagarin, A., Hussain, S., Yang, L.T.: Distributed hierarchical search for balanced energy consumption routing spanning trees in wireless sensor networks. *J. Parallel Distrib. Comput.* **70**(9), 975–982 (2010)
7. Ren, F., Zhang, J., He, T., Lin, C., Das, S.K.: EBRP: energy-balanced routing protocol for data gathering in wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **22**(12), 2018–2125 (2011)
8. Ye, M., Li, C.F., Chen, G.H., Wu, J.: EECS: An energy efficient clustering scheme in wireless sensor networks. In: *IEEE International Performance Computing and Communications Conference (IPCCC)*, 535–540 (2005)
9. Wei, D., Jin, Y., Vural, S., Moessner, K., Tafazolli, R.: An energy-efficient clustering solution for wireless sensor networks. *IEEE Trans. Wirel. Commun.* **10**, 3973–3983 (2011)
10. Challal, Y., Ouadjaout, A., Lasla, N., Bagaa, M., Hadjidj, A.: Secure and efficient disjoint multipath construction for fault tolerant routing in wireless sensor networks. *J. Network Comput. Appl.* **34**(4), 1380–1397 (2011)
11. Lai, Y., Chen, H.: Energy-efficient fault-tolerant mechanism for clustered wireless sensor networks. *IEEE ICCCN.* 272–277 (2007)
12. Kuila, P., Jana, P.K.: An energy balanced distributed clustering and routing algorithm for wireless sensor networks. In: *2nd IEEE International Conference Parallel Distributed and Grid Computing (PDGC)*, pp. 220–225 (2012)
13. Azharuddin, Md, Kuila, P., Jana, P.K.: Energy efficient fault tolerant clustering and routing algorithms for wireless sensor networks. *Comput. Electr. Eng.* (2014). <http://dx.doi.org/10.1016/j.compeleceng.2014.07.019>
14. Yu, J., Qi, Y., Wang, G.: An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks. *J Control Theor. Appl.* **9**(1), 133–139 (2011)
15. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **1**(4), 660–670 (2002)

Mining Community-Based Top- k Experts and Learners in Online Question Answering Systems

S. Rao Chintalapudi and M.H.M. Krishna Prasad

Abstract Online Question Answering Systems are very popular and helpful for programming community. In these systems, users can post questions, answer the questions, collaboratively tag the questions, and vote for quality answers. This paper implements a link structure-based Top- k Experts and Learners finding algorithm using Stanford Network Analysis Project (SNAP) Library. Experiments are done on real data taken from Stack Overflow that mainly focuses on computer programming and the results show that link analysis techniques are more suitable for analyzing online question answering systems.

Keywords Graph mining · Social network analysis · Expert finding · Recommendation system · Community-based question answering system

1 Introduction

Online Question Answering Systems are very beneficial to the programming community which has gained more popularity in recent years. Here, users ask questions related to different programming languages such as C, C++, HTML, Java, Python, PHP, and JavaScript. These are used to share the knowledge on open platform and produces large volume of knowledge. Finding experts from this valuable knowledge in a particular domain or community is an important problem [1] and used in many real world applications viz. selecting best answers and routing new questions to the appropriate experts. Expert finding is a task that finds users who can provide good number of high quality and reliable answers. This problem is very useful in Natural

S.R. Chintalapudi (✉) · M.H.M. Krishna Prasad
Department of CSE, University College of Engineering Kakinada (A),
JNTUK, Kakinada, Andhra Pradesh, India
e-mail: srao.chintalapudi@gmail.com

M.H.M. Krishna Prasad
e-mail: krishnaprasad.mhm@gmail.com

language Processing (NLP) and Information Retrieval (IR) areas. The sites like Yahoo! Answers [2], Stack Overflow [3], Stack Exchange [4], and Quora [5] are the popular websites in this category. The primary focus of this paper is to find Top- k experts and learners in a particular domain (e.g., Java, C++, python... etc.) using link analysis methods.

The existing online Question Answering Systems [6] are not meeting the user expectations for the following reasons: (1) poor quality in finding expertise whenever a new question comes, that leads higher latency. (2) Contains some low quality answers often receives lower ratings. (3) Before posting a question, unable to browse archived similar questions and their respective answers. In fact, all these problems are closely related and the fundamental solution is how to model expertise and topics.

Two traditional approaches for finding experts in Online Question Answering Systems are topic-based algorithms [7–10] and authority-based algorithms [11–16]. The topic-based algorithms are designed based on latent topic model techniques such as Latent Dirichlet Allocation (LDA). These algorithms fit the content into a probabilistic model and then predict the required values. The authority-based algorithms construct the network between users based on their question and answering and predictions are done using authority model.

The rest of the paper is organized as follows. Section 2 describes different existing approaches related to expert finding concept. Top- k experts and learners finding using Pagerank and Hypertext-Induced Topic Selection (HITS) is discussed in Sect. 3. Section 4 provides systematic experimental setup and dataset preparation downloaded from Stack Overflow. The Results and implementation details were discussed in Sect. 5 and lastly, Sect. 6 concludes our study.

2 Related Work

This section briefly reviews some of the related works on expert finding algorithms in Online Question Answering System.

Most of the existing expert finding algorithms are based on the link analysis approaches viz. Pagerank and HITS. Pagerank Algorithm (PR) [17, 18] finds experts using link structure between the users in Online Question Answering System, which is similar to ranking pages on the web. HITS algorithm [19] is also a link structure-based method to find experts in OQAS using estimated ranking scores.

Expertise Rank (ER) [12] is a link structure-based approach similar to Pagerank Algorithm and assigns ranks to the experts by taking number of users involved in expertise network into account. Here, expertise network can be constructed using question ask–answer relationship between users. It considers the magnitude of interaction between users; so, it represents OQA system as directed weighted graph. It cannot be used for authoritative users.

Topic Sensitive Probabilistic Model (TSPM) [7, 14] is an extension of Pagerank Algorithm for finding experts in online Question Answering System. It takes both link structure and topic similarity between users into account and uses LDA-based latent topic model.

Authority Rank (AR) [6] algorithm calculates user authority based on number of best answers. This approach uses in-degree method for computing number of best answers. It constructs a directed weighted graph, where edge weight represents the magnitude of interaction between asker and answerer, but it fails to identify authority nodes accurately.

3 Top-*k* Experts and Learners Finding Algorithm

Link Analysis techniques are widely used for finding experts and relative authority of each node in OQA Systems, and identifying a suitable model plays a major role for this problem. Pagerank and HITS algorithms are prominent approaches that are used to find relative authority of a node in a network.

Initially, Pagerank algorithm assigns equal Pagerank values to all the nodes using the formula $1/n$ where n is number of nodes in the network. Identify the nodes from which the current node has in-links. Compute out-degree of these nodes and if the out-degree is greater than zero then update the pagerank score by dividing with its out-degree. Now, multiply the current values of page rank with damping factor (i.e., $\beta = 0.85$) and add leaked pagerank. Repeat the iterations until the difference between previous and current iteration is below threshold. If the sum of the ranks of in-links for a node is high, then the node will have higher Pagerank value. It considers both nodes having many in-links and nodes having few highly ranked in-links. This algorithm will provide interesting results when the relationship between users is around a single subject such as Java.

HITS algorithm computes *hubs* and *authorities* in a network, where *hubs* are the special nodes that contain collections of links to *authorities*. A good *hub* is a node pointed to good *authorities* while good *authority* is a node pointed to by good *hubs*. In the case of online question answering systems, experts/best answerers are considered as authorities and learners/askers are considered as hubs. In this algorithm, two scores can be computed for each node namely, authority score and hub score. This algorithm was originally designed to rank the web pages but the proposed approach uses the same for online question answering systems to mine Top-*k* experts and learners. The overview of Top-*k* Java Experts and Learners Finding Algorithm for Stack Overflow data is shown in Fig. 1.

3.1 Algorithm

The following algorithm takes original dataset in XML format as input and produce Top-*k* Experts and Learners. It utilizes the link analysis algorithms viz. Pagerank,

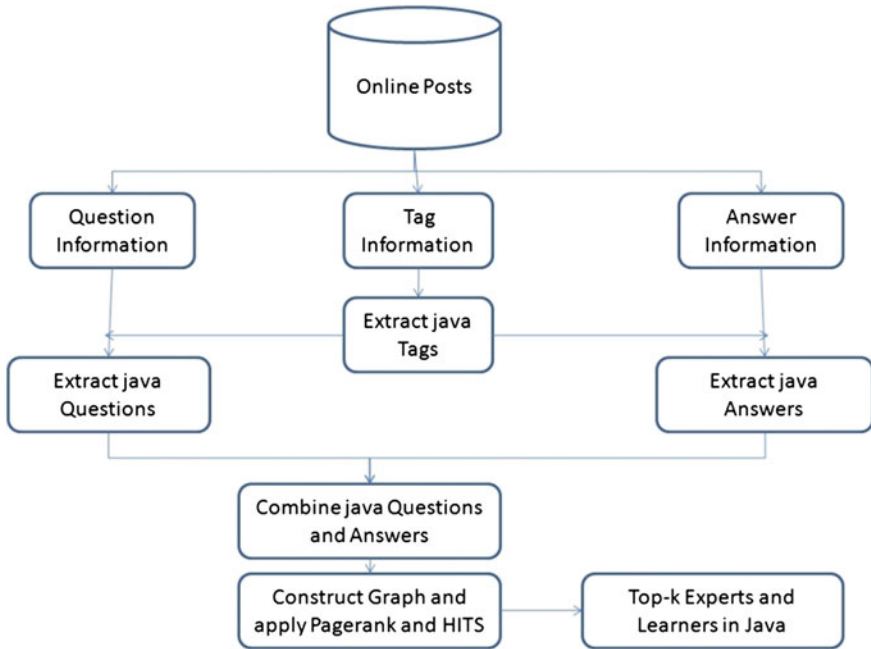


Fig. 1 Overview of Top- k java experts and learners finding algorithm for stack overflow data

HITS in order to find rank scores. The steps involved in solving experts finding problem are described as follows.

Input: *onlineposts.xml*.

Output: *user ids* and *rank scores*.

Step 1: Extract all the question posts and accepted answers.

Step 2: Extract all the answer posts.

Step 3: Extract java question ids from input file.

Step 4: Extract only java posts using java question ids.

Step 5: Combine java posts and answers based on respective user ids of accepted answers.

Step 6: Construct a graph with users as nodes and their communication as edges.

Step 7: Apply Pagerank and HITS algorithms on the graph.

In the above algorithm, node represents owner id and the edge represents association between question owner and accepted answer owner. HITS returns both Top- k experts and learners as it uses the concept of *authorities* and *hubs* where as Pagerank returns only Top- k experts.

4 Experiments

4.1 Experimental Setup

This algorithm is experimented on a machine with Intel core i7-3770 CPU 3.40 GHz, 4 GB of primary memory, 500 GB of secondary memory and ubuntu 14.04 LTS 64-bit operating system. Python 2.7.6 is used for the implementation of the algorithm and python version of SNAP library [20] is installed for graph analytics.

4.2 Dataset

The real dataset is used from popular online question answering website Stack Overflow that focuses on computer programming and it is publicly available from Stack Overflow archives [21]. It consists of 1,98,81,020 posts from July 2008 to May 2014 with the size of 5.2 GB compressed/26 GB uncompressed. Questions and Answers are extracted to different files from the original dataset based on their tag information for simplifying the computation and this preprocessing step reduces the dataset size into Mega Bytes. In specific, authors have used java tag information to extract java posts for the current application. The summary of dataset is shown in Table 1 and sample records of the original dataset are shown in Fig. 2.

Table 1 Stack overflow dataset summary

Item description	Count
Posts	1,98,81,020
Questions	72,14,697
Answers	1,26,09,623
Java posts	6,32,493

Fig. 2 Sample records of original dataset provided by stack overflow

```

Sample Records in Onlineposts.xml
Questions Record Format:
<row Id="4" PostTypeId="1"
OwnerUserId="8" AcceptedAnswerId="7"
Tags="&lt; c#&gt; &lt; winforms&gt; &lt; forms&gt;
&lt; opacity&gt;" ... />

Answers Record Format:
<row Id="12" PostTypeId="2" OwnerUserId="1" ... />

#PostTypeId="1" represents Question
#PostTypeId="2" represents Answer
    
```

4.3 SNAP Python API

SNAP is an open source and general purpose software for analyzing large scale networks. It is used to study web data and social networks. It provides huge number of built in functions to simplify network analytics and is a large repository for real world network data sets such as facebook, twitter, and amazon. This software is implemented using both C++ and Python. Python API is used for the Top- k experts and Learners finding algorithm.

```

mhm@mhm: ~/OQAS
Top-K Java Experts based on PageRank Algorithm
id 992484, pagerank 0.013981
id 135152, pagerank 0.010006
id 22656, pagerank 0.007104
id 139985, pagerank 0.005521
id 157882, pagerank 0.004597
id 57695, pagerank 0.003737
id 571407, pagerank 0.003572
id 70604, pagerank 0.003502
id 218978, pagerank 0.002918
id 103154, pagerank 0.002459

Top-K Java Experts based on HITS Algorithm
id 22656, authority rank 0.604724
id 157882, authority rank 0.298699
id 571407, authority rank 0.283907
id 57695, authority rank 0.269305
id 139985, authority rank 0.249401
id 203907, authority rank 0.183517
id 522444, authority rank 0.102917
id 131872, authority rank 0.099855
id 230513, authority rank 0.099155
id 276052, authority rank 0.098242

Top-K Java Learners based on HITS Algorithm
id 892029, hub rank 0.073364
id 1194415, hub rank 0.059551
id 359862, hub rank 0.056876
id 648138, hub rank 0.055593
id 470184, hub rank 0.053179
id 384706, hub rank 0.052113
id 802050, hub rank 0.050122
id 44330, hub rank 0.048781
id 130758, hub rank 0.048099
id 454049, hub rank 0.048083
Thu Feb 26 17:51:56 IST 2015 ... END analysis of StackOverflow
mhm@mhm:~/OQAS$

```

Fig. 3 Screen shot of results that includes Top- k java experts finding using pagerank and Top- k java experts and learners using HITS

5 Results

Top-*k* experts and learners finding algorithm for Stack Overflow data is implemented in Python using SNAP Library. Here, two major link analysis approaches Pagerank and HITS are used; former one is used for finding Top-*k* experts and later algorithm is used to find both Top-*k* experts and learners as it uses the concept of authorities and hubs. The screenshot of the result is shown in Fig. 3. Here, Top-10 experts are displayed at the top based on Pagerank score and higher Pagerank score represents that the user has more expertise in that subject. The HITS algorithm is also applied on the same dataset to find Top-10 experts and learners in java and the results are shown in middle and bottom of the screenshot, respectively. Higher authority and hub rank score corresponds to expertise and learning capacity of the user in the specified subject because authorities are considered as answerers and hubs are considered as askers.

6 Conclusion and Future Work

This paper implements a Top-*k* Experts and Learners finding algorithm for Online Question Answering Systems using link analysis techniques such as Pagerank and HITS. The experiments are conducted on real world dataset taken from Stack Overflow. Experimental results revealed that link analysis techniques are more suitable for analyzing online question answering systems.

In future, this algorithm may suffer from big data problem as the size of the dataset in online question answering systems is increasing rapidly, which leads scarcity of main memory to store and analyze the data. So, one need to adopt distributed and parallel programming environment. Hence, authors planned to implement the same algorithm using hadoop Map Reduce frame work.

References

1. Zhao, Z., Zhang, L., He, X., Ng, W.: Expert finding for question answering via graph regularized matrix completion. *IEEE Trans. Knowl. Data Eng.* **99**, 1–12 (2014)
2. Yahoo Answers: <http://answers.yahoo.com>
3. Stack Overflow: <http://www.stackoverflow.com>
4. Stack Exchange: Hot Questions. <http://www.stackexchange.com>
5. Quora—Your Best Source for Knowledge: <http://www.quora.com>
6. Liu, Y., Minghui, Q., Swapna, G., Feida, Z., Jing, J., Huiping, S., Zhong, C.: CQARank: jointly model topics and expertise in community question answering. In: *CIKM'13*, pp. 99–108. ACM, San Francisco (2013)
7. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums: the case of yahoo! Answers. In: *SIGKDD'08*, pp. 866–874. ACM (2008)

8. Jurczyk, P., Agichtein, E.: Discovering authorities in question answer communities by using link analysis. In: CIKM'07, pp. 919–922. ACM (2007)
9. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: WWW'07, pp. 221–230. ACM (2007)
10. Zhu, H., Chen, E., Xiong, H., Cao, H., Tian, J.: Ranking user authority with relevant knowledge categories for expert finding. *World Wide Web*, 17, pp. 1081–1107. Springer US (2014)
11. Deng, H., King, I., Lyu, M. R.: Formal models for expert finding on dblp bibliography data. In: ICDM'08, pp. 163–172. IEEE (2008)
12. Guo, J., Xu, S., Bao, S., Yu, Y.: Tapping on the potential of q&a community by recommending answer providers. In: CIKM'08, pp. 921–930. ACM (2008)
13. Liu, X., Croft, W.B., Koll, M.: Finding experts in community based question-answering services. In: CIKM'05, pp. 315–316. ACM (2005)
14. Zhou, G., Lai, S., Liu, K., Zhao, J.: Topic sensitive probabilistic model for expert finding in question answer communities. In: CIKM'12, pp. 1662–1666. ACM, USA (2012)
15. McCallum, A., Wang, X., Corrada-Emmanuel, A.: Topic and role discovery in social networks with experiments on enron and academic email. *J. Artif. Intell. Res.* **30**, 249–272, AI Access Foundation, USA (2007)
16. Zhao, T., Bian, N., Li, C., Li, M.: Topic-level expert modeling in community question answering. *SDM*, pp. 776–784, SIAM (2013)
17. Page, L., Brin, S., Motwani, R., Winograd, T.: The Pagerank citation ranking: bringing order to the web. *Stanford Digital Library Technologies Project* (1998)
18. Berkhin, P.: A survey on pagerank computing. *Internet Math.* **1**, 73–120 (2005)
19. Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *J. ACM.* **46**, 604–632. ACM (1999)
20. SNAP: Stanford Network Analysis Project. <http://snap.stanford.edu>
21. Stack Overflow Dataset: <https://archive.org/download/stackexchange/stackoverflow.com-Posts.7z>

An Approach for Creating Framework for Automated Question Generation from Instructional Objective

Syaamantak Das, Rajeev Chatterjee and Jyotsna Kumar Mandal

Abstract Generating questions for evaluation and assessment of learner's knowledge, in an e-learning system, is usually a tedious job for the instructor. There are approaches being made to automatically generate questions from content of an e-learning course. However, there is a general procedure of using only the content provided by the instructor, to generate questions, which are applicable on these automated question generation systems. Here a proposed approach to develop questions from the instructional objectives of an e-learning system, irrespective of the e-learning content, is presented. These questions which are being generated are aimed to satisfy the instructional objectives of the instructor and learning objectives to the learner. This reduces the necessity of questions being generated from a particular content and provides a proposed methodical framework to generate questions having similar instructional and learning objectives.

Keywords Question generation · Item generation · Assessment · Instructional objectives

1 Introduction

In an e-learning course, a given question set consists of a number of questions called items. There must be enough number of questions in the given set to avoid repetition. The increase in number of items may make the job tedious for the

S. Das (✉) · R. Chatterjee

Department of Computer Science and Engineering, NITTTR, Kolkata, India
e-mail: syaamantak.das@gmail.com

R. Chatterjee

e-mail: chatterjee.rajeev@gmail.com

J.K. Mandal

Department of Computer Science and Engineering, University of Kalyani, West Bengal, India
e-mail: jkm.cse@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_51

527

instructor to form questions. Thus, generation of questions based on instructional objective, if automated, may be beneficial to the instructors.

1.1 Scope of the Proposed Method

The proposed method is based on test cases for generation of questions for basic school level mathematics. Mathematics, being a well-structured subject in syntax and semantic parameter, is chosen as the basis for test cases. Instructional objectives of basic school level mathematics are simple and comprehensive, matching to the requirement of a sample instructional objective.

1.2 Category of Questions and Assessment

Questions can be of multiple types [1], such as factual, convergent, and divergent. Among them, the factual, convergent, and evaluative types for question generation are considered in the proposed research work. Factual questions require direct answers based on given facts. They are answered in either correct/incorrect or in one word. Convergent questions have answers in a given range of accuracy. Evaluative questions require higher levels of cognitive knowledge. For an e-learning system, aforesaid categories of questions are suitable for assessment. An example would be solving mathematics problems which require knowledge and application of multiple mathematical functions. Considering divergent or combined type of questions for automatic generation is currently out of scope of the proposed work. Assessment can be divided into multiple types such as diagnostic, formative, summative, and criteria referenced. Currently formative, summative and to a certain extent, criteria referenced questions are considered. A formative assessment assesses a students' performance at regular interval throughout the instruction process. Summative assessment measures students' performance at the end. Criteria referenced measure a students' performance against a specific objective.

The paper is organized as follows. In Sect. 2, review of existing works on related topics is given. Section 3 explains the proposed approach and methodology. Section 4 provides a sample result. Section 5 analyzes the generated questions, Sect. 6 describes the limit, scope, and future work and Sect. 7 is conclusion.

2 Review of Existing Work

Alsubait et al. [2], in their paper, proposed a method to use currently available higher quality ontologies as a base for mining of analogies in multiple choice-based question. Papasalouros et al. [3] proposed a method that is based on the

domain-specified ontologies. The proposed method is not dependent of lexicons or any other linguistic-based resources. Ali et al. [4] in their paper proposes a question generation system method that generates questions. An advanced system for question generation from text was proposed by Heilman [5] in his doctoral thesis. Skalban et al. [6] in their paper gave example of how questions can be generated from videos combined with text. Huang et al. [7] in their paper present an automated quiz question generation based on video clips. Iwane et al. [8] in their paper propose a framework which is developed using action of learner and using an expository text’s context knowledge. Paul et al. [9] generates a template that is dynamic in nature for question paper generation. Kale and Kiwelekar [10] in their paper present design of an algorithm that can be used for generation of question paper template. Al-Yahya [11] in his paper describes a system, an engine which is based on domain ontologies. Johnson et al. [12] in their paper introduce a new kind of open source quiz generator that is platform independent.

3 The Proposed Methodology

The proposed methodology is divided into following steps:

- Step 1: Writing an instructional objective as per Audience, Behavior, Condition, and Degree (ABCD) model shown in Fig. 1.
- Step 2: Mapping keyword present in “Behavior” part of instructional objective with Bloom’s Table of Action verb.
- Step 3: Obtaining the mapped keyword’s question template.
- Step 4: Concatenate the question template with Instructional objective’s Condition and Degree, respectively, to generate the question framework.
- Step 5: Selecting the values to be used as operand.

XML have been used for the implementation of the technique. XML is defined by a set of rules for encoding documents in a format that is both human-readable as well as machine-readable. For example, the student will be able to solve addition of numbers up to two decimal places. Here, this simple English sentence is broken down as per ABCD model using XML.

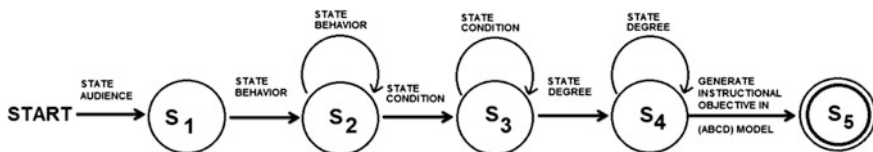


Fig. 1 State diagram of automation for creating instructional objective in ABCD model

E.g., <instructional_objective>

<audience> the student </audience>
 <behavior> will be able to solve </behavior>
 <condition> addition of numbers </condition>
 <degree> up to two decimal places </degree>

</instructional_objective>

“Solve” is the keyword in “Behavior” part of instructional objective. In this case, “solve” is mapped with Bloom’s Taxonomy of Action Verbs [13]. It is mapped to application level of cognitive domain. The keyword “solve” has the question template [14, 15]—“Apply, calculate, complete, show, solve, modify, etc.” The question template is concatenated with the condition of instructional objective which is “addition of numbers.” It is the further concatenated with the degree of the instructional objective which is “up to two decimal places.” Thus, the final question framework which is generated is (question template)+(conditions of instructional objective) + (degree of instructional objective). The state diagram for question generation is shown in Fig. 2. E.g., <solve> <addition of numbers> <up to two decimal places.> similarly for subtraction and multiplication the following question framework is generated—

- <Calculate> <subtraction of numbers> <up to two decimal places.>
- <Show> <multiplication of numbers> <up to two decimal places.>

Based on the final question format, the operand is selected. An operand is the part of a computer instruction that specifies data that is to be operated or manipulated. In mathematics, an operand is the object of a mathematical operation. The objective is to select the operand based on random values. This can generate maximum possible items in the given range of numbers. Two values for the two operands were taken, within a pre-defined range denote by L for the lowest value and H for the highest value. A check is put on for any repetition by storing the values in a repository. Any repeated values will be discarded. The range of number needs to be sufficiently large.

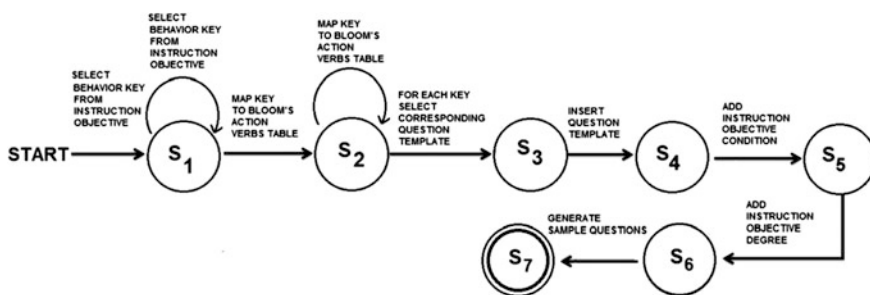


Fig. 2 State diagram of automation of question generation

3.1 *Model for Question Generation*

Elementary mathematics was used as the model framework for question generation using question template with bloom's keyword along with condition and degree parameters of ABCD model. C program has been used to implement it, by creating a function called "question generator framework," which includes question template consisting keywords from blooms taxonomy, condition of instructional objective, and degree of instructional objective as parameters. In this particular case, five blooms taxonomy keywords were used for mathematics, which are "Apply," "Calculate," "Complete," "Solve," and "Show" as question template. Four task keywords addition, subtraction, multiplication, and division on numbers were used as condition for instructional objective.

3.2 *Algorithm for the Proposed Work*

Begin:

Step 1: **For** each instructional objective in course, create question template in ABCD question format.

Step 2: Each ABCD question format consist of three parts - question template, condition and degree.

Step 3: Insert 'audience' part of instructional objective. // example of audience - Student.

Step 4: Insert 'behavior' part of instructional objective. // example of behavior - able to solve.

Step 5: Insert 'condition' part of instructional objective. // example of condition - addition of numbers.

Step 6: Insert 'degree' part of instructional objective. // example of degree - upto 2 decimal place.

End of For loop. //Step 1.

Step 7: **For** each instructional objective, select the 'behavior' keyword and compare it with blooms action verb.

Step 8: If the 'behavior' keyword matches with blooms action verb
then put bloom's action verb into the question template.

Step 9: Select the 'condition' and insert it into condition part of ABCD question format.

Step 10: Select the 'degree' and insert it into degree part of the ABCD question format;

Step 12: Create final question by concatenating the question template with condition and degree of the instructional objective.

Step 13: Generate random value for operands in each question

End of for loop. // Step 7.
End.

4 Results

A sample set of input instructional objectives is given in Table 1. Figure 3 shows the output questions. Table 2 presents the information regarding generation of item with level of accuracy.

5 Analysis of the Generated Questions

The questions that were generated were tested for syntax and semantics. Majority of the questions successfully passed the grammatical and syntactical benchmark when compared to similar questions available in textbooks of school level. However, the further refining of the question template will be required so that the language quality of the questions that is being generated can be developed. In this proposed work, the approach was to use the ABCD model to creating questions from the instructional objectives.

Table 1 A sample set of input instructional objective

Sl. no.	Audience	Behavior	Condition	Degree
1.	The student	Will be able to solve	Addition of numbers	Up to two decimal places
2.	The student	Will be able to solve	Subtraction of numbers	Up to two decimal places
3.	The student	Will be able to solve	Multiplication of numbers	Up to two decimal places

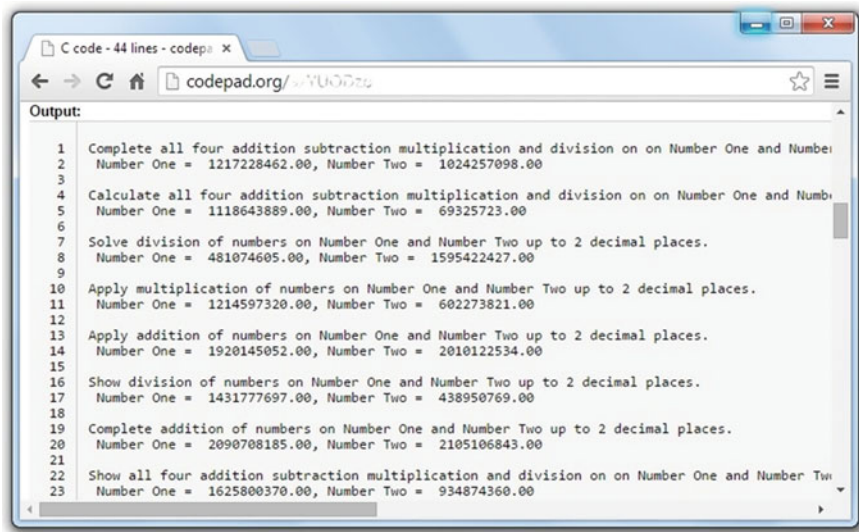


Fig. 3 Sample screenshot showing questions generated with random values for operands

Table 2 Sample tests information regarding generation of item with level of accuracy

No. of question template	No. of conditions in instructional objective	No. of degree in instructional objective	No. of operands	No. of questions generated	No. of questions correctly generated	Remarks
5	5	5	2	50 (Test case)	49	98 % accurate

6 Limit, Scope, and Future Possibility of the Approach

The proposed approach of creating a framework for automated question generation based on instructional objective is currently limited to elementary level of mathematics. This approach is predominantly dependent on instructional objectives created as per ABCD model. With complex instructional objectives, the approach may

not yield the desired and appropriate results. Depending upon the category of assessment required, better and advanced algorithms based on the proposed approach may be able to create a framework for a variety of subjects and assessment methods.

7 Conclusion

This proposed work is a new approach for automated generation of questions which is not limited to the content of the course and may be used as one of the approaches for the process. The method goes well with subjects like mathematics and technological courses, but the scope is limited for subjects of linguistics and humanities.

References

1. Inamullah, H.M.: A study of lower-order and higher-order questions at secondary level. *Asian Soc. Sci.* **7**(9), 149 (2011)
2. Alsubait, T., Parsia, B., Sattler, U.: Automatic generation of analogy questions for student assessment: an ontology-based approach. *Res. Learning Technol.* **20**, (2012)
3. Papasalouros, A., Kanaris, K., Kotis, K.: Automatic generation of multiple choice questions from domain ontologies. In: *IADIS e-Learning Conference*, pp. 427–434 (2008)
4. Ali, H., Chali, Y., Hasan, S.: Automatic question generation from sentences. In: *Proceedings of QG2010: The Third Workshop on Question Generation*, pp. 58–67 (2010)
5. Heilman, M.: Automatic factual question generation from text. Ph.D. Dissertation, Carnegie Mellon University (2011)
6. Skalban, Y., Le, A., Ha., Specia, L., Mitkov, R.: Automatic question generation in multimedia-based learning. In: *COLING (Posters)*, pp. 1151–1160 (2012)
7. Huang, Y.T., Tseng, Y.M., Sun, Y.S., Chen, M.C.: TEDQuiz: Automatic Quiz Generation for TED Talks Video Clips to Assess Listening Comprehension. In: *IEEE 14th International Conference on Advanced Learning Technologies*, pp. 350–354. IEEE Press, Athens (2014)
8. Iwane, N., Chunming, G., Yoshida, M.: Question generation for learner centred learning. In: *IEEE 13th International Conference on Advanced Learning Technologies*, pp. 330–332. IEEE Press, Beijing (2013)
9. Paul, D.V., Naik, S.B., Rane, P., Pawar, J.D.: Use of an evolutionary approach for question paper template generation. In: *IEEE Fourth International Conference on Technology for Education*, pp. 144–148. IEEE Press, Hyderabad (2013)
10. Kale, V.M., Kiwelekar, A.W.: An Algorithm for question paper template generation in question paper generation system. In: *International Conference on Technological Advances in Electrical, Electronics and Computer Engineering*, pp. 256–261, IEEE Press, Konya (2013)
11. Al-Yahya, M.: OntoQue: A question generation engine for educational assessment based on domain ontologies. In: *11th IEEE International Conference on Advanced Learning Technologies*, pp. 393–395. IEEE Press, Athens (2011)
12. Johnson, K., Hall, T., O’Keeffe, D.: Generation of Quiz Objects (QO) with a Quiz Engine Developer (QED). In: *IEEE International Workshop on Wireless and Mobile Technologies in Education*, pp. 120–122. IEEE Press (2005)

13. Blooms Taxonomy: http://en.wikipedia.org/wiki/Bloom's_taxonomy
14. Shorser, L.: Blooms taxonomy interpreted for mathematics (1999). <http://www.math.toronto.edu/writing/BloomsTaxonomy.pdf>
15. Bloom, B.S.: Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. Longmans, Green, New York (1956)

Adaptive Output Feedback Control System Design for Low-Cost Electronic Knee Prosthesis

Salman Shaikh and Akshay Malhotra

Abstract Electronic knees provide a wide range of mobility for the amputees. These can often cover K2 and K3 grades of mobility. To achieve a wider mobility, these electronic knees depend upon feedback sensors which provide real-time data to a microcontroller. Various methods are used to calculate the joint motion and knee angle which include magnetic encoder, electro goniometer, and inertial measurement units. The goal of this paper is to develop a sensor to measure angular change of the “stump” socket using accelerometer and ultimately that of the thigh movement, and to use it as feedback signal to the microcontroller in the electronic knee. Sensing the thigh position directly can help in achieving a more natural gait. This feedback system can be used instead of the passive feedback systems currently used in the electronic knees. Results show the potential of low-cost sensing method as a reliable feedback system which can reduce complexity of the hardware as well as the algorithms currently used in the development of modern electronic knee, and hence the overall cost of the electronic knee.

Keywords Prosthesis · Microcontroller · Electronic knee · Angle measurement

1 Introduction

A major cause of lower limb amputation is injuries resulting from accidents [1, 2]. Road accidents in urban areas and agricultural and railway accidents in rural areas are the major causes of such injuries in India [2]. With there being a rapid rise in the

S. Shaikh (✉) · A. Malhotra

Department of Electronics and Telecom Engineering, SIT, Lavale, Pune, India
e-mail: salman.mit@gmail.com

A. Malhotra

e-mail: akshaymalhotra@sitpune.edu.in

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_52

number of the road accidents in India, around 5.5 lakh people needed lower limb prosthesis as of 2010 estimates [3, 4]. For transfemoral amputees, the prosthetic knee provides the lost mobility. These mobility grades are classified as K1 (indoor walker), K2 (restricted outdoor walker), K3 (unrestricted outdoor walker), and K4 (unrestricted outdoor walker with rigorous demands). Low-cost knees provide limited mobility, whereas high-cost electronic knees have been able to provide a wider range of mobility for the amputees [3]. More accurate gait, accurate swing phase, better control, and better adaptability at walking speeds were the advantages that soon became the strong point of this commercially available knee [5]. Sensors play an important role in these systems, as they provide data from the various biomechanical actions that take place in a gait cycle. Angular movement of the joints, foot pressure, and speed of the movement of limbs are the basic factors that need to be sensed and measured. These complex feedback sensors and their associated control designs make the current available electronic knees very costly and out of reach for most of the amputees. While most of the electronic knees will measure the knee angle for feedback sensing, the hip angle can actually provide a more accurate feedback that can be used in a variable damping prosthetic knee [1, 6]. The hip angle produces a periodic movement that shows a correlation with the knee angle over one gait cycle. Thus, the measurement of the hip angle for feedback can help in achieving a normal gait cycle even at variable walking speed [1, 7]. Reducing the number of sensors reduces the complexity in design and cost of the electronic knee. This paper presents a system for real-time acquisition of the stump angle and its correlation with the knee angle, thus, enabling the design of a low-cost electronic knee with K1 and K2 grade of mobility, targeted but not restricted to Indian amputees.

2 Human Gait

Figure 1 shows a single stride of the human gait cycle. The simplest model of walking comprises two legs moving in sagittal plane alone. Gait cycle consists of two phases: stance phase and swing phase. The period when both of the lower extremities are in contact with ground is called the “stance” phase, and the period when only one of the extremities is in contact with the ground is called the “swing” phase.

Stance phase is also called as “period of double support” and contributes to 60 % of the human gait cycle. This whole process leads center of mass to move in an arc and is highest at “mid stance” and lowest at “heel strike” and “toe off.”

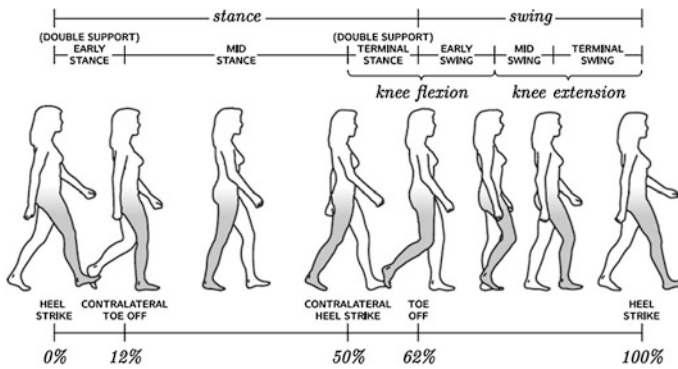


Fig. 1 Human gait cycle

3 Electronic Knee

Prosthesis industry grew steadily post-World War II. Electronic actuators and EMG signals were used to achieve a volitional control of the electronic knee. It resulted in a more esthetic gait and also provided stability over uneven terrain [7]. Attempts to improve the mobility during stance phase of level walking focused either on knee controller design or on ankle design [8]. Programmed microprocessor was used to recognize common gait patterns from the angular knee movement and strain sensors feedback [8]. The pre-set values in the microprocessor lacked the smooth gaits necessary for variable speed and walking levels of the subject [9, 10]. Biologically inspired adaptive control of knee prosthesis allowed flexion during stance [11]. A robust control system for an electronic knee which was auto-adaptive, speed adaptive, and allowed stance flexion and detection of stairs was developed [12]. Variable torque magnetorheological (MR) knee was developed giving some improvements over the electrohydraulic knee [13–15]. Local mechanical sensing allowed the amputee to walk with an increased level of biological realism as compared to mechanical passive prosthetic knee [13]. Hybrid passive-active prosthesis added capability to walk backward, stand up from a seat, ascend ramps, and stairs ramp over ramp [16]. Two series-elastic actuators along with a variable damper were used in agonist-antagonist active knee prosthesis [17]. System for onboard charging of the batteries for a longer operating duration was developed to improve the reliability of electronic knee [18]. Research continues to develop new designs of electronic knees which provide a better mobility and are less costly as compare to present commercial electronic knees.

3.1 Sensors in Electronic Knee

Sensors form the key component of any feedback system. Feedback signals from knee, ankle, hip, and feet help the control system achieve a smoother gait.

The complexity in the system design depends upon the number of sensing units used, and the algorithms involved.

3.1.1 Angle Sensor

Potentiometer, optical encoder, or magnetic encoder attached at a pivotal point measures the angle between two links directly. Indirectly, an accelerometer can be used to measure the tilt and determine an angular change at the joint. The angle sensor is used to measure the hip angle and knee angle in sagittal plane.

3.1.2 Torque Sensor

Human gait involves weight loading activity at stance phase of the gait. Torque sensor can either be bulky or expensive. Torque feedback is necessary to achieve a better stability of the gait.

3.1.3 Foot Pressure Sensor

The foot pressure/contact sensor provides heel strike and toe-off event in the gait. Like torque feedback, the feedback from foot position helps achieve a better stability during the dynamic gait cycle.

3.2 Actuators

Electronic knees use various actuators to carry out the knee action. These mainly include electrohydraulic, pneumatic, MR, or linear actuators. Electrohydraulic actuators offer better functional and safety-related advantages over the other actuators used in an electronic knee [4]. Clinical effectiveness of prosthetic component is dependent on the reliability of the different functions offered to the amputee in daily life. The cost factor of electronic knee varies greatly depending upon the type of actuator used. The research over the years can be concluded to a major point, and the sensors and the actuators used were crucial in obtaining a natural gait with wider mobility.

4 Methodology

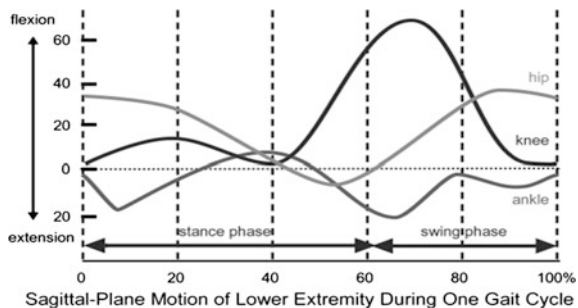
There are various electronic knees commercially available for the amputees ranging from hydraulic, pneumatic, and MR knee. These products provide wider mobility grades for the amputees. The accuracy and efficient working of these electronic

knees depend largely on the feedback systems incorporated in these designs. The complexity of the hardware and the associated algorithms make these devices highly expensive, with the costs going as high as \$45,000, whereas a basic electronic knee can cost around \$12,000 [3]. To develop a cost-effective electronic knee, the development of sensing modality should comprise minimal hardware and a less complex yet reliable algorithm.

4.1 Technology Principle

An electronic knee is basically a feedback control system with three major components: a sensor system, microcontroller unit, and an electrohydraulic actuator [19]. It helps provide automatic flexion and extension as well as damping while loading of the body weight, hence maintaining a proper and stable gait even for various walking speeds. Figure 2 shows the degree of variations of the hip, knee, and ankle joints over one gait cycle. The knee angle feedback provides information about the flexion angle of the prosthesis. The fact that in above knee amputees, the knee is absent implies that the system has to rely on indirect measurement of the knee angle in order to obtain an accurate feedback. The prosthesis thus requires more complex algorithms and depends on more than one sensor [20]. To lower the cost of an electronic knee, the less number of sensors with reliable feedback can be used to provide basic K1 and K2 mobility grades. For this, the feedback from the stump is acquired using an accelerometer. The hip/stump provides a volitional control and can easily be used to detect the thigh angle and the gait movement [1]. This would also reduce the algorithm complexity and help in developing a low-cost prototype with a fair amount of mobility. Figure 2 shows the degree of variations of the hip, knee, and ankle joints over one gait cycle. In the module design presented in this paper, thigh angle sensor is mounted on the stump externally (Fig. 3). For this, the primary criterion was to develop a robust system in a changing environment. The sensor that is discreetly mounted on the stump should provide a robust and appropriate base for the long-term reliable monitoring.

Fig. 2 Angular changes at limb joints in lower extremity over one gait cycle



4.2 Interfacing

The accelerometer (ADXL345) is capable of measuring static acceleration of gravity in tilt sensing measurement as well as the dynamic acceleration resulting from motion, shock, and vibration. It is mounted on the stump directly to detect the thigh movement in sagittal plane. The output amplitude of sensor does not need any amplification and is enough to drive the input of the microcontroller (LPC 21xx series). The built in ADC is used to further process the output from the accelerometer. Signal conditioning can include low-pass filtering to remove high-frequency noise content. The design presented herein does not include any LPF. The sensitivity required in the input measurement is kept at medium to avoid high-frequency noise generated due to human errors.

4.3 Angle Sensor

The knee joint separates the angle sensor and the actuator (Fig. 3). The microcontroller sits in the knee housing, and a flexible connector is used for connectivity with the sensor. The sensor is an accelerometer that is used to measure tilt changes occurring during the gait cycle. The flexion and extension action of the actuator (replicating the knee) can be easily achieved, as decided by an appropriate algorithm to achieve a stable gait, even at various walking speeds. The extension and flexion of the hip can thus be used to bring out the flexion and extension in the lower limb using any electronic actuator. Whenever the range of motion (ROM) exceeds the above values, actions like sitting and standing up can be predicted (Fig. 4).

Fig. 3 Sensor placement

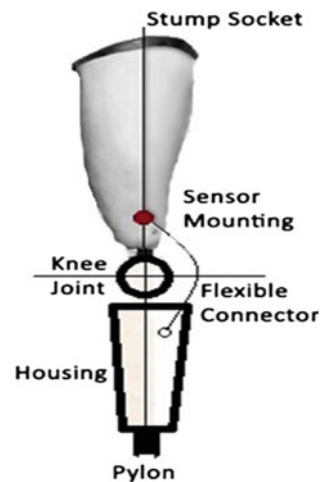
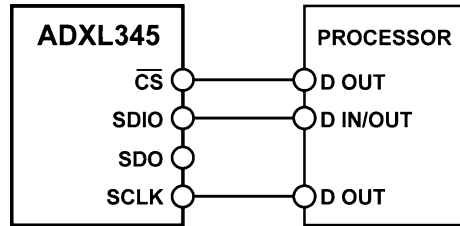


Fig. 4 Sensor interfacing



5 Properties of the System

5.1 Feedback Measurement

The feedback control system incorporated the accelerometer attached to the thigh on the garment fabric of a sound person for trial (mounted on stump in case of an amputee). The sensor was connected to the microcontroller with a flexible wire connector. The subject was then allowed to walk on the treadmill. A periodic output waveform is obtained at the output. This gives the tilt values in the two-dimensional plane, here the sagittal plane. These values are then compared with the pre-set lookup values, and a corrective action is taken. A typical movement pattern (Fig. 5) and output at static conditions (Fig. 6) of this system are shown. There was a slight noise present which can be reduced by fixing the accelerometer in the stump socket and also by using a low-pass filter. The output shows a periodic pattern and correlates well with the movement of the hip in the sagittal plane. The system shows an acceptable steady response at static gait, sitting, standing, or in squat position.

5.1.1 Circuitry Testing

The circuit was tested under static and dynamic conditions. Initial tests were performed to check for the precision and stability of the circuitry in a changing environment. For the static conditions of standing and sitting, the circuit was tested for the baseline drifts over a period of 20 s. The root-mean square error of the output signal at static conditions was calculated to be 0.045 V. For measurement under dynamic conditions, the sensor was temporarily attached to the thigh on a garment, and readings were taken. The maximum output obtained during the active gait cycle was at the heel strike. The baseline drift in dynamic environment was more but within acceptable limits. This test was repeated on the same subject at various time intervals. The RMSE value thus obtained was between 0.008 and 0.007 V. This shows that the feedback system is sensitive and can provide a reliable feedback for the electronic knee joints.

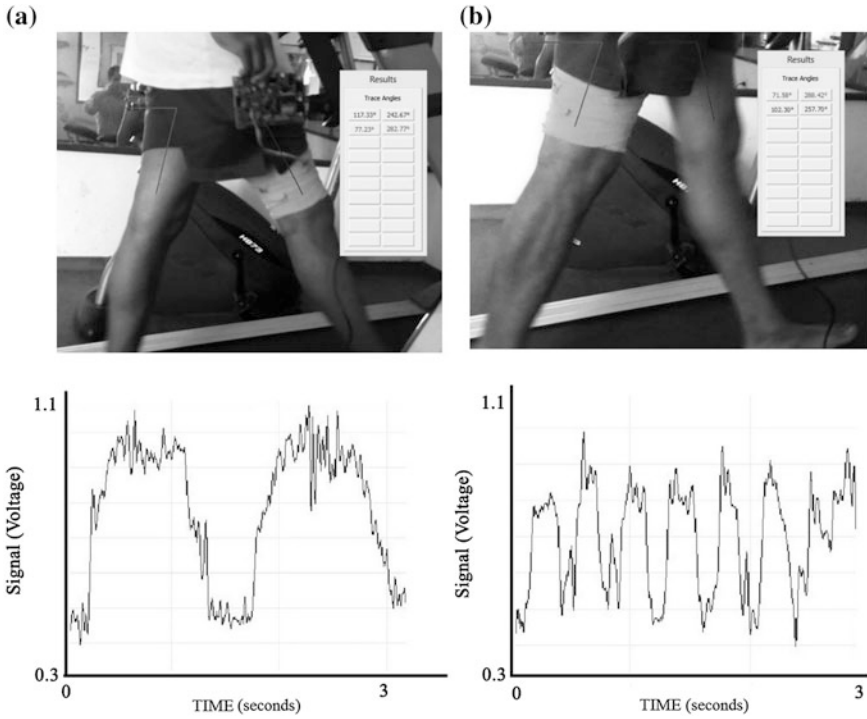


Fig. 5 Testing of the module on a normal subject. Variations in the output of feedback sensors. **a** Slow walking speed. **b** Fast walking speed

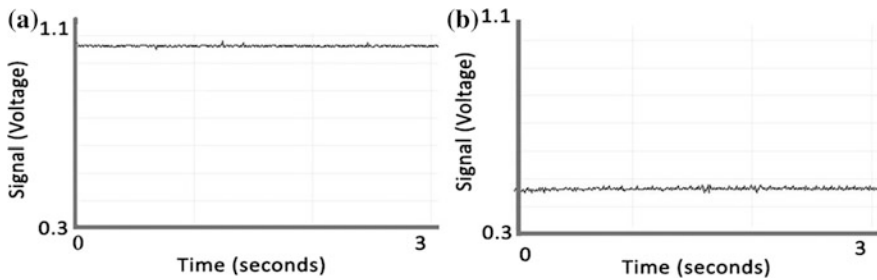


Fig. 6 Output at the static conditions of gait. **a** Standing. **b** Sitting condition output

5.1.2 Algorithm

The sensor needs one-time calibration making it easier for the trials on various subjects. The calibration scale is divided over the heel strike and toe-off period of the gait cycle. Region over the minimum and maximum voltage range output was selected for optimum output acquisition. The frequency of the change at output

helps determine the various walking speeds. The walking speed/thigh movement (S) is given by a differential equation:

$$S = dV/dt \quad (1)$$

where V is the output of the accelerometer.

Depending upon this feedback, the corrective action is taken. The above equation is used not only to predict the gait but also to provide a key feedback for upslope/downslope walking, sitting, and standing up.

6 Results

Figures 5 and 6 show the output signal that can be used as a feedback for electronic knee. The periodic pattern helps relate the position of the thigh and can be used to bring extension and flexion at the knee actuator enabling walking at various speeds.

The tilt sensor measures the movement resulting from the thigh swing in the sagittal plane. The hip joint makes a range of motion (ROM) of 38.9° . The maximum extension occurs at pre-oscillation phase at -16.2° , and maximum flexion is 22.7° , which is observed at the heel strike [1].

The development cost for this prototype of is only around \$200 and could thus be incorporated in the design of a low-cost electronic knee.

With foot pressure sensors and load sensors, this system can provide a reliable feedback system for the electronic knee helping cover basic mobility of K2 and K3 grades.

7 Discussion

This paper describes the design process for measurement of the hip angle in the sagittal plane. The goal is to develop a low-cost yet efficient and reliable sensing modality that could further be used in development of a low-cost electronic knee model that provides a wider mobility and thus bridges the economic as well as the functional gap that currently exists between high end mechanical knees and the basic electronic knee. Good initial results have been obtained using the accelerometer for the angle measurement. Offset errors were present but within the acceptable limit of 0.7 V. Focus was to obtain a clear waveform which could correlate with the natural gait movement of the thigh. The results obtained are satisfactory and can provide reliable feedback at various walking speeds. Using the value of S , the algorithm can make necessary actions and can cover a wider range of mobility grades. Clinical

setting uses the traditional methods of joint motion evaluation using goniometer to record the range of motion. But this static measurement differs a lot over the values obtained by the dynamic assessment. Although the current design lacks a signal conditioning unit, the results obtained showed less noise interference, except for the high-frequency noise component present due to the sensitivity of the accelerometer and also the fabrication/placement of sensor to some extent. A dedicated signal conditioning unit could make the signal smoother. As the system measures the hip angle, the dataset was recorded on a sound subject at various levels of walking within a laboratory environment. The next phase would be to test this system under real-world conditions.

References

1. Lambrecht, B.: Design of a hybrid passive-active prosthesis for above-knee amputees. PhD Thesis, Dept. Mech. Eng., University of California (2008)
2. Mohan, D.: A report on amputees in India. *Orthot. Prosthet.* **40**(1), 16–32 (1982)
3. Jain, S., Saroj, R.: Private communication. Prosthesis Workshop, Sancheti Hospital, Pune (2013)
4. Kumar, N., Sohni, S.: Low cost prototype development of electronic knee. *J. Sci. Ind. Res.* **69**, 444–448 (2010)
5. Bellmann, M., Schmalz, T., Blumentritt, S.: Comparative biomechanical analysis of current microprocessor-controlled prosthetic knee joints. *Clinic. Impl. Basic Res.* (2010). <http://www.ncbi.nlm.nih.gov/pubmed/20382300>
6. Ha, K.H., Atakan Varol, H.: Volitional control of a prosthetic knee using surface electromyography. *IEEE Trans. Biomed. Eng.* **58**(1) (2011)
7. Dyck, W.R.: A voluntarily controlled electrohydraulic prosthesis. *Bull. Prosthet. Res.* (1975). Spring: 169:86. www.ncbi.nlm.nih.gov/pubmed/1203614
8. Stein, J., Herr, H.: Stance phase control of AK prosthesis: Knee control Vs. SACH foot design. *J. Biomech.* **20**(1), 19–28 (1987)
9. James, K.B.: System for controlling artificial knee joint action in AK prosthesis. U.S. Patent 5571205, 5 Nov 1996
10. Zlatnik, D.: Intelligently controlled above knee prosthesis, Institute of robotics. ETH-Zurich (1997)
11. Wilkenflod, A.J.: Biologically inspired autoadaptive control of a knee prosthesis. PhD Thesis, Dept. Elect. Eng. and Comp. Science, MIT (2000)
12. Deffenbaugh, B.: Electronically controlled prosthetic knee. U.S. Patent 6764520, 20 July 2001
13. Herr, H.: User-adaptive control of a magnetorheological prosthetic knee. *Indus. Robot: Int. J.* **30**(1), 42–55 (2003)
14. Heiðar Guðmundsson, K.: Design of a magnetorheological fluid for an mr prosthetic knee actuator with an optimal geometry. PhD Thesis, Mech. Eng., University of Iceland (2011)
15. Gudmundsson, K.H., Jonsdottir, F., Olafsson, S.: Multi-objective optimization of a magneto-rheological prosthetic knee actuator. In: 19th International Conference on Adaptive Structures and Technologies, Oct 2008
16. Sup, F.: Upslope walking with a powered knee and ankle prosthesis: initial results with an amputee subject. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**(1) (2011)
17. Martinez-Villalpando, E.: Agonist-antagonist active knee prosthesis. *J. Rehabil. Res. Dev.* **46**(3) (2009)

18. Andrysek, J., Chau, G.: An electromechanical swing-phase-controlled prosthetic knee joint for conversion of physiological energy to electrical energy: feasibility study. *IEEE Trans. Biomed. Eng.* **54**(12) (2007)
19. Merkle, D.: Fields of applications of electro hydraulics. *Electro Hydraulics-Basic Level*, Esslingen, Festo Didactic (2004)
20. Zhang, F.: Towards design of a stumble detection system for artificial legs. *IEEE Trans. Neural Syst. Rehabil. Eng.* **19**(5) (2011)

Low-Leakage, Low-Power, High-Stable SRAM Cell Design

Soumitra Pal, Y. Krishna Madan and Aminul Islam

Abstract This paper proposes a technique for designing low-leakage stable SRAM cell which can mitigate impact of V_t (threshold voltage) variation. The architecture of the proposed transmission gate-based 9-transistor SRAM cell (TG9T) is almost similar to that of 7-transistor SRAM cell (7T) except the access transistors, which are replaced with transmission gates. In this study, various key design metrics like noise margin, leakage current, and hold power are simulated for both cells and compared. The proposed design provides $1.25\times$ lower leakage current and $1.46\times$ higher SINM (static current noise margin) while bearing $3.8\times$ penalty in WTI (write trip current) compared with 7T. Proposed design exhibits its robustness by achieving $1.1\times$ tighter spread in hold power compared to 7T.

Keywords MOSFET · Transmission gate · SINM · WTI · Leakage current · Hold power

1 Introduction

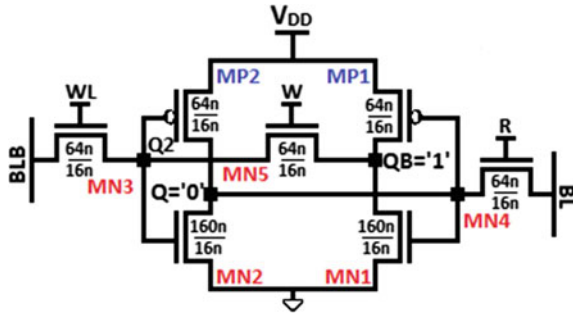
Performances of the upcoming microprocessors are being improved generally through increase in the capacity of on-chip memory. Growth in these processors speed is more than memories speed, and this gap increases by time [1]. To synchronize both of the speeds, hierarchical usage of memory levels is introduced. This means most frequently used data/instructions are placed nearer to processor such

S. Pal (✉) · Y. Krishna Madan · A. Islam
Electronics and Communication Engineering, Birla Institute of Technology,
Mesra, Ranchi 835215, Jharkhand, India
e-mail: soumitra10028.13@bitmesra.ac.in

Y. Krishna Madan
e-mail: krishna1285.11@bitmesra.ac.in

A. Islam
e-mail: aminulislam@bitmesra.ac.in

Fig. 2 7T SRAM cell



2 Proposed TG9T SRAM Cell

Proposed cell contains only one read/write port and does not require many changes when compare to 7T cell, except replacing access transistors MN3/4 in 7T cell with TGs. There are two additional controls, WLB and RB, to switch ON the PFETs (MP3/4) by applying exact complementary control signals (WL, R) of their respective NFETs (MN3/4).

Shifts in the threshold voltage (V_t), due to several fabrication parameters and RDF, are inversely proportional to square root of device area, which can be known from $\sigma_{V_t} \propto 1/\sqrt{WL}$. Here, W and L are effective channel width and length, respectively. Thus, the simple solution to improve the variability is to increase the device area. Therefore, for fair, both the cells are maintained at iso-device area, and design metrics are estimated and compared to show the improved variability of TG9T over 7T.

In this paper, 5000 Monte Carlo simulations are used while maintaining widths as shown in Fig. 1 to meet the desired robustness with lengths of all FETs set to 16 nm. Bit-line capacitances remain unchanged even though access NMOS is replaced with TG as they bear same drain diffusion area.

3 Simulation Results and Comparison

This section presents simulation results of various key design metrics such as read stability, leakage current, and hold power of both 7T and TG9T. To achieve reasonably high accuracy, Monte Carlo simulations are performed with sample size of 5000. All simulations are performed in super-threshold region with assumption of some parameters such as channel length, channel width, oxide thickness, channel doping concentration, and threshold voltage having independent Gaussian distribution with 3σ variation of 10 % [10].

3.1 Noise Margin

In read operation, SRAM cell is most vulnerable to noise, and data retention during this operation is the most important constraint. Due to the technology scaling, cell becomes less stable during lower supply voltages. Therefore, both cells are used in super-threshold region to avoid this condition. Increase in variability and increase in leakage currents are also undesirable outcomes of technology scaling which also affect the cell stability. Noise margin during read operation (read stability) is usually defined by RSNM (read static noise margin) [11] as maximum DC noise voltage that can be allowed without changing the stored bit, by the SRAM cell [12–14]. During read operation, node, which stores ‘0’ in 7T SRAM cell rises to a voltage higher than ground, because of voltage dividing effect between pulldown transistor and access transistor corresponding to ‘0’ storage node. Cell ratio ($\beta_{\text{ratio}} = \beta_{\text{pulldown}} / \beta_{\text{access}}$) is a factor on which RSNM is highly depended.

To explain the cell stability in terms of RSNM, most used technique is plotting butterfly curve. Butterfly curve can be obtained by measuring and plotting the voltages at storage nodes through placing two DC noise voltage sources in series with two inverters by removing cross-couple connection in between them [15]. In ideal case, gain of two cross-coupled inverters tends to infinite value. As a result, RSNM calculation through butterfly curves is limited to a maximum value of $0.5V_{\text{DD}}$. An additional drawback with butterfly curve approach is the requirement of additional mathematical calculations using measured data due to the absence of inline testers [16].

An alternative approach for calculating read stability is through “*N*-curve” which can be measured through inline testers [16]. *N*-curve gives information regarding both voltages and currents, which can overcome the problems of V_{DD} scaling, described for SNM. *N*-curve can be extracted (Fig. 3) by maintaining both the bit-lines at V_{DD} , and all other word lines (WL, WLB, RB, R, W) in read mode. A voltage sweep is inserted at node storing ‘0,’ and corresponding current is measured.

Characteristic *N*-curve meets ZERO line at three points A, B, and C as shown in Fig. 4. Positive peak between A and B gives static current noise margin (SINM). For better-read stability, absolute value of SINM should be high. In addition to read stability, write ability is also important for an SRAM cell. This can also be determined from the same *N*-curve in terms of WTI (write trip current). Negative Peak between B and C gives WTI. For better write ability, absolute value of the WTI should be low. Proposed TG9T achieves larger SINM value than 7T by offering penalty in WTI (see Fig. 4).

3.2 Leakage Current

Leakage current in proposed TG9T is less than that of 7T even though access TG in TG9T occupies same area as of access transistor MN4 in 7T. This can be observed

Fig. 3 Proposed TG-based 9T SRAM cell (TG9T) with setup for extracting *N*-curve during read mode

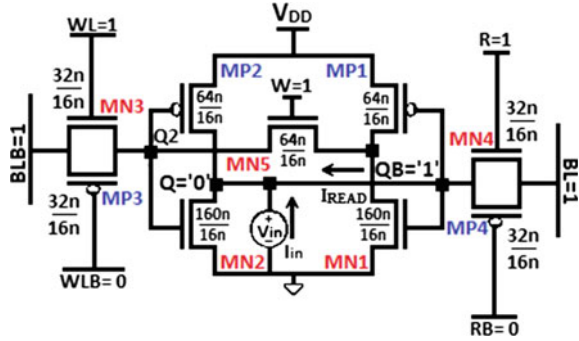
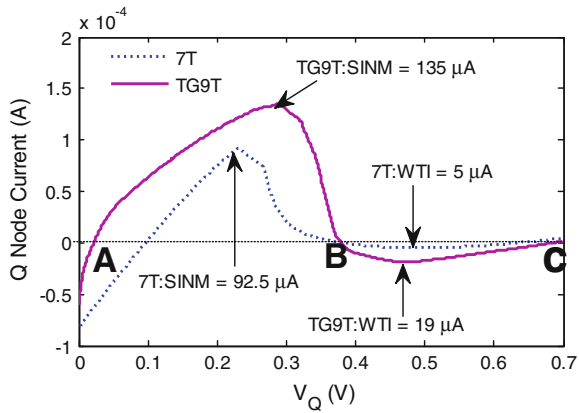


Fig. 4 Static current noise margin (SINM) and write trip current (WTI) versus internal node voltage @ nominal $V_{DD} = 0.7$ V



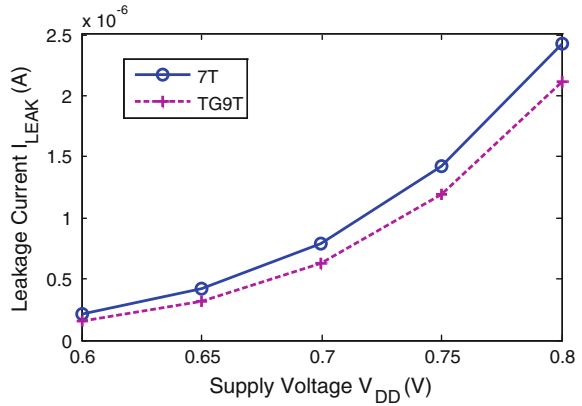
from Table 1 and Fig. 5 that at all V_{DD} , leakage current of TG9T is slightly less than the 7T.

This phenomenon can be explained by considering hot carrier injection mechanism. Hot carrier injection in solid state electronic devices is a phenomenon where a “hole” or an “electron” gains sufficient kinetic energy to overcome a potential barrier necessary to break an interface state. Leakage current is due to the current flow across gate from either drain or source in MOSFETs. Gate contains oxide interface, which is deposited directly on the silicon surface. Due to the high electric field near this Si/SiO₂ surface, electrons in NMOS and holes in PMOS can gain sufficient kinetic energy and enter oxide layer. This happens to be more likely in NMOS than in PMOS, since electrons have lower effective mass than holes and lower potential barrier (3.1 eV for electrons and 4.5 eV for holes) [17].

Lower leakage current helps achieve higher column height; in other words, it increases ability to insert more number of SRAM cells in a single column, which decreases area occupied by SRAM.

Table 1 Comparison of leakage current

Supply voltage V_{DD} (V)	Leakage current (I_{LEAK}) of TG9T (μA)	Leakage current (I_{LEAK}) of 7T (μA)
0.60	0.155 (1)	0.209 (1.35)
0.65	0.319 (1)	0.415 (1.30)
0.70	0.631 (1)	0.788 (1.25)
0.75	1.190 (1)	1.420 (1.19)
0.80	2.110 (1)	2.430 (1.15)

Fig. 5 Leakage Current (I_{LEAK}) versus supply voltage (V_{DD})

3.3 Hold Power

Hold power or Leakage power of both SRAM cells is evaluated using 5000 Monte Carlo simulations in SPICE. For easier comparison, simulated data at 0.7 V are plotted in Fig. 7. Leakage power depends on the effective width of the device, which in this case, is equally maintained for both the cells. From Fig. 6, it can be observed that, both the cells having almost same amount of hold power. It can be noticed that hold power of TG9T is slightly less than that of 7T at all respective V_{DD} (see Fig. 6). This happens due to the presence of PMOS in TG9T at larger number than in 7T, and thus, leakage current is reduced, which in turn decreases hold power in TG9T.

From Fig. 7, it can be observed that maximum value of TG9T is lesser than that of 7T, which also proves that the mean value of leakage power is improved in TG9T. The above distribution curves of TG9T and 7T cross at $1 \mu W$. Based on the available data, 14.74 % of statistical samples of 7T have more than $1 \mu W$, whereas in the case of TG9T, it is only 10.3 %.

Fig. 6 Hold power (H_{POWER}) versus supply voltage (V_{DD})

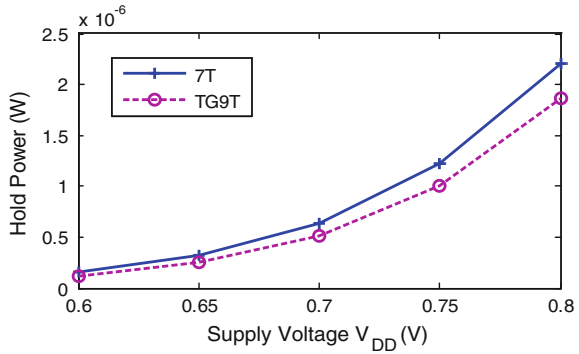
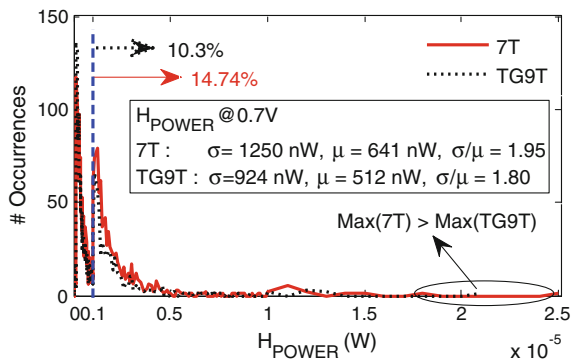


Fig. 7 Comparison of hold power distribution of TG9T and 7T @ nominal $V_{DD} = 0.7$ V



4 Conclusion

This paper proposes a transmission gate-based 9T SRAM cell. It offers high stability, low leakage current, and low power dissipation during hold mode which are key parameters for a robust SRAM cell. Due to low leakage current, the more number of SRAM cells can occupy a single column. Therefore, area occupying by the SRAM cell is also decreased. Due to the presence of TGs in proposed circuit, it mitigates impact of V_t variations and improves the variability of the design metrics.

References

1. Hennessy, J.L., Patterson, V.: Computer architecture: a quantitative approach, Chap. 5. Morgan Kaufman (2006)
2. Jain, S.K., Agarwal, P.: A low leakage and SNM free SRAM cell design in deep sub micron CMOS technology. In: 2006 19th International Conference VLSI Design, pp. 3–7 Jan 2006
3. Athe, P., Dasgupta, S.: A comparative study of 6T, 8T and 9T Decanano SRAM cell. IEEE Symposium Industrial Electronics and Applications (ISIEA), vol. 2, pp. 889–894 Oct 2009

4. Kim, T.-H., Liu, J., Kim, C.H.: A voltage scalable 0.26 V, 64 kb 8T SRAM with V_{min} lowering techniques and deep sleep mode. *IEEE J. Solid-State Circuits* **44**(6), 1785–1795 (2009)
5. Chang, L., et al.: An 8T-SRAM for variability tolerance and low-voltage operation in high-performance caches. *IEEE J. Solid-State Circuits* **43**(4), 956–963 (2008)
6. Liu, Z., Kursun, V.: Characterization of a novel nine-transistor SRAM cell. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **16**(4), 488–492 (2008)
7. Calhoun, B.H., Chandrakasan, A.: A 256-kb sub-threshold SRAM design for ultra-low-voltage operation. *IEEE J. Solid-State Circuits* **42**(3), 680–688 (2007)
8. Chang, I.J., Kim, J.-J., Park, S.P., Roy, K.: A 32 kb 10T sub-threshold SRAM array with bit-interleaving and differential read scheme in 90 nm CMOS. *IEEE J. Solid-State Circuits* **44**(2), 650–658 (2009)
9. Aly, R.E., Bayoumi, M.A.: Low-power cache design using 7T SRAM Cell. *IEEE Trans. Circuits Syst.—II: Express Briefs* **54**(4) (2007)
10. Vaddi, R., Dasgupta, S., Agarwal, R.P.: Device and circuit co-design robustness studies in the subthreshold logic for ultralow-power applications for 32 nm CMOS. *IEEE Trans. Electron Dev.* **57**, 654–664 (2010)
11. Seevinck, E., et al.: Static-noise margin analysis of MOS SRAM cells. *IEEE J. Solid-State Circuits* **SC-22**(5), 748–754 (1987)
12. Islam, A., Hasan, M., Arslan, T.: Variation resilient subthreshold SRAM cell design technique. *Int. J. Electron.* **99**(9), 1223–1237 (2012)
13. Pal, S., Bhattacharya, A., Islam, A.: Comparative study of CMOS- and FinFET-based 10T SRAM Cell in Subthreshold regime. In: *IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT)*, pp. 507–511, May 2014
14. Islam, A., Hasan, M.: Leakage characterization of 10T SRAM cell. *IEEE Trans. Electron Dev.* **59**(3), 631–638 (2012)
15. Islam, A., Hasan, M.: Variability aware low leakage reliable SRAM cell design technique. *Microelectron. Reliabil.* **52**(6) (2012)
16. Wann, C. et al.: SRAM cell design for stability methodology. In: *Proceedings IEEE VLSI-TSA*, pp. 21–22, Apr 2005
17. Taur, Y., Ning, T.H.: *Fundamentals of Modern VLSI Devices*. Cambridge University Press, New York (2009)

Optimizing Routes in Mobile Ad Hoc Networks Using Genetic Algorithm and Ant Colony Optimization

Pankaj Uttam Vidhate, R.S. Bichkar and Yogita Wankhade

Abstract A network formed with the collaboration of mobile nodes which can communicate among themselves is called as Mobile Ad hoc NETWORK (MANET). These networks are infrastructure less in nature causing the mobile nodes to act like routers and which in turn forward the packets from the source node to the destination node. Different protocols have been used to maintain connectivity between mobile nodes. Continuous movement of nodes, radio transmission and low battery power of mobile nodes can lead to break the connectivity between nodes. Thus, the performance of network may depend upon the protocol used for routing purpose. To measure quality of service (QoS) of the network, various factors can be used like ratio of packet delivery, end to end delay, control and routing overhead, and distance in terms of nodes present between the source and destination nodes. To search an optimized path between source and destination node pair, different optimization methods are applied. In this paper, the proposed algorithm uses ant colony optimization (ACO) Technique to explore most outstanding feasible paths in collaboration with genetic algorithm (GA) which assists to give globally optimal solution among feasible paths generated by the ACO. The experiments carried out use the AODV protocol with GA-API protocol in terms of packet delivery, delay required in end to end communication and energy consumption.

Keywords MANET · QoS · GA-API · ACO · AODV

P.U. Vidhate (✉) · R.S. Bichkar · Y. Wankhade
G.H. Raisoni College of Engineering & Management, Wagholi, Pune, India
e-mail: pankajvidhate@gmail.com

R.S. Bichkar
e-mail: bichkar@yahoo.com

Y. Wankhade
e-mail: yogita16.wankhade@gmail.com

1 Introduction

MANET is structured using mobile nodes which converse with each other with the help of wireless links, without relying on predefined infrastructure. Nodes use intermediary nodes to forward their messages if they are not in each other's radio propagation range. The absence of infrastructure in MANETs leads to enact the nodes as routers. The non-requirement of infrastructure at the time of deployment of network results in the use of MANETs in the situations like search and disaster rescuing operations, battlefields and crowd control. The QoS can be determined by end to end delay, control and routing overhead, ratio of packet delivery, path distance from source node to destination node. QoS can be increased by diverse techniques like reservation of resources, minimization in delay to deliver packets, minimizing probability of link failure, availing substitute paths in case of link or node failure.

API algorithm is a special case of ACO which is instigated by the behavior of ants called as 'apicalis' ants. API algorithm has been used in this paper to search the feasible paths, in assistance with genetic algorithm (GA) which in turn imparts globally optimal path in between source node and destination node, amongst the paths provided by API algorithm.

2 Related Work

The QoS is an important aspect with respect to routing algorithms, but finding out the shortest path with number of performance metrics is an NP-complete problem. Barolli et al. [1] have applied GA for QoS routing in MANETs. They have done comparative study between GA based routing algorithms like GAMAN and GLBR.

Mutation and crossover, which are the well known operators of GA, have been used by Sharma and Sinha [2] to find the efficient and shortest routing algorithm in Mobile ad hoc networks (MANET). Mutation is used to preserve and introduce diversity in the solutions whereas crossover used to generate the new population while finding out shortest routing path. QoS routing protocol was designed for MANET using GA that employed specialized encoding, initialization, crossovers, mutations, fitness functions, selections and searching of route to improve the structure of routing protocols in MANETs [3].

Genetic algorithm is used for selection and ranking of all feasible routes in MANET where ranking depends upon priority code of links, which is calculated by routing protocols with the help of different parameters and metrics [4].

Abdullah et al. [5] described the QoS routing scheme that selects route between source-destination pair of nodes using GA. The node has to fulfill the node bandwidth, end to end delay, node delay as well as node connectivity index. It described different GA operations like initialization of population, fitness calculation, crossover process, mutation and selection used to select QoS route. Genetic

based routing approach is used to optimize the routing in MANETs to improve QoS of routing to generate optimized path [6].

Sardar et al [7] proposed On Demand QoS routing algorithm for MANET using ant colony optimization (ACO) which is based on Swarm Intelligence. The algorithm gives effective routes for communication in MANETs with the introduction of new QoS parameters like link stability time, available buffer size with respect to ACO.

Kolavali and Bhatnagar [8] studied about an application on a multi-stage shortest path problem by proposing four variants of a multi-timescale algorithm for ACO. Also they have studied the performance of different algorithms in the same framework.

To provide support related to QoS, measures like delay, hop-count, energy, bandwidth etc. are essential to face the problems and to get capable for real time applications for MANET. It is required for routing protocols to integrate with QoS metrics for finding routes and their maintenance to support QoS. Sensarma and Majumder [9] have studied concerns and challenges involved with existing QoS aware routing protocols and use of ACO for QoS aware routing in MANETs. Also they did comparative study of different existing ACO based QoS aware routing schemes to examine the strengths and weaknesses of these routing schemes.

These algorithms suggested in the literature provide good performance usually for small MANETs. However, their performance is degraded while working with large MANETs.

3 Proposed Approach for Genetic Algorithm with ACO

The proposed system consists of the modules like API—an ACO searching module, GA-optimum path search module and Combination of these two to select the best optimum path for transferring data.

3.1 Data Understanding

The main purpose of the proposed system is to find out an optimal route from source node to destination node. The MANET is characterized as a graph in which nodes are represented as vertices and the communication links as edges. In this work, different attributes are considered like node speed, energy of nodes, memory at node required to store path information as well neighboring node information (Fig. 1).

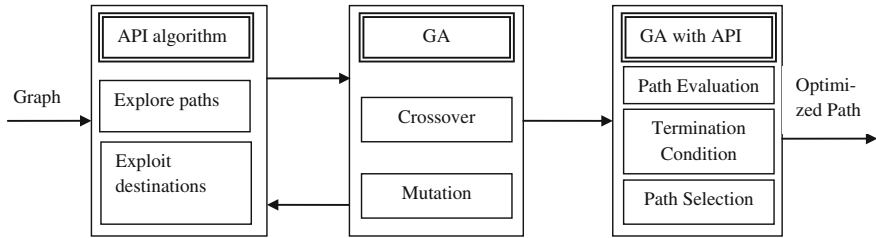


Fig. 1 System architecture

3.2 API Search Algorithm

The Graph is used to explore the paths available from source to destination node. This algorithm helps in generating initial population required for GA. Step by step description of the API algorithm is as follows:

1. Initialization: set the algorithm parameters.
2. Choose source node and destination node.
3. Exploration:
 - a. Find out the possible paths between source node and destination node with beacon messages.
 - b. If the number of paths explored is less than the predefined number then exploit the new paths.
 - c. Else disregard the destination node.
4. Termination condition:
 - a. If unable to find the path in between source and destination node after predefined number of consecutive failures then stop searching for next path.

3.3 Genetic Algorithm

A step by step GA is as follows:

1. Initialize the population:

$P_i = P(s_n, d_n)^n$, where p_i is the individual i of the population, with $i = 1, \dots, m$, and P is a path from source node to destination node in the range $[s_n, d_n]$, bounded by n dimension, and m is the number of potential parents or it can be population size.
2. Determination of fitness score of each individual in the population and select parents according to the fitness score. The individual with higher fitness score are selected as parents.

$\Theta(p_i) = f(p_i)$, where $\Theta(p_i)$ is the fitness score of the individual p_i . f indicates the fitness function.

where f can be calculated by,

$$f(p_i) = \alpha P_i(E) + \beta P_i(H_{\text{count}}). \tag{1}$$

where,

α and β are the normalizing factors ranging from 0 to 1

$$P_i(E) = 1 - \left(\frac{\max(E(m_1, \dots, m_n)) - \text{cur}(E(m_s, \dots, m_d))}{\max(E(m_1, \dots, m_n))} \right) \tag{2}$$

$\max(E(m_1, \dots, m_n))$ indicates maximum energy available in all nodes

$\text{cur}(E(m_s, \dots, m_d))$ indicates available energy in the nodes between source and destination

$$P_i(H_{\text{count}}) = \left(\frac{\sum_{i=1}^n N - \text{cur}(A(s, \dots, d))}{\sum_{i=1}^n N} \right) \tag{3}$$

$\text{Total}(N(1, \dots, n))$ gives total number of nodes in the network

$\text{cur}(A(s, \dots, d))$ gives available number of nodes in between source and destination

3. Selection of parents as per their fitness scores.

The individuals who scored less through fitness function will be selected as parents. As energy need to be maximized and the number of nodes in between source and destination node need to be minimized.

4. Variance Assignment.

a. Apply crossover operator on selected parents.

- i. List up the set of nodes say N_c in both $p1$ and $p2$ paths which were selected as parents. N_c should exclude source and destination nodes. N_c contains potential crossing sites.
- ii. Select node j as a crossing site for the list of N_c .
- iii. Do crossover the paths by exchanging all the nodes after the crossing site j .

An overview of the crossover operator is shown in the Fig. 2, in which crossover operator is applied on paths $p1$ and $p2$ at the crossing site 4 then the new offspring are generated after exchanging the subroutes.

b. Apply mutation operator on selected parents.

Path p1	2	6	4	3	8	
Path p2	2	7	3	9	8	

Child p1	2	6	4	3	9	8
Child p2	2	7	3	8		

Fig. 2 Example of crossover

Mutation operator is applied on the randomly selected solution from the population, which does small random changes in the solution. The node which is used for mutation is called as mutation node.

- i. Select the mutation node m_i from nodes in the parent path randomly.
 - ii. Select a node n_j from the neighbors of mutation node m_i .
 - iii. Generate a random from source node to node n_j and from node n_j to destination node.
 - iv. If there is a duplication of nodes in the newly created offspring path then discard that path, and do not perform mutation on it. Otherwise these two routes are connected with each other to form a new mutated path.
5. Determine the fitness score for each variance.
 6. Rank the solution in the descending order of their fitness score.
 7. Repeat: Go to the step 3 until pre-specified time exhausted or an acceptable solution is found.

3.4 Genetic Algorithm with API Algorithm

The proposed GA with API algorithm, combines the results of API algorithm and the GA [10]. This algorithm works for continuous improvement in the solution. When GA is in passive state, API algorithm searches for the available paths and when API algorithm is in passive state, it provides the explored paths to the GA to make its population.

3.5 AODV Protocol

The standard protocol used for comparing the proposed protocol is AODV (Ad Hoc On Demand Distance Vector) routing protocol which is reactive type of routing protocol. In this protocol routes are set up on-demand, as and when needed. The advantage of this protocol is that routing overhead is decreased and the disadvantage is a possible delay required to acquire desired route.

4 Comparison Between AODV and GA with API

4.1 Success Ratio

Success Ratio is the percentage of success in finding the routes between specified source node and destination node. The Graph shows the performance of the packet delivery from source to destination nodes with the increasing number of nodes from 30 to 150 nodes. With the least number of nodes all the algorithms works same. As the number of nodes increases in the network the performance of AODV decreases as compared to the GA-API and GA-API with Energy (Fig. 3).

4.2 Packet Delay

Packet delay is the time required for the packet to get delivered at the destination node from the source node. As the number of nodes increase in the network, the time required for the packet to reach from source node to the destination node also increases as of AODV is concerned. In contrast packet delay for GA-API and GA-API energy is initially (when number of nodes in the network are less in number) more as compared to AODV and as the network size increases the packet delay increases gradually but not such as AODV (Fig. 4).

4.3 Energy Consumption

Energy consumption gives the value of average energy consumed by nodes which takes part in finding the route between source node and destination node. As the network size increases AODV performance gets worst. AODV consumes more energy for finding optimized route in densely populated network. GA-API energy algorithm gives good result (optimized path) with least consumption of energy as compared to GA-API and AODV (Fig. 5).

4.4 Performance Analysis

AODV gives optimized path when the network size is small but as the network size increases the performance of AODV gets inferior. GA-API works well when the size of the network is huge but it consumes more energy to find the optimized path than that of GA-API energy because the energy parameter is considered in that algorithm.

Fig. 3 Success ratio

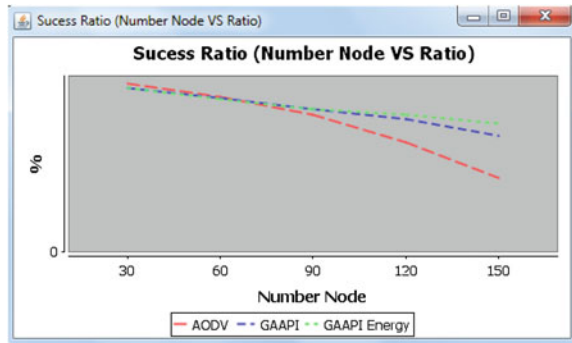


Fig. 4 Packet delay

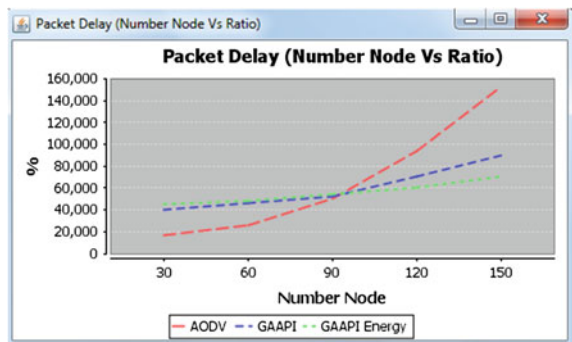
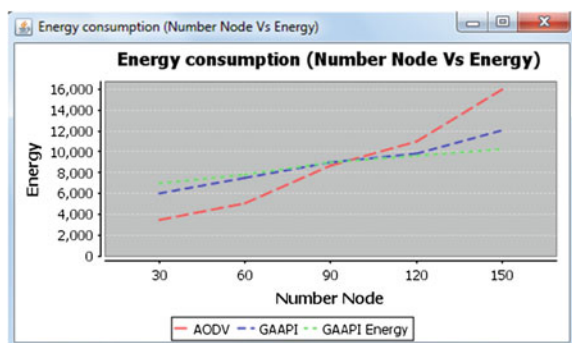


Fig. 5 Energy consumption



5 Conclusion

Our proposed technique uses GA along with ACO technique which results in providing the optimized path in between source node and destination node with minimum efforts as per the energy concerned and with minimum packet delay when the size of network is huge. The proposed algorithm works well when the networks are large and complex with greater dimensions.

References

1. Barolli, A., Takizawa, M., Khafa F.: Application of genetic algorithms for QoS routing in mobile ad hoc networks., pp. 250–259. IEEE (BWCCA) (2010)
2. Sharma, A., Sinha, M.: Influence of crossover and mutation on the behavior of genetic algorithms in mobile ad-hoc networks. In: International Conference on Computing for Sustainable Global Development, pp. 895–899. IEEE (2014)
3. Abdullah, J.: Multi-objectives GA-based QOS routing protocol for mobile ad hoc network. In: IJGDC, pp. 57–68 (2010)
4. Afridi, M.I.: Selection and ranking of optimal routes through genetic algorithm in a cognitive routing system for mobile ad hoc network. In: ISCID, pp. 507–510. IEEE (2012)
5. Abdullah, J., Ismail, M.Y., Cholan, N.A., Hamzah, S.A.: GA-based QoS route selection algorithm for mobile ad-hoc networks. In: NCTT-MCP, pp. 339–343. IEEE (2008)
6. Siwach, V., Singh, Y., Seema, Barak, D.: An approach to optimize QoS routing protocol using genetic algorithm in MANET. In: IJCSMS, pp. 149–153 (2012)
7. Sardar, A., Singh, A., Sahoo, R., Majumder, R., Sing, R., Sarkar, R.: An efficient ant colony based routing algorithm for better quality of services in MANET. In: ICT & CI, pp. 233–240. Springer (2014)
8. Kolavali, S., Bhatnagar, S.: Ant colony optimization algorithms for shortest path problems. In: NET-COOP, pp. 37–44. Springer (2008)
9. Sensarma, D., Majumder, K.: A comparative analysis of the ant based systems for QoS routing in MANET. In: International Conference, SNDS, pp. 485–496. Springer (2012)
10. Yang, S., Cheng, H., Wang, F.: Genetic Algorithms With immigrants and memory schemes for dynamic shortest path routing problems in mobile ad hoc networks. In: Transactions on Systems, Man, and Cybernetics, pp. 52–63. IEEE (2010)

Concatenation Technique for Extracted Arabic Characters for Efficient Content-based Indexing and Searching

Abdul Khader Jilani Saudagar and Habeeb Vulla Mohammed

Abstract This research paper demonstrates the work accomplished in the last phase of the ongoing research project with an objective of developing a system for moving Arabic video text extraction for efficient content-based indexing and searching. The novelty of this paper is the technique used for concatenation of the individual stand alone Arabic characters which are extracted and recognized from image frames. Unicode format of Arabic characters is used for concatenation of extracted characters which is never done before. The concatenated characters are written into the text file in incessant way. This text files are indexed using Lucene and search for the desired string is done in a faster and precise manner.

Keywords Concatenation · Extraction · Recognition · Indexing and searching

1 Introduction

Arabic language is the most widely used language in the world. Plenty of Arabic literature is available in the form of images containing essential information. Apart from images the news videos display a lot of flash news in the form of scrolling text which are either superimposed or embedded into video clips. This text needs to be extracted and amass by appropriate indexing and searching techniques for future use. The work carried out in [1] efficiently splits the news video into image frames for the purpose of extraction of scrolling text. The Arabic text from the image frames is extracted adeptly with high percentage of accuracy in recognition of Arabic characters [2].

A.K.J. Saudagar (✉) · H.V. Mohammed
College of Computer and Information Sciences, Al Imam Mohammad Ibn Saud Islamic University (IMSIU), Airport Road, Riyadh 13318, Saudi Arabia
e-mail: saudagar_jilani@ccis.imamu.edu.sa

H.V. Mohammed
e-mail: habeebvulla@ccis.imamu.edu.sa

The extracted characters need to be concatenated to form strings and sentences before writing them to text files. Few works in the past demonstrate the concatenation of extracted Arabic characters but each one has its limitations. Kashidahs' cut point center of gravity method [3] uses the chosen segmented characters are aligned in the concatenation step. The non-connectable characters are added with small white space and the new image is saved in the files. One of the traditional way for handwritten word recognition is by using Hidden Markov Model HMMs [4]. The word is constituted by concatenation of HMMs characters. This way of concatenation is used for Amharic language, the national language of Ethiopia. The work in [5] is related with issues in Arabic orthography during the process of morphology analysis and part-of-speech tagging. A grouping algorithm [6] uses one connected character at an instant and tries to form a group from existing groups. A likely merge is found when the connected character and a group are in close proximity to each other. The main focus of this paper is to present a technique for concatenation of extracted Arabic characters.

As data repositories are packed with plenty of files, it is difficult to obtain the required information. Proper indexing techniques are necessary in order to get the necessary information in swift way. Different methods for information recuperation and content scrutiny in natural language are proposed in the past and one such method is stem-based extensible method [7]. For Arabic word shape indexing an Arabic word descriptor (AWD) is proposed in [8]. An application [9] for indexing Arabic legal language in Lebanon is developed based on N-grams. In order to automatically create and index books written in Arabic language a method is proposed in [10, 11]. It completely depends on text summarization and abstraction processes for collecting main topics and statements. An approach to exploit a lexical resource (Arabic Wordnet) for indexing the documents is designed in [12]. Lexicon-free system [13] is developed for executing text queries on offline printed and handwritten Arabic documents for indexing and searching. A multi-linear term-phrasing practice to take out notion descriptors for automatic indexing, followed by notion space of allied descriptors are produced using a co-occurrence analysis is explained in [14]. AltaVista's personal version [15] is used for creating and indexing Arabic HTML records. A semi-automatic approach is proposed in [16] for indexing and searching old Arabic documents without resorting to recognize their contents. An approximate string matching technique based on Levenshtein distance for indexing and searching corrupted document images without identifying the textual molds is attempted in [17]. A system for Arabic information retrieval ARABIRS [18], and a report on TREC-11 is discussed in detail in [19]. Alternative choices are explored in [20] for indexing terms using existing electronic text collection and a newly developed collection of printed Arabic documents. New techniques, for organizing and compressing inverted index structures [21], morpho-semantic method [22] based on semantic links of the morphological forms and concept-based information retrieval [23] are explained by different authors for indexing and searching purposes.

Searching techniques in Holy Quran based on text-based, stem-based, synonyms-based, and regular expressions are discussed in [24, 25]. Dynamic Time

Warping (DTW) and Hidden Markov Models (HMM) are the two algorithms proposed in [26] for searching words of interest in Arabic handwritten and antique documents. A versatile system known as CEDARABIC is illustrated for searching a repository of scanned handwritten Arabic documents in [27]. String matching [28] is performed on disjoint text substrings of span q , during search time, fraction of the text is filtered out for similar occurrence of the pattern.

Regardless of lots of research undergone for concatenation of extracted characters in the process of archiving of image text, the problem of concatenation of Arabic text, and searching for the text of interest, with maximum accuracy is still a limitation and, very few of them have handled Arabic script.

2 Methodology

There are 28 characters in the Arabic language. Each character is written in any one of the four forms (Stand alone, Initial, Medial, Terminal). Each form is different from the another in calligraphy. Few characters cannot be represented in all the four forms. Figure 1 shows the Arabic character set and those characters which cannot be written in all forms.

In this project after successfully splitting the Arabic news video into frames [1], the scrolling Arabic text is extracted and recognized [2]. The extracted characters are individual, separated and are in any one of the four forms. These characters need to be concatenated, form sentences and should be written into text files for further indexing process. Considering a sample sentence as shown in the below Fig. 2. The recognized individual characters based on their serial number in Fig. 1, its corresponding image is written into the array using `BufferedImage` class, starting from $n - 1$ position. If the white space between one character and another character is greater than five pixels then it is considered as one blank space and a null character is inserted into the array [2] as shown in Fig. 3.

Using Java programming every character in the array is replaced by their corresponding Unicode format as shown in the Fig. 4 from Fig. 5.

The Unicode type of concatenation gives accurate string, sentence formation. This approach of concatenation has never implemented before especially for Arabic characters.

Upon concatenation the sentence is written to text file as shown in Fig. 6a which is created in the folder for the first time with system date as the name of the file as shown in Fig. 6b. All the sentences formed on that particular day are appended into the same text file. A new text file is created for writing the extracted text from frames after concatenation on the following day. All the text files which are created are indexed using Lucene indexing library [29].

The whole process is as shown below with the help of block diagram (Fig. 7).

S.No	Standalone	Terminal	Medial	Initial
1	ا	آ		
2	ب	ب	ب	ب
3	ت	ت	ت	ت
4	ث	ث	ث	ث
5	ج	ج	ج	ج
6	ح	ح	ح	ح
7	خ	خ	خ	خ
8	د	د		
9	ذ	ذ		
10	ر	ر		
11	ز	ز		
12	س	س	س	س
13	ش	ش	ش	ش
14	ص	ص	ص	ص
15	ض	ض	ض	ض
16	ط	ط	ط	ط
17	ظ	ظ	ظ	ظ
18	ع	ع	ع	ع
19	غ	غ	غ	غ
20	ف	ف	ف	ف
21	ق	ق	ق	ق
22	ك	ك	ك	ك
23	ل	ل	ل	ل
24	م	م	م	م
25	ن	ن	ن	ن
26	ه	ه	ه	ه
27	و	و		
28	ي	ي	ي	ي

Fig. 1 Arabic character set

جامعة الإمام محمد بن سعود الإسلامية

Fig. 2 Sample sentence

ج ا م عة ا إ م ا م م ح م د ب ن س ع و د ا إ س ل ا م ية

Fig. 3 Array with individual characters

062F	06A4	06A5	0627+06A4	0633	0623+06A4	0627	062F	0643	0639	0633	0640	0628	062F	0643	062D	0643	0643	0627	0643	0623+06A4	0627	0629	0639	0643	0627	062C
------	------	------	-----------	------	-----------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----------	------	------	------	------	------	------

Fig. 4 Array in unicode format

Arabic

ا	س	ع	ص	□	□	□	□	□	□	□	ف	،	ر	ه	ح
0600	0601	0602	0603	0604	0605	0606	0607	0608	0609	060A	060B	060C	060D	060E	060F
س	م	ن	د	ط	□	□	□	□	□	؛	□	ء	؟		
0610	0611	0612	0613	0614	0615	0616	0617	0618	0619	061A	061B	061C	061D	061E	061F
□	ء	آ	أ	ؤ	إ	ئ	ا	ب	ة	ت	ث	ج	ح	خ	د
0620	0621	0622	0623	0624	0625	0626	0627	0628	0629	062A	062B	062C	062D	062E	062F
ذ	ر	ز	س	ش	ص	ض	ط	ظ	ع	غ	□	□	□	□	□
0630	0631	0632	0633	0634	0635	0636	0637	0638	0639	063A	063B	063C	063D	063E	063F
-	ف	ق	ك	ل	م	ن	ه	و	ى	ي	ء	ء	ء	ء	ء
0640	0641	0642	0643	0644	0645	0646	0647	0648	0649	064A	064B	064C	064D	064E	064F
ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	ء	□
0650	0651	0652	0653	0654	0655	0656	0657	0658	0659	065A	065B	065C	065D	065E	065F
٠	١	٢	٣	٤	٥	٦	٧	٨	٩	٪	،	،	*	ب	و
0660	0661	0662	0663	0664	0665	0666	0667	0668	0669	066A	066B	066C	066D	066E	066F
ا	أ	أ	إ	ء	أ	ؤ	ئ	ى	ث	ث	ب	ت	ت	ب	ت
0670	0671	0672	0673	0674	0675	0676	0677	0678	0679	067A	067B	067C	067D	067E	067F
پ	خ	خ	ج	ج	خ	چ	چ	ڈ	د	د	ڈ	ت	د	ڈ	ڈ
0680	0681	0682	0683	0684	0685	0686	0687	0688	0689	068A	068B	068C	068D	068E	068F
ڈ	ژ	ز	ر	ر	ر	ر	ر	ز	ژ	ژ	پس	پس	پس	ص	ظ
0690	0691	0692	0693	0694	0695	0696	0697	0698	0699	069A	069B	069C	069D	069E	069F
غ	ف	ف	ف	ف	ف	ف	ق	ق	ك	ك	ك	ك	ك	ك	ك
06A0	06A1	06A2	06A3	06A4	06A5	06A6	06A7	06A8	06A9	06AA	06AB	06AC	06AD	06AE	06AF
گ	گ	گ	گ	گ	ل	ل	ل	ل	ن	ن	ن	ن	ن	ه	چ
06B0	06B1	06B2	06B3	06B4	06B5	06B6	06B7	06B8	06B9	06BA	06BB	06BC	06BD	06BE	06BF
ه	ه	ه	ه	و	و	و	و	و	و	و	ى	ى	ى	ى	و
06C0	06C1	06C2	06C3	06C4	06C5	06C6	06C7	06C8	06C9	06CA	06CB	06CC	06CD	06CE	06CF
ي	ي	ے	ے	ہ	ط	ق	م	لا	ح	ء	س	○	○	ء	ء
06D0	06D1	06D2	06D3	06D4	06D5	06D6	06D7	06D8	06D9	06DA	06DB	06DC	06DD	06DE	06DF
ء	ء	ء	س	ء	ء	ء	ء	ن	ﷻ	ء	ء	ء	ء	ذ	ز
06E0	06E1	06E2	06E3	06E4	06E5	06E6	06E7	06E8	06E9	06EA	06EB	06EC	06ED	06EE	06EF
٠	١	٢	٣	٤	٥	٦	٧	٨	٩	پس	ض	غ	ء	چ	ھ
06F0	06F1	06F2	06F3	06F4	06F5	06F6	06F7	06F8	06F9	06FA	06FB	06FC	06FD	06FE	06FF

Fig. 5 Arabic unicode chart

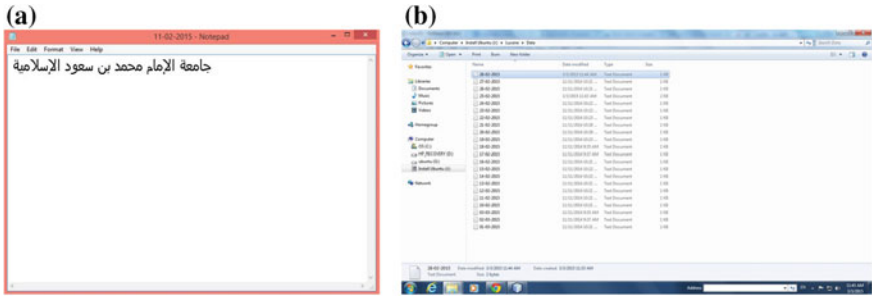
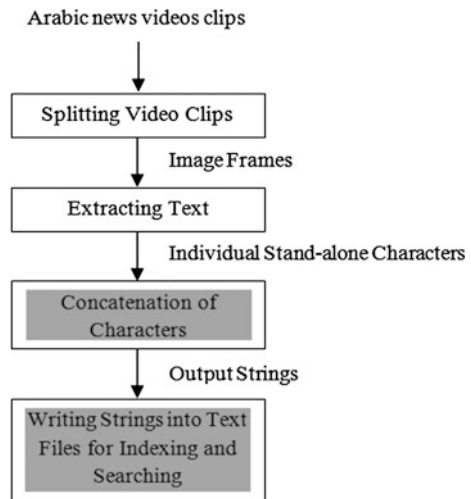


Fig. 6 a Text file containing sentence. b Text files in folder

Fig. 7 Block diagram



3 Results

This work is deployed on Java platform using NetBeans IDE 8.0.1 and tested on 50 video clips split into image frames containing Arabic sentences. The coding, execution, and testing are carried on Intel i7×64 processor with 16 GB RAM and 1 TB hard disk. On total 22 files were created on different dates and are indexed by Lucene. A text based searching approach is used for searching a particular string, for example, “العالم”, the number of its occurrences, file details are known as shown in Fig. 8.

The correctness of the concatenated strings is evaluated using the equation as shown below.

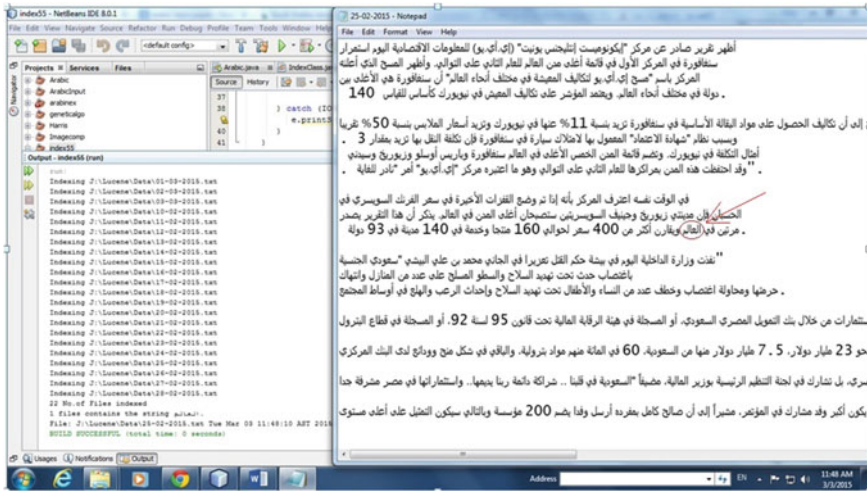


Fig. 8 Output screen

$$\text{Precision} = \frac{\text{Number of true concatenated strings}}{\text{Total number of strings}} \tag{1}$$

4 Conclusion

The proposed concatenation technique yields results with an accuracy of 98.5 % in amalgamating individual, stand alone extracted Arabic characters in string, sentence formation. The proposed technique can be applied to other languages like Urdu, Hindi, Persian, Chinese etc.

Acknowledgments This research is supported by King Abdulaziz City for Science and Technology (KACST), Saudi Arabia, vide grant no. AT-32-87.

References

1. Saudagar, A.K.J., Mohammed, H.V.: A comparative study of video splitting techniques. In: 23rd International Conference on Systems Engineering, pp. 783–788. Springer International Publishing, Switzerland (2015)
2. Saudagar, A.K.J., Mohammed, H.V., Iqbal, K., Gyani, Y.J.: Efficient Arabic text extraction and recognition using thinning and dataset comparison technique. In: International Conference on Communication, Information & Computing Technology, pp. 1–5. IEEE Press, New York (2015)

3. Elarian, Y.S., Al-Muhtaseb, H.A., Ghouti, L.M.: Arabic handwriting synthesis. In: 1st International Workshop on Frontiers in Arabic Handwriting Recognition. <http://hdl.handle.net/2003/27562> (2010)
4. Assabie, Y., Bigun, J.: HMM-based handwritten amharic word recognition with feature concatenation. In: 10th International Conference on Document Analysis and Recognition, pp. 961–965. IEEE Press, New York (2009)
5. Buckwalter, T.: Issues in Arabic orthography and morphology analysis. In: Workshop on Computational Approaches to Arabic Script-based Languages, pp. 31–34 (2004)
6. Amin, A.: Recognition of printed Arabic text based on global features and decision tree learning techniques. *Pattern Recogn.* **33**, 1309–1323 (2000)
7. Harmanani, H., Keirouz, W., Raheel, S.: A rule-based extensible stemmer for information retrieval with application to Arabic. *Int. Arab. J. Inf. Techn.* **3**, 265–272 (2006)
8. Chherawala, Y., Cheriet, M.: Arabic word descriptor for handwritten word indexing and lexicon reduction. *Pattern Recogn.* **47**, 3477–3486 (2014)
9. Mahmoud, R., Majed, S.: Improving Arabic information retrieval system using n-gram method. *WSEAS Trans. Comput.* **10**, 125–133 (2011)
10. Al-Molijy, A., Hmeidi, I., Alsmadi, I.: Indexing of Arabic documents automatically based on lexical analysis. *Int. J. Nat. Lang. Comput.* **1**, 1–8 (2012)
11. Wedyan, M., Alhadidi, B., Alrabea, A.: The effect of using a thesaurus in Arabic information retrieval system. *Int. J. Comput. Sci. Issues* **9**, 431–435 (2012)
12. Abderrahim, M.A., Abderrahim, M.E.A., Chikh, M.A.: Using Arabic wordnet for semantic indexation in information retrieval system. <http://arxiv.org/ftp/arxiv/papers/1306/1306.2499.pdf>
13. Chan, J., Ziftci, C., Forsyth, D.: Searching off-line arabic documents. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1455–1462. IEEE Press, New York (2006)
14. Lin, C.H., Chen, H.: An automatic indexing and neural network approach to concept retrieval and classification of multilingual (Chinese-English) documents. *IEEE T. Syst. Man. Cy.* **26**, 75–88 (1996)
15. Moukdad, H., Large, A.: Information retrieval from full-text arabic databases: can search engines designed for English do the job? *Libri.* **51**, 63–74 (2001)
16. Kefali, A., Chemmam, C.: A semi-automatic approach of old arabic documents indexing. http://ceur-ws.org/Vol-825/paper_83.pdf
17. Sari, T., Kefali, A.: A search engine for Arabic documents. <https://hal.archives-ouvertes.fr/hal-00334402/document>
18. Yacine, E.Y.: Towards an Arabic web-based information retrieval system (ARABIRS): stemming to indexing. *Int. J. Comput. Appl.* **109**, 16–21 (2015)
19. Savoy, J., Rasolofo, Y.: Report on the TREC-11 experiment: Arabic, named page and topic distillation searches. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.13.8419&rep=rep1&type=pdf>
20. Darwish, K., Oard, D.W.: Term selection for searching printed Arabic. In: 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 261–268 (2002)
21. He, J., Yan, H., Suel, T.: Compact full-text indexing of versioned document collections. In: 18th ACM conference on Information and Knowledge Management, pp. 415–424 (2009)
22. Al-Tayyar, M.S.: Arabic information retrieval system based on morphological analysis (AIRSMA). <https://www.dora.dmu.ac.uk/bitstream/handle/2086/4126/DX221482.pdf?sequence=1>
23. Mazari, A.C., Aliane, H., Alimazighi, Z.: A conceptual indexing approach for Arabic texts. In: ACS International Conference on Computer Systems and Applications (AICCSA), p. 1. IEEE Press, New York (2013)
24. Al-Taani, A.T., Al-Gharaibeh, A.M.: Searching concepts and keywords in the HolyQuran. http://www.nauss.edu.sa/En/DigitalLibrary/Researches/Documents/2011/articles_2011_3088.pdf

25. Arara, A., Smeda, A., Ellabib, I.: Searching and analyzing Arabic text using regular expressions e–Quran case study. *Int. J. Comput. Sci. Electron. Eng.* **1**, 627–631 (2013)
26. Saabni, R., El-Sana, J.: Keyword searching for Arabic handwritten documents. <http://www.iapr-tc11.org/archive/icfhr2008/Proceedings/papers/cr1134.pdf>
27. Srihari, S.N., Ball, G.R., Srinivasan, H.: Versatile search of scanned Arabic handwriting. In: *Arabic and Chinese Handwriting Recognition*. LNCS, vol. 4768, pp. 57–69. Springer, Heidelberg (2008)
28. Navarro, G., Sutinen, E., Tanninen, J., Tarhio, J.: Indexing text with approximate q-Grams. In: *Combinatorial Pattern Matching*. LNCS, vol. 1848, pp. 350–363. Springer, Heidelberg (2000)
29. Lucene: <http://lucene.apache.org/core/>

Harmonic Elimination Using DSTATCOM

Shabana Urooj and Pikasha Sharma

Abstract There has been an increasing awareness admits customers and utilities regarding superior quality and reliable electric power. The awareness has been triggered by a stupendous increase in the number of nonlinear loads such as power electronic devices. Due to nonlinear loads and low power factor the gravest power quality problems today are voltage sags and swells and harmonic distortion. Therefore, devices like DSTATCOMs are extensively used in power system applications and are able to provide efficient reactive power support to maintain voltage stability. Simulation of DSTATCOM is performed using simpowersystems. Analysis of instantaneous reactive power theory and instantaneous active and reactive current techniques is done which are used for extracting the reference current in the control of the DSTATCOM. These compensation techniques recuperate the ability of DSTATCOM to sustain voltage stability in steady-state and transient conditions. Investigation of the simulation results for both the techniques is done.

Keywords Distribution static compensator (DSTATCOM) · Harmonics elimination · Instantaneous reactive power theory · Instantaneous active and reactive current theory

1 Introduction

The extensive use of power electronic devices in the distribution networks introduces perturbations into the power system causing power quality issues and challenges [1]. To counter these issues a variety of solutions are reported in the

S. Urooj · P. Sharma (✉)
Department of Electrical Engineering, Gautama Buddha University,
Greater Noida, UP, India
e-mail: pikasha29@gmail.com

S. Urooj
e-mail: shabanaurooj@ieee.org

literature. DSTATCOM is one of them, it is a power electronic converter used for either series or parallel compensation. Digital microprocessor-based controllers are used to optimize their performance [2]. The synergy between the device and the network is analysed with the help of simulation. The motive of this work is to develop a simulation model of DSTATCOM using simulink's power system block set in MATLAB [3].

Further, investigation of instantaneous reactive power theory and instantaneous active and reactive current techniques is done. The technique is used for generating the reference current for the control of DSTATCOM. The schemes discussed will be simulated under MATLAB environment using simpowersystem tool box. Obtained simulation results are showing the utility of these techniques for DSTATCOM.

2 System Configurations

Figure 1 shows the matlab-based model of the DSTATCOM [4], the system has a nonlinear load comprising of a series rlc branch connected to a three phase full wave diode-based converter. DSTATCOM model is simulated and the reference

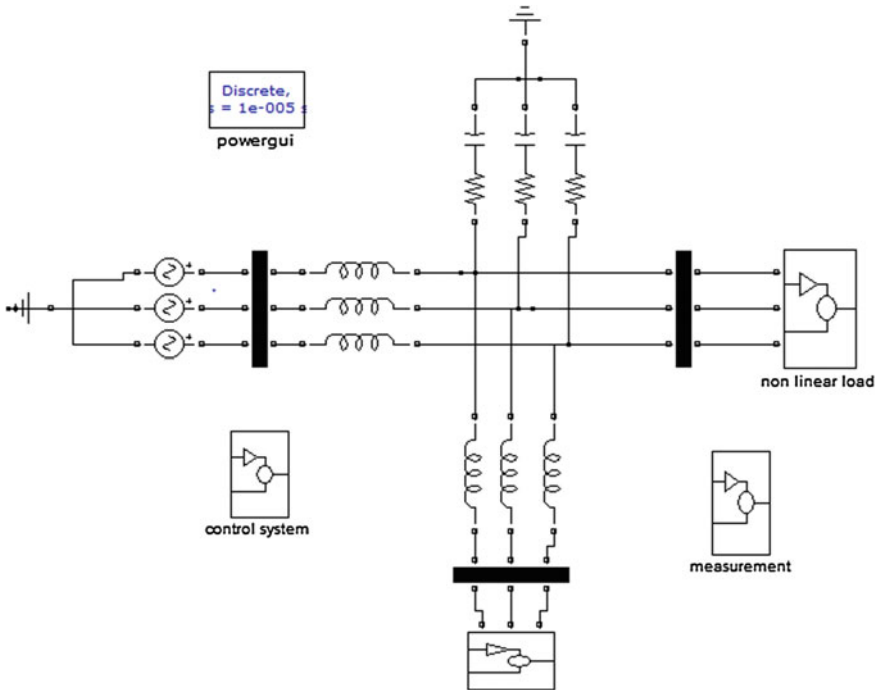


Fig. 1 MATLAB-based model of the DSTATCOM

current for the control of DSTATCOM is generated using $p-q$ (instantaneous active and reactive power theory) and SRF (instantaneous active and reactive current theory) based theories.

3 Instantaneous Reactive Power Theory

The instantaneous reactive power is obtained by the product of the instantaneous value of the current and the voltage waveforms. The DSTATCOM proposed, incorporates hysteresis technique for switching. This compensator can be used eliminate some harmonic currents by including proper control techniques and tuning of the pi controllers [5].

abc —coordinates are transformed into $\alpha - \beta$ coordinates using Clarks Transform according to the given matrix:

$$e_\alpha = \sqrt{\frac{2}{3}}e_a - \frac{1}{\sqrt{6}}e_b - \frac{1}{\sqrt{6}}e_c \tag{i}$$

$$e_\beta = \frac{1}{\sqrt{2}}e_b - \frac{1}{\sqrt{2}}e_c \tag{ii}$$

where the α and β axes are orthogonal. e_α and i_α lie along the α axis and e_β and i_β along the β axis.

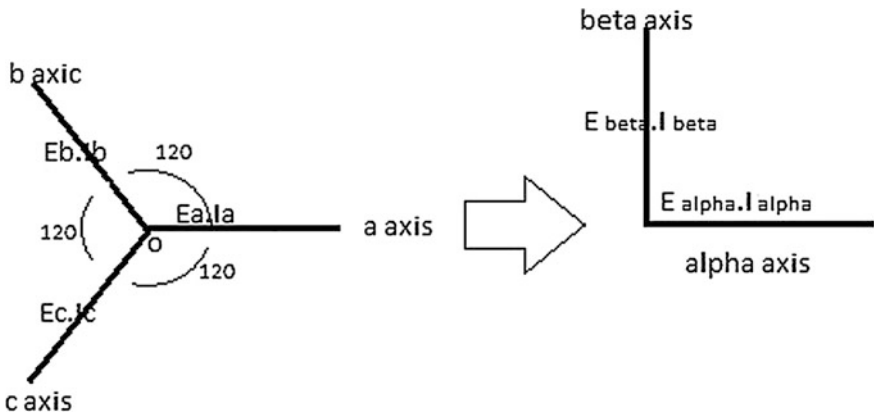


Fig. 2 Instantaneous space vectors and $\alpha - \beta$ coordinates (source [6])

Based on the Fig. 2 real and imaginary instantaneous powers are expressed as [6]:

$$p = e_\alpha i_\alpha + e_\beta i_\beta \quad (\text{iii})$$

$$p = e_a i_a + e_b i_b + e_c i_c \quad (\text{iv})$$

$$q = e_\alpha i_\beta - e_\beta i_\alpha \quad (\text{v})$$

The instantaneous imaginary power vector is perpendicular to the $\alpha - \beta$ plane[6].

3.1 Interpretation of Instantaneous Reactive Power

$$i_\alpha = \frac{e_\alpha}{e_\alpha^2 + e_\beta^2} p - \frac{e_\beta}{\sqrt{V_\alpha^2 + V_\beta^2}} q \quad (\text{vi})$$

$$i_\beta = \frac{e_\beta}{e_\alpha^2 + e_\beta^2} p - \frac{e_\alpha}{\sqrt{V_\alpha^2 + V_\beta^2}} q \quad (\text{vii})$$

i_α and i_β are separated into instantaneous active and reactive current components [7, 8]

$$\alpha - \text{axis instantaneous active current } i_{\alpha p} = \frac{e_\alpha}{e_\alpha^2 + e_\beta^2} p \quad (\text{viii})$$

$$\alpha - \text{axis instantaneous reactive current } i_{\alpha q} = \frac{-e_\beta}{e_\alpha^2 + e_\beta^2} q \quad (\text{ix})$$

$$\beta - \text{axis instantaneous reactive current } i_{\beta p} = \frac{e_\beta}{e_\alpha^2 + e_\beta^2} p \quad (\text{x})$$

$$\beta - \text{axis instantaneous reactive current } i_{\beta q} = \frac{e_\alpha}{e_\alpha^2 + e_\beta^2} q \quad (\text{xi})$$

$$p_\alpha = e_\alpha i_{\alpha p} + e_\alpha i_{\alpha q} = e_\alpha i_{\beta p} \quad (\text{xii})$$

$$p_\beta = e_\beta i_{\beta p} + e_\beta i_{\beta q} = e_\beta i_{\beta p} \quad (\text{xiii})$$

The instantaneous real power in the power circuit is given as follows [6]:

$$p = p_\alpha + p_\beta = \frac{e_\alpha^2}{e_\alpha^2 + e_\beta^2} p + \frac{e_\beta^2}{e_\alpha^2 + e_\beta^2} p + \frac{-e_\beta e_\alpha}{e_\alpha^2 + e_\beta^2} q + \frac{e_\beta e_\alpha}{e_\alpha^2 + e_\beta^2} q \quad (\text{xiv})$$

Addition of the third and fourth terms of the equation gives the answer as zero [6]

$$p = e_a i_{ap} + e_a i_{\beta p} \triangleq p_{\alpha p} + p_{\beta p} \tag{xv}$$

$$0 = e_a i_{aq} + e_a i_{\beta q} \triangleq p_{\alpha q} + p_{\beta q} \tag{xvi}$$

where,

$$\alpha - \text{axis instantaneous reactive current } p_{\alpha p} = \frac{e_x^2}{e_x^2 + e_\beta^2}$$

$$\alpha - \text{axis instantaneous reactive power } p_{\alpha q} = \frac{-e_\beta e_x}{e_x^2 + e_\beta^2} q \tag{xvii}$$

$$\beta - \text{axis instantaneous active power } p_{\beta p} = \frac{e_\beta^2}{e_x^2 + e_\beta^2} p \tag{xviii}$$

$$\beta - \text{axis instantaneous reactive power } p_{\beta q} = \frac{e_\beta e_x}{e_x^2 + e_\beta^2} q \tag{xix}$$

3.2 Control Strategy

The instantaneous reactive power on the generator side caused due to the presence of nonlinear loads on the distribution end is eliminated by the instantaneous reactive power produced by the compensator [2]. The compensator uses PWM or Hysteresis techniques for switching purposes with no component for storage of energy, because P_c i.e., real power from compensator is always zero [5] (Fig. 3).

$$i_\alpha = \frac{e_x}{e_x^2 + e_\beta^2} \bar{p} + \frac{e_\beta}{e_x^2 + e_\beta^2} \tilde{p} + \frac{-e_\beta}{e_x^2 + e_\beta^2} \bar{q} + \frac{-e_\beta}{e_x^2 + e_\beta^2} \tilde{q} \tag{xx}$$

where \bar{p} and \bar{q} are the dc part of the instantaneous real and reactive power and \tilde{p} and \tilde{q} are the ac part of the real and reactive power.

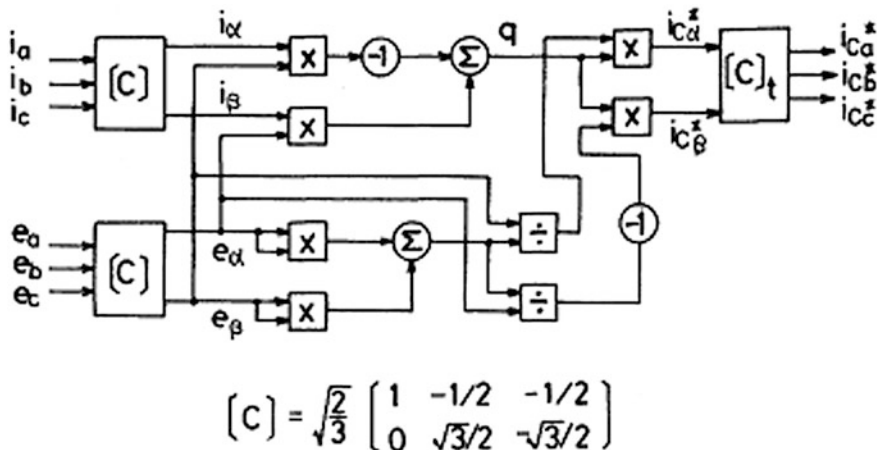


Fig. 3 Compensation scheme, source [2]

4 Id-Iq Method or Instantaneous Reactive Current Method

The i_d - i_q method is based on a synchronous rotating frame theory. In this technique, the reference current for the control DSTATCOM are obtained from the instantaneous active and reactive current components (i_{Ld} and i_{Lq}) of the nonlinear load in a two-step procedure. First, by using park transformation the current on the load side in the a-b-c reference frame is modified to the $\alpha - \beta$ reference frame [5]. Secondly alpha-beta reference frame components are converted into SRF (Fig. 4).

A d - q coordinate system has been defined such that the d and q axes rotate at the angular frequency ω , in the $\alpha - \beta$ plane. The direct and quadrature current components can be written as:

$$I_{Ldq} = I_{Ld} + jI_{Lq} \tag{xxi}$$

$$i_{Ld} = i_{L\alpha}\cos\theta + i_{L\beta}\sin\theta \tag{xxii}$$

$$i_{Lq} = -i_{L\alpha}\sin\theta + i_{L\beta}\cos\theta \tag{xxiii}$$

In the new reference frame the real component of the current is the direct axis current (i_d) and the imaginary component is called the quadrature axis component (i_q) [9].

$$V_d = \sqrt{V_\alpha^2 + V_\beta^2}, \quad V_q = 0 \text{ (as } V_q \text{ is the reference frame)}$$

Fig. 4 *abc* phases and $i_d - i_q$ reference frame. *Source* [10]

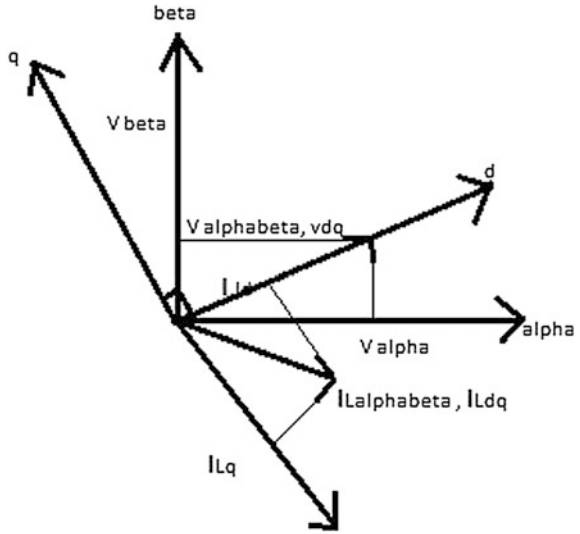
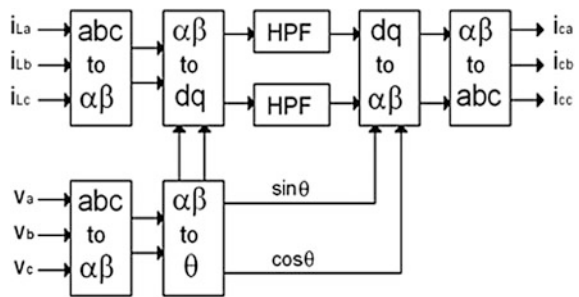


Fig. 5 *abc* to *dq* transformation, *source* [10]



The stationary reference frame load voltage vectors are:

$$i_{Ld} = \frac{V_\alpha}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{L\alpha} + \frac{V_\beta}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{L\beta} \tag{xxiv}$$

$$i_{Lq} = -\frac{V_\beta}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{L\alpha} + \frac{V_\alpha}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{L\beta} \tag{xxv}$$

the first harmonic positive sequence current is transformed to dc quantity, and other current components constitute the oscillatory parts. After removal of the DC-component of i_{Ldq} by using high-pass filters, the compensation current is obtained as: (Fig. 5)

$$I_{c\alpha} = -\frac{V_\alpha}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{Ld} + \frac{V_\beta}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{Lq} \quad (\text{xxvi})$$

$$I_{c\beta} = -\frac{V_\beta}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{Ld} - \frac{V_\alpha}{\sqrt{V_\alpha^2 + V_\beta^2}} i_{Lq} \quad (\text{xxvii})$$

Compensation powers derived from the p-q method are the parts other than the average power, the nonlinear load powers can be calculated in d-q reference frame as:

$$p_L = V_d i_{Ld} = \overline{p_L} + \widetilde{p_L} \quad (\text{xxviii})$$

$$q_L = -V_d i_{Lq} = \overline{q_L} + \widetilde{q_L} \quad (\text{xxix})$$

Under balanced conditions and the sinusoidal mains voltages are

$$p_{Cpq} = p_{Cdq} = -\overline{V_d} \widetilde{i}_{Ld} \quad (\text{xxx})$$

$$q_{Cpq} = q_{Cdq} = -\overline{V_d} \widetilde{i}_{Lq} \quad (\text{xxxix})$$

For unbalanced and non-sinusoidal mains voltages, the oscillatory part of the mains voltages is considered [11, 12].

5 Matlab Simulation of System

Figure 6a, b shows the simulation models for these theories that. The model is made using simmulink block sets of MATLAB software. The simulation of the model is done in discrete mode with a step size of 10^{-5} .

6 Results and Discussions

The outputs of the DSTATCOM are studied for both the two methods of control techniques, and the following observations are rendered based on these results.

6.1 Control of DSTATCOM by IRPT

The considered load is nonlinear and resistive reactive in nature with 0.8 power factor. There is a delay in compensation for the source current and that is because of

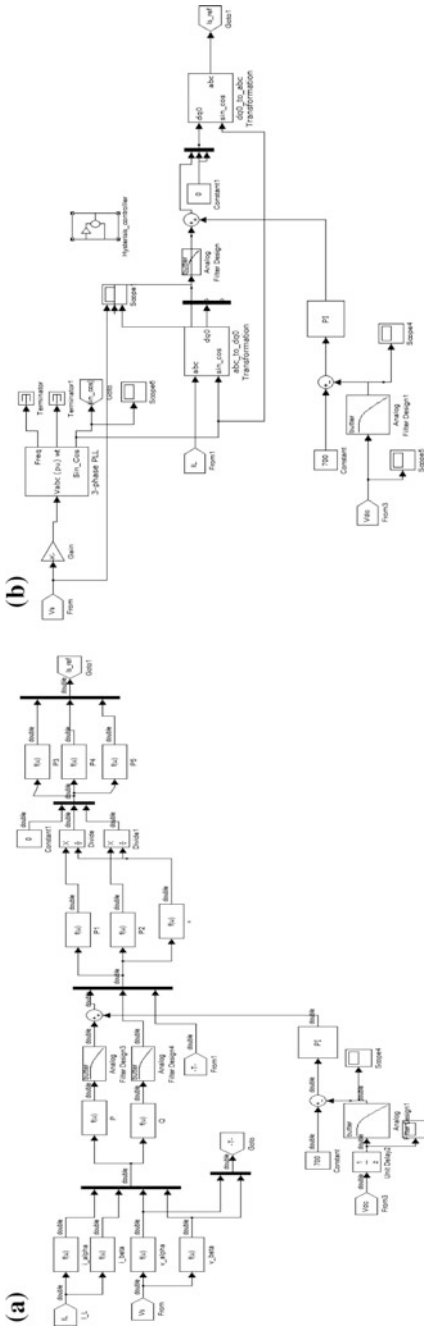


Fig. 6 MATLAB-based model for reference current extraction by using a IRLP, b SRF

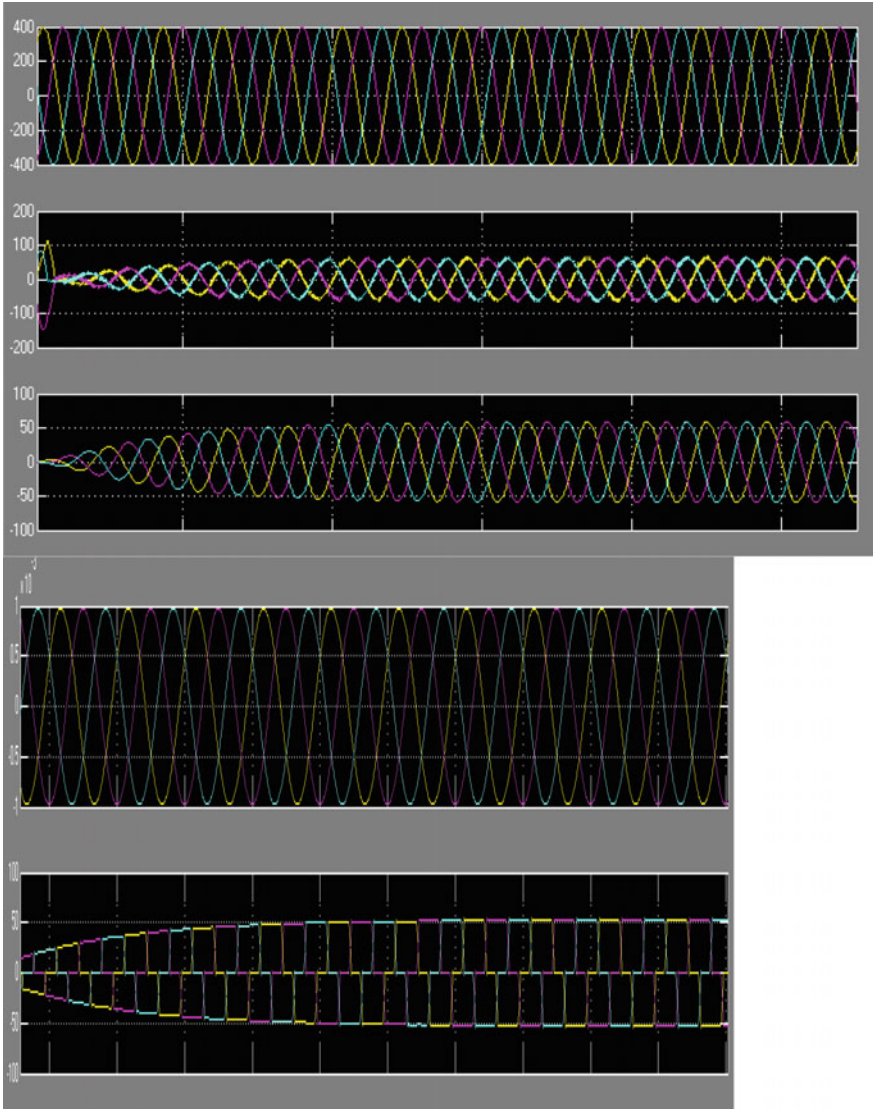


Fig. 7 Waveforms of (1) source voltage, (2) source current, (3) reference current and dc voltage for a system using DSTATCOM for compensation

the presence of a series linear power filter used for filtering harmonics. In the instantaneous reactive power theory PI control of the voltage components is performed to derive the reference current. This is then used to calculate the real and imaginary power by Clark's transformation. Any kind of voltage unbalance will result into imprecise calculation of the reference current [9] (Fig. 7).

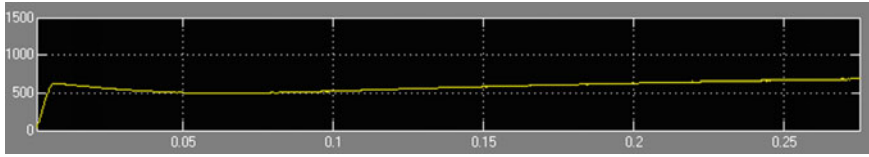


Fig. 8 Waveform of (1) source voltage and (2) source current for a system using a DSTATCOM

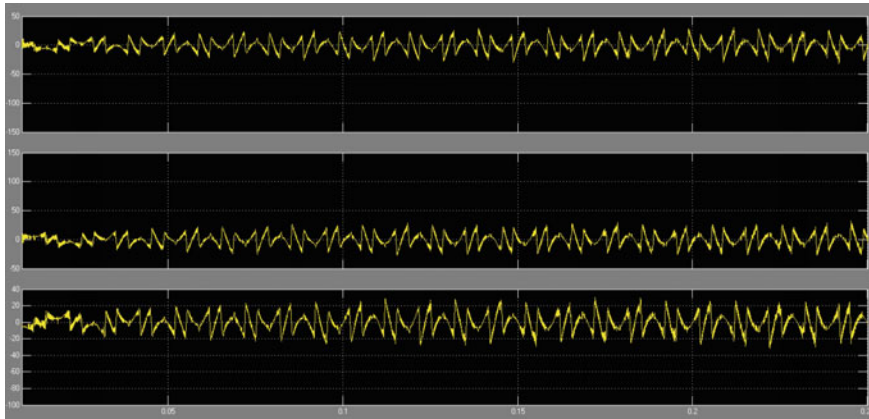


Fig. 9 Compensating current waveforms provided by IRPT controlled DSTATCOM for harmonics elimination

It can be seen that there are harmonics present in the source current when a DSTATCOM is not used as a compensator. However, with the use of DSTATCOM in Fig. 8 the source voltage becomes sinusoidal in nature as it is made to follow the reference current generated through the instantaneous reactive power control scheme.

6.2 Control of DSTATCOM by SRF Theory

Figure 9 shows the output of a DSTATCOM controlled by Synchronous reference frame theory. In the second case load is same as that used for simulation of IRPT technique. And there is a delay in the compensation of current due to presence of linear series filter [10]. The current in stationary abc reference frame is directly transformed into synchronous reference frame. Proper tuning of PI controller and the PLL is indispensable for the accuracy in the calculation of the reference current (Figs. 10, 11 and 12).

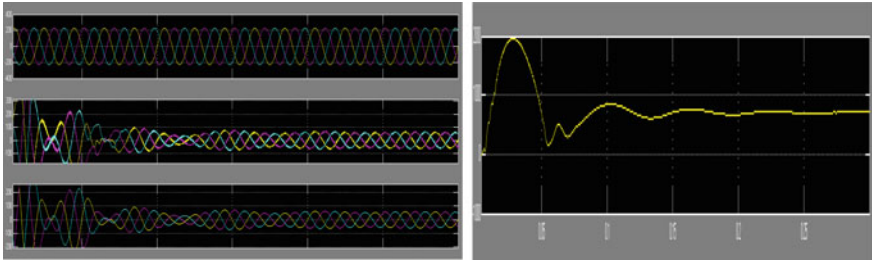


Fig. 10 Waveforms of (1) source voltage, (2) source current, (3) reference current and dc voltage for a system using DSTATCOM for compensation

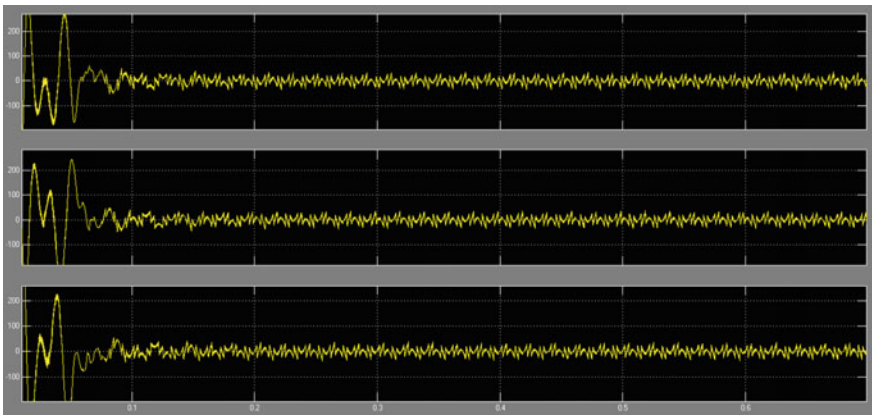


Fig. 11 Compensating current waveforms provided by srf controlled DSTATCOM for harmonics elimination

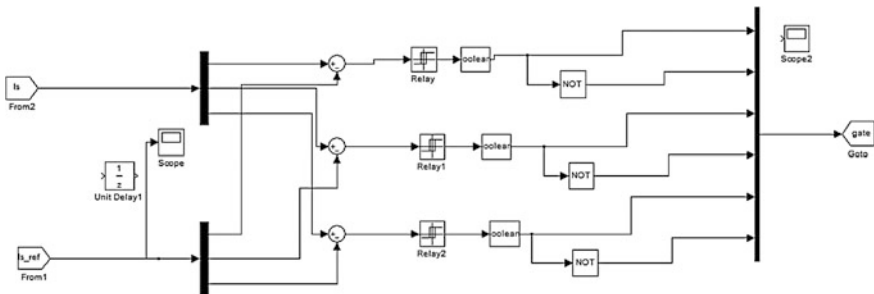


Fig. 12 Hysteresis current controller used for for gating pulse generation in IRPT and SRF techniques

In both the control techniques a hysteresis controller is used in order to generate the gate pulses for the controlled operation of the voltage source converter. The upper and the lower limits taken for the controller are 0.2 and -0.2 .

7 Experimental Results

7.1 Selection of DC Voltage

The dc bus voltage is maintained around 750 V for SRF-based DSTATCOM control technique and 1000 V for IRPT technique. The reference dc bus voltage can be obtained by the formula:

$$V_{dc\text{ref}} = \frac{\sqrt{2}V_c}{m}$$

where, m is the amount of modulation applied on the voltage signal, it is assumed to be 1 and V_c is the Ac voltage produced by the DSTATCOM for compensation of the source voltage [9].

7.2 Selection of DC Capacitor

In accordance with principle of energy conversion difference of the energy stored in the capacitor due to minimum voltage and set reference voltage is equal to the three phase compensating power supplied by compensator.

$$\frac{1}{2} C_{dc} [(V_{dc\text{ref}}^2) - (V_{dc\text{min}}^2)] = 3V(aI) t$$

In both the IRPT and SRF technique the DC capacitance is taken to be 1000 μF . The minimum reference voltage is set to be 340 V in both the cases [9] (Fig. 13).

7.3 FFT Analysis for All the Three Cases

The total harmonic distortions level in the source current in case of an uncompensated system is 24.81 %, THD level in the source current when using a DSTATCOM compensated system using IRPT-based control technique is 5.08 %, THD level for the source current when using a DSTATCOM compensated system using SRF-based control technique is 4.06 %.

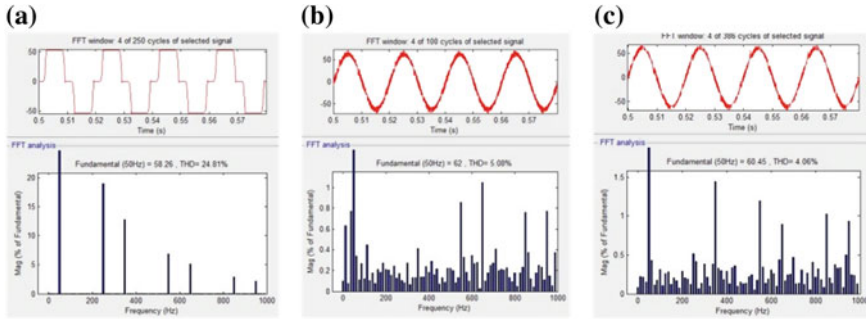


Fig. 13 a Shows FFT analysis of source current for uncompensated system, b for IRPT-based control, c for SRF-based control

8 Conclusion

By using IRPT and SRF control techniques the performance of the DSTATCOM is optimized. It has been proved that the source current can be made to follow the reference current which is sinusoidal in nature by using appropriate reference current generation techniques. If comparison between above two techniques is done then it can be said that for SRF theory the THD level in source current is less than that in the IRP theory. However, there is not much difference and both can be considered as equally reliable and useful.

References

1. Hussain, M., Zaman, M.U., Iqbal, S., Urooj, S.: Performance investigation of a single phase unidirectional buck converter for improved power quality. In: IEEE International Conference on Computing for Sustainable Global Development INDIACom2014 IEEE, held at BVICAM New Delhi India during 5–7 March 2014. doi:[10.1109/IndiaCom.2014.6828148](https://doi.org/10.1109/IndiaCom.2014.6828148)
2. Zaveri, T., Bhalja, B., Zaveri, N.: Simulation and analysis of control strategies for STATCOM. In: ACEEE International Journal on Control and Instrumentation, vol. 1, no. 1. ACEEE International Journal (2010)
3. Suresh, M., Patnaik, S.S., Suresh, Y., Panda, A.K.: Comparison of two compensation control strategies for shunt active power filter in three-phase four-wire system. In: IEEE conference on innovative Smart Grid technologies, pp. 1–6. IEEE PES, CA (2011)
4. Sing, B., Solanki, J.: A comparative study of control algorithms for DSTATCOM for load compensation. In: IEEE International Conference on Industrial Technology, vol. 56, pp. 2745. IEEE Transactions (2009)
5. Akagi, H., Kanazawa, Y., Nabae, A.: Instantaneous reactive power compensators comprising switching devices without energy storage components. In: IEEE Transactions on Industry Applications, vol. 1A-20, pp. 625–630. IEEE Transactions (1984)
6. Singh, B., Adya, A., Mittal, A.P., Gupta, J.R.P.: Modelling and control of DSTATCOM for three-phase, four-wire distribution systems. In: IEEE Industry Applications Conference, 2005. Fourtieth IAS Annual Meeting, vol. 4, pp. 2424–2434. IEEE conference record of (2005)

7. Akagi, H., Nabae, A.: The p-q theory in three-phase systems under non-sinusoidal conditions. In: *European Transactions on Electrical Power*, vol. 3, pp. 31. IEEE Transactions (1993)
8. Shabana, U., Das, S.: Harmonic convergence for balanced and unbalanced nonlinear load using hybrid filter. *Int. J. Converg. Comput.* **1**(2), 109–117 doi:[10.1504/IJCONVC.2014.063740](https://doi.org/10.1504/IJCONVC.2014.063740)
9. Chang, W., Shee, T.: A comparative study of active power filter compensation approaches. In: *IEEE Power Engineering Society Summer Meeting*, vol. 2, pp. 1017. IEEE Press, Chicago (2002)
10. Gyugyi, L. et al.: Principles and applications of static, thyristor-controlled shunt compensators. In: *IEEE Transactions on Power Apparatus and System*, vol. PAS-99, pp. 1795–1804. IEEE Transaction (1978)
11. Yao, X., Fangxing, L.: Adaptive PI control of statcom for voltage regulation. In: *IEEE transactions on power delivery*, vol. 29, pp. 1002–1011. IEEE Transaction (2014)
12. Urooj, S., Khan, M.: Sub-optimal controller for airframe model of harrier AV-8A VTOL Aircraft. *South East Asian Journal of Mathematics and Mathematical Sciences*, published in vol. 6, no. 3 (2008)

Fault Tolerance in Wireless Sensor Networks: Finding Primary Path

Pritee Parwekar and Sireesha Rodda

Abstract Wireless sensor networks (WSN) are prone to be affected by faults which may be caused due to a variety of reasons, namely hardware malfunction, software problems, inadequate energy resources, and range or environmental hazards. A WSN is required to be equipped to handle such situations else would suffer an overall curtailment of the lifetime and ultimately not meet the required goal. Tolerance to faults thus forms one of the guiding parameters in WSN design. In this paper, we have proposed a method to find a reliable routing protocol using fuzzy logic based on Link Quality Indicator (LQI), Received Signal Strength Indicator (RSSI), and number of hops to the base station. Additional to the primary reliable path, each node has a secondary path which will be alternate path for each sensor in case of failure of the primary path. Implementation of this approach has been done as TinyOS module and evaluated through TOSSIM simulations. The experimental results show promising results in terms of packet delivery and reliability of the network.

Keywords Fuzzy logic · Received signal strength indicator · Link quality indicator

1 Introduction

The functional benefits of wireless sensor networks have now been well understood and accepted. This has resulted in extensive research toward developing low cost, sustainable, and highly reliable wireless sensor networks. With this quest, energy

P. Parwekar (✉)

Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, India
e-mail: pritee.cse@anits.edu.in

S. Rodda

GITAM University, Visakhapatnam, India
e-mail: sireesha@gitam.edu

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_57

593

conservation has become almost an undoubted primary goal, while throughput and fault tolerance have found second place yet are significant factors to realize the final goal of a sustainable and reliable WSN.

A network of sensors that work independent of a common centralized administrator is a typical characteristic of a wireless sensor network. Each one of the sensors nodes performs the tasks like a router. They should be able to adapt to the changing topology, i.e., the routing tables should change with a change in their connection topology. Routing in WSNs is usually classified based on the network structure, as flat-based, hierarchical, and location-based. The nodes in the network are assigned equal or same functional roles in flat-based routing. The nodes in the network are assigned differing roles in a hierarchical-based routing architecture. Location-based routing uses sensor node positions used to route network data. An adaptive routing protocol is able to adapt to current network conditions by controlling system parameters like signal strength, nodes energy, or relative positions of sensors. Also, these protocols can be further classified into multipath-based protocols, query-based protocols, negotiation-based protocols, QoS-based protocols, or coherent-based routing techniques depending on the protocol operation. Depending on the method of the source finds a route to the destination the above routing protocols can be additionally classified into three categories viz. the proactive, the reactive, and the hybrid [1]. In proactive protocols, all the routes are computed and stored before they are actually needed while in the reactive protocols, these routes are computed in realtime. Hybrid protocols optimally use both these ideas. Compared to the reactive protocols, it has been found that table-driven routing protocols are more promising due to the static sensor nodes. Reactive protocols tend to be resource hungry due to the activities like setup and Route. The data from the sensor is collected and further processed by a central node. This is yet another class called as cooperative routing. This aims at reducing the energy use and thereby the route cost. A few other protocols depend on time and position information.

The Wireless Sensor Networks [2] are required to be protected from typical faults like sensor failure due to exhaustion of energy, thereby, failure of a particular path or the complete network using a Fault-tolerant mechanism. Considering the typical issues with Wireless Sensor Networks a handful of fault-tolerant techniques have emerged. Data is transmitted through multiple paths from source to destination to provide for data redundancy. The quick recovering technology is employed to restore broken paths and thereby, the robustness of the network is increased. The source sends the data as several copies across multiple paths to the destination, so as to ensure that the data is delivered with a certain degree of reliability. However, the cluster head (CH) has a task of grading the sensors based on several parameters and this makes the cluster head quite energy hungry. Another unresolved issue is the balance of energy consumption and dependability for the CH. Employing two or three paths to ensure the accuracy of the data, identifying faulty, or dead sensors in a particular path and the shifting over to an alternate path facilitates an end-to-end transmission. The mechanism shows an improvement in the end-to-end message arrival rate but the fault tolerance for each sensor has not yet been addressed.

The present paper is based on path reliability for creating an alternate path for every sensor on the path of data transmission and the same is simulated to obtain results.

As we go along, the Sect. 2 brings out the related work and previous studies on the subject. Section 3 has been earmarked to explain the protocol in detail and shows how the implementation has been undertaken of the proposed protocol. The simulation results and the efficiency of the proposed method have been discussed in Sect. 4 and finally in Sect. 5 we conclude.

2 Related Work

Though there are a lot many protocols which have been studied in our course of research, however, only a certain cluster-based protocols have been reflected in the related work section. Low Energy Adaptive Clustering Hierarchy (LEACH) protocol was introduced by Heinzelman et al. [3] for sensor networks. LEACH is a cluster-based protocol, employing distributed cluster formation which selects a few sensor nodes at random to function as cluster heads (CHs) and then reassign this role in rotation so as to maintain an even distribution of the energy load among the sensors in the network. The CH nodes of the cluster compress the data arriving from nodes, and further send the same as an aggregated packet to the BS thereby, minimizing the amount of information transmitted to the BS. LEACH employs a TDMA/code-division multiple access (CDMA) MAC to reduce intercluster and intracluster collisions. However, data collection is centralized and performed periodically. When there is a need for continuous monitoring by the sensor network, the LEACH protocol is found to be the most beneficial. The primary-backup path routing algorithm (RPBLR) [2] protocol, selects a backup path for all sensors for reliable route to reach destination. RPBLR uses weight factor, link delay, and volume of data for calculating the backup route. Thus between the sink and the node, there are no cluster heads as intermediaries. Having an alternative path other than a primary path is advantageous so as to better load balance the network with no single node being used for the transmission and thereby, preserving the energy of the node and balance the energy of all the nodes in the network. This also facilitates secure transmission, increased reliability toward ensuring that the message reaches from the source to the destination. Al-Karaki and Kamal [4] has described advantages of multipath routing in detail. Further, the author has brought out the design issues, challenges, and evaluation metrics of multipath routing protocols with their resource constraints. Paper [5] has also proposed two options for finding the alternate path in sensor network to have a best communication in the network, with better energy utilization. It uses a concept of finding the multiple paths by using disjoint sets of a node and another technique used is finding the alternate path using partial disjoint sets of a node which is called as braided multipath. These options of disjoint path are used to find an alternative path in case of any node

failure in the network. Author [6] has used two steps to find a multipath which are formulation of optimized cluster and subsequently finding a tree-based multipath path which will have fault free communication. This distributed multipath fault-tolerant routing method has proposed to solve a heady traffic problem on each node and to avoid a hole in network.

The research in wireless sensor networks is focused on the development of routing protocols that provide application-specific service guarantees. Hammoudeh and Newman [7] paper presented a clustering protocol called as ROL, for route optimization and load balancing in sensor network. This protocol uses different quality measures to develop an application-specific protocol for the network. This protocol looks for the better life of the network. ROL is an Optimization and Load-balancing protocol which employs Quality of Service (QoS) metrics to meet application requirements. ROL combines many application requirements and thereby provides a comprehensive solution to achieve a prolonged network life, on time message delivery and augments network robustness.

In [8], the authors used fuzzy logic to find the location of sensor network. Fuzzy logic is an easy method which tries to emulate the thinking process of the human brain. The only problem in sensor network with fuzzy logic is to save the fuzzy rules. Sensors have limited memory resource. In fuzzy logic this memory is used to save the rule-based system to make a decision. Alwan and Agarwal [9] talks about the survey of fault tolerance routing protocol used in wireless sensor networks. A comparative study is given of retransmission and replication-based routing protocols are with memory consumption, energy usage, recovery time, and overhead.

Multipath routing is the most used fault tolerance mechanism. Multiple paths are identified between the sources and the sink. But this process consumes additional energy and generates an additional network traffic. Other benefits include a better load-balancing and bandwidth aggregation. Based on the path finding methods, the routing protocols are classified as Proactive, reactive, and hybrid routing [1]. The fault tolerance mechanisms are also classified into retransmission and replication types.

2.1 Retransmission

Retransmission is the most popular method in which the data packets are retransmitted to the sink. Among the available multiple path, the protocol selects one with minimum hop count and/or minimum energy consumption. The sink node finally acknowledges the packet to the source. The source retransmits the data packet if the acknowledgment is not received prior timeout. The packet loss rate in Wireless Sensor Networks is quite large as compared to other types of networks. Therefore, the link level retransmissions have gained popularity.

2.2 Replication

Fault tolerance can also be provided in WSNs by introducing redundancy into packet delivery [10]. This is achieved by sending the original packet over different paths as copies and thereby, negating network failure due to local path failures. This is a popular replication mechanism.

2.3 Causes of Fault in a Wireless Sensor Network

A fault may occur due to one or more of the following reasons:

- connectivity of node. The physical connectivity between the nodes is breached due to poor range or other interferences from outside the network.
- hardware fault in circuit. One of the nodes experiences a breakdown.
- low battery/energy. The node is dead due to drained out battery power
- attack in sensor network. The network has experienced an attack due to which a node or a few nodes are not available to the network.

Any fault in the sensor network will adversely affect the availability of the network or may have a detrimental effect on the efficiency of the network. To maintain a minimum level of availability and efficiency, it is prudent to identify and rectify the faults in a timely and efficient manner and on the go [11].

Most of the work in the literature [12, 13] have used Heinzelman's energy dissipation radio model [3] for sensor networks. It is simple abstraction for a theoretical radio model and the results are not realistic. Many current platforms (micaZ, Telos, Shimmer Mote, Tmote Sky, Zolertia Z1 mote, and Intel Mote2) use the same radio chip, the CC2420 [14]. This is the more realistic model and used in many wireless networks.

In our proposed approach we use TinyOS [15] module and TOSSIM [16] for extensive simulations. TinyOS uses CC2420 radio chip, and we use the CC2420 power consumption model in our approach and evaluated with the existing work [12].

3 Proposed Approach

In this paper, we propose reliable routing algorithm using fuzzy logic for WSNs. It relies on the Receive Signal Strength Indicator (RSSI), Link Quality Indicator (LQI), and the number of hops to the base station. The LQI is used to indicate how strong the communication link is. RSSI provides a measure of the signal strength at the receiver. The "quality" or "correctness" of the signal does not become a significant issue. The RSSI value is measured in dBm and express the signal power.

The values typically range from -45 to -100 dBm. The lower value is determined by the receiver input threshold, and the upper by the airborne signal strength. The LQI value reflects the link quality seen from the receiver side. The LQI value correlates with the Packet Reception Rate (PRR) and is therefore a very important figure in routing protocols. LQI does not care about the actual signal strength but a stronger signal is likely to be more cleaner from noise and therefore more correctly received. LQI provides an average correlation value, and the values are predominantly between 110 and 50, and corresponding to maximum and minimum quality frames, respectively.

The input fuzzy variables are the RSSI, LQI, and the number of hops to the base station in the proposed algorithm. The cost of the link between two sensor nodes is the output variable [13]. We decompose both the input space of RSSI and LQI and output space of cost into three triangular-shaped membership functions: low, medium, and high, respectively. The membership functions Less and More are defined on number of hops input variable. The Less is a triangular membership function while More is a trapezoidal membership function.

Figures 1, 2 and 3 show the membership graphs for input variables RSSI, LQI and number of hops to the base station respectively. Figure 4 shows the membership graph for output variable cost. Using Fuzzy rules the weights are calculated based on the Table 1.

In order to simulate the WSNs, we should provide the characteristics of both the radio, and the environment (channel) where they are placed. We use the radio propagation model path loss to account for the environment factors. The model chooses a distance d_0 in which the lognormal path-loss model [17]. The log-scale propagation model generalizes far field and measures.

Algorithm 1: GETLQI

Input: An RSSI register value: rssi_reg

Output: lqi value

1. $\text{rssi_reg} = (\text{rssi_reg} > 255 - 45) ? 255 : \text{rssi_reg} + 45$
2. $x = \text{rssi_reg}$
3. $p1 = -0.6475, p2 = 3.761, p3 = -5.997$
4. $p4 = 3.418, p5 = 105.9$
5. $x = x - 222.6$
6. $x = x / 10.61$ //and give a rough estimate of what the LQI would have been for this rssi
7. $\text{lqi} = (p1 * \text{pow}(x, 4.0)) + (p2 * \text{pow}(x, 3.0)) + (p3 * \text{pow}(x, 2.0)) + (p4 * x) + p5$
8. $\text{lqi} = ((\text{lqi} > 0) ? \text{lqi} : 0.0)$
9. return lqi

The path loss $PL(d)$. The propagation path-loss factor n indicates the rate of path loss increase with the distance. It can be found from the measurements that at a value d , the path loss $PL(d)$ at a particular location is random and distributed lognormally (normal in dB) about the mean distance-dependent value. That is,

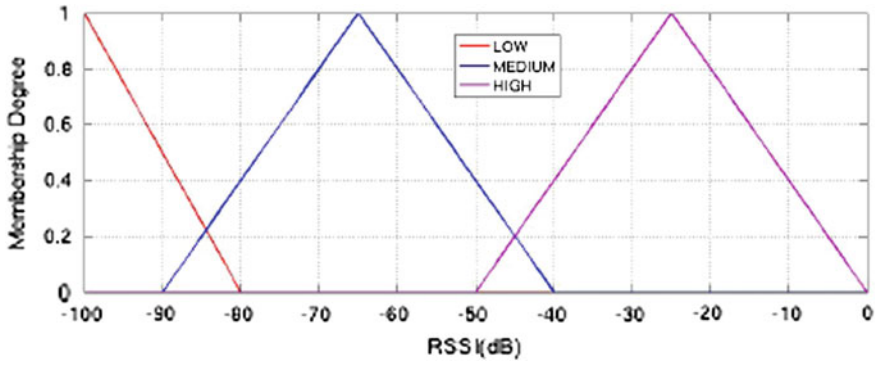


Fig. 1 Membership function of RSSI

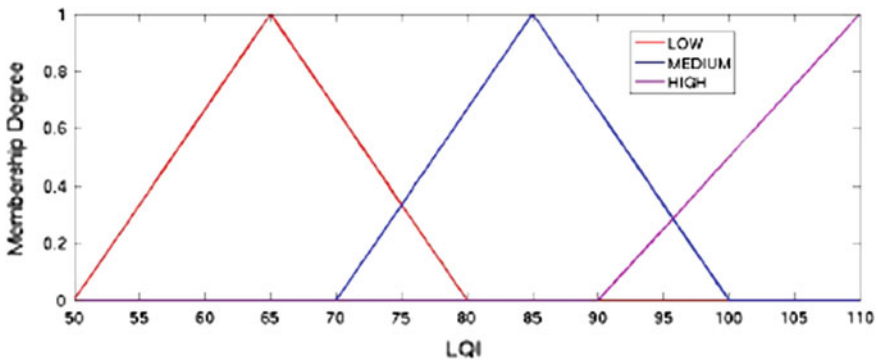


Fig. 2 Membership function of LQI

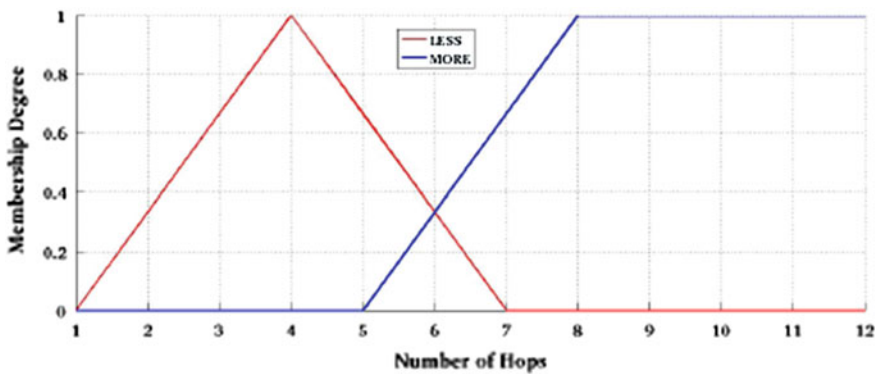


Fig. 3 Membership function of no. of hops

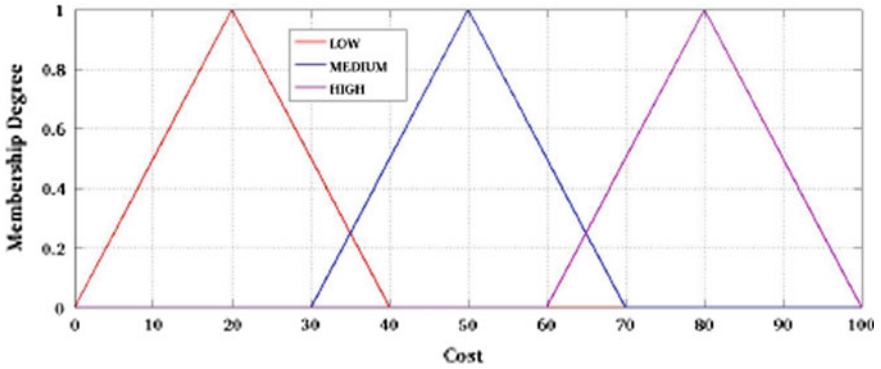


Fig. 4 Membership function of cost

Table 1 Fuzzy rule-based table

Rule	RSSI	LQI	No. of hops	Weight
1	Low	Low	Less	Low
2	Low	Medium	Less	Low
3	Low	High	Less	Medium
4	Medium	Low	Less	Low
5	Medium	Medium	Less	Medium
6	Medium	High	Less	High
7	High	Low	Less	Medium
8	High	Medium	Less	High
9	High	High	Less	High
10	Low	Low	More	Low
11	Low	Medium	More	Low
12	Low	High	More	Medium
13	Medium	Low	More	Low
14	Medium	Medium	More	Low
15	Medium	High	More	Medium
16	High	Low	More	Low
17	High	Medium	More	Medium
18	High	High	More	Medium

$$PL(d) = PL(d0) + 10\eta \log 10(d/d0) + X\sigma \tag{1}$$

where $X\sigma$ is a zero-mean Gaussian distributed random variable (in dB) with standard deviation σ (also in dB). Here we use $d0 = 8$ and $n = 3.3$.

We have implemented this approach as a TinyOS module and evaluated through TOSSIM simulations. In order to get RSSI and LQI values in TOSSIM, the TOSSIM code is modified and LQI is calculated using RSSI value. RSSI is the estimate of the signal power. It is calculated over eight symbol periods and is stored

in the RSSI VAL register. To calculate LQI from RSSI, we used linear polynomial model with coefficients (95 % confidence bounds). The polynomial is,

$$\text{Poly}(x) = p_1x^4 + p_2x^3 + p_3x^2 + p_4x + p_5 \quad (2)$$

where x is the RSSI value received from anchor node. A curve fitting technique is employed and a fourth polynomial curve is best fitted with the data, RSSI to LQI, in a least squares sense. Algorithm 1 will find the LQI value from RSSI.

TOSSIM uses the CC2420 radio and it consumes 62 mW in Receiving/listening mode. In the transmit mode, multiple TX power level with different power consumption (e.g., 0 dBm consumes 57.42 mW). In our simulations, we used the default transmission power for the CC2420, CC2420 DEF RF POWER = 31, which corresponds to maximum power (0 dB). In the sleep mode, the radio consumes 1.4mW.

Our proposed routing algorithm works as follows: if a source node desires to send a packet to base station then it searches for the route on the fly and establishes an optimized connection to transmit and receive the packet. The route request packets are flooded through the network to facilitate route discovery. The nodes which possess a valid route to the base station replies with a route reply packet to the source. Our approach uses the RSSI, LQI and number of hops to the base station from the route reply packets and establishes the route from source to base station based on fuzzy logic. Each node maintains two paths one is primary and another one is secondary to their neighboring nodes. The primary route is the one who is having the best edge weight and the secondary path having the next best value, respectively. In our approach, we used the source node id, sequence number and ttl to avoid loop formation.

4 Simulation and Results

The performance of our proposed approach implemented as a TinyOS module and evaluated through extensive TOSSIM simulations. TOSSIM simulates the behavior and interactions of networks at the network bit granularity and not on the packet level. And more importantly, the same nesC code without tweaking is able to run on actual motes (hardware) and also in the TOSSIM simulator environment. TOSSIM requires topology and noise file as inputs and the topology file is generated using log-distance path-loss model. The noise file is taken from TOSSIM source code. We evaluated our approach with 25, 50, 75, and 100 nodes. The nodes were deployed randomly in the 100 m by 100 m field with base station 1 is located at the center. And the topology for 25 nodes network as shown in Fig. 5.

The performance of our proposed approach was compared with [12]. The results that are presented in Figs. 6 and 7. Figure 6 shows the percentage of dead nodes with respect to total number of transmissions. The dead nodes are the nodes in the network whose remaining energy is zero. We observe that our approach shows the

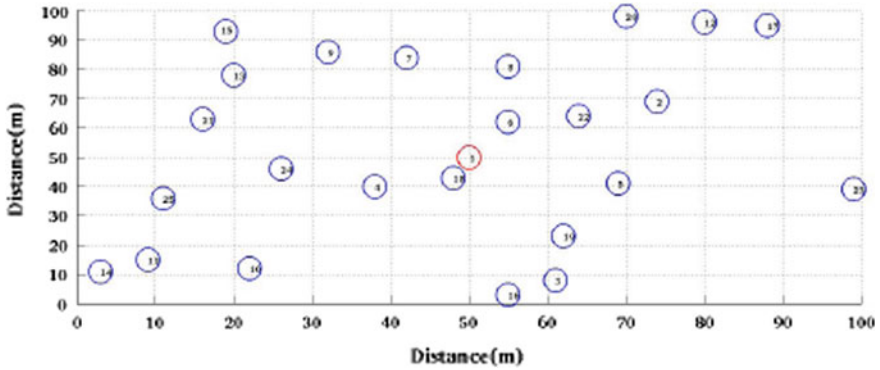


Fig. 5 Simulation environment

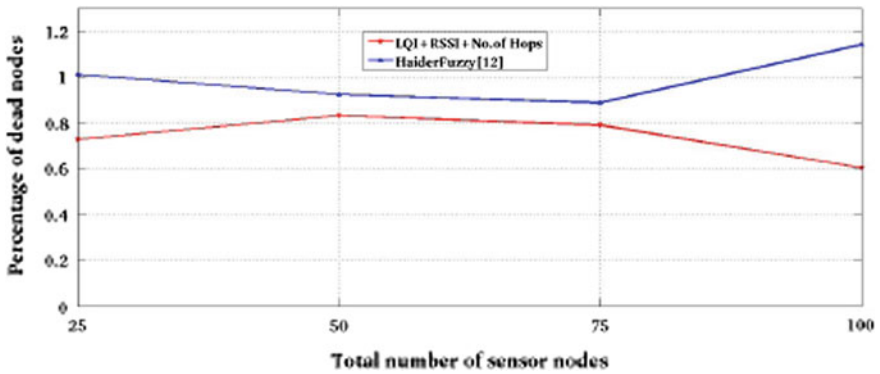


Fig. 6 Performance of our approach versus Haider Fuzzy

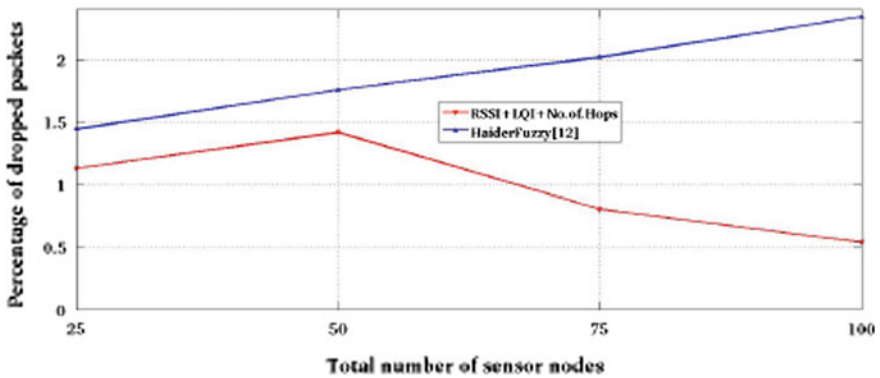


Fig. 7 Performance of our approach versus Haider Fuzzy

better reliability of the network compare to the existing approach [12]. Figure 7 shows the packet drop percentage with respect to the total number of transmissions. We see that our approach consistently gives better results and the drop rate is much lower than the existing approach [12] even with an increase in the number of sensor nodes. In our approach each node uses the secondary path in case the primary path fails so the packet reception rate is high. Whereas in [12] the drop rate increases when the network size is large.

5 Conclusion

In this paper, we proposed a fuzzy-based reliable routing algorithm for wireless sensor networks. The RSSI, LQI and number of hops to the base station is used to estimate the path from source to the base station. Fuzzy logic system is the main component of the proposed scheme. The edge weight of each node in the path is found out using Mamdani fuzzy system. Our approach is implemented as a TinyOS module and evaluated using TOSSIM simulator. Our simulation results highlight the high packet delivery and reliability of the network.

References

1. Akkaya, K., Younis, M.: A survey on routing protocols for wireless sensor networks. *Ad hoc Networks* **3**(3), 325–349 (2005)
2. Weipeng, J., Qu, W., Yaqiu, L., Qianlong, Z.: A reliable primary backup routing algorithm in wireless sensor network. *Phys. Procedia* **24**, 1462–1468 (2012)
3. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy efficient communication protocol for wireless microsensor networks. In: *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, 2000, p. 10. IEEE (2000)
4. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey. *IEEE Wirel. Commun.* **11**(6), 6–28 (2004)
5. Ganesan, D., Govindan, R., Shenker, S., Estrin, D.: Highly-resilient, energy-efficient multipath routing in wireless sensor networks. *ACMSIGMOBILE Mobile Comput. Commun. Rev.* **5**(4), 11–25 (2001)
6. Chanak, P., Banerjee, I., Rahaman, H.: Distributed multipath fault tolerance routing scheme for wireless sensor networks. In: *2013 Third International Conference on Advanced Computing and Communication Technologies (ACCT)*, pp. 241–247. IEEE (2013)
7. Hammoudeh, M., Newman, R.: Adaptive routing in wireless sensor networks: Qos optimisation for enhanced application performance. *Information Fusion* (2013)
8. Parwekar, P., Reddy, R.: An efficient fuzzy localization approach in wireless sensor networks. In: *2013 IEEE International Conference on Fuzzy Systems (FUZZ)*, pp. 1–6. IEEE (2013)
9. Alwan, H., Agarwal, A.: A survey on fault tolerant routing techniques in wireless sensor networks. In: *SENSORCOMM'09. Third International Conference on Sensor Technologies and Applications*, 2009, pp. 366–371. IEEE (2009)
10. Kim, S., Fonseca, R., Culler, D.: Reliable transfer on wireless sensor networks. In: *IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, 2004, pp. 449–459. IEEE (2004)

11. Raj, A.S.A., Ramalakshmi, M.K., Priyadharsini, M.C.: A survey on classification of fault tolerance techniques available in wireless sensor network. In: International Journal of Engineering Research and Technology, vol. 3, no. 1 (January-2014). ESRSA Publications (2014)
12. Haider, T., Yusuf, M.: A fuzzy approach to energy optimized routing for wireless sensor networks. Int. Arab J. Inform. Technol. (IAJIT) **6**(2), 179–186 (2009)
13. Manjunatha, P., Verma, A., Srividya, A.: Fuzzy based optimized routing protocol for wireless sensor networks. In: Advances in Wireless Sensors and Sensor Networks, pp. 273–282. Springer (2010)
14. Chipcon inc: <http://www.chipcon.com>
15. Tinyos: <http://tinyos.net/>
16. Levis, P., Lee, N., Welsh, M., Culler, D.: Tossim: Accurate and scalable simulation of entire tinyos applications. In: Proceedings of the 1st international conference on Embedded networked sensor systems, pp. 126–137. CM (2003)
17. Rappaport, T.S. et al.: Wireless Communications: Principles and Practice, vol. 2. Prentice Hall PTR, New Jersey (2002)

PSO-Based Multiple-sink Placement Algorithm for Protracting the Lifetime of Wireless Sensor Networks

C. Srinivasa Rao P., Haider Banka and Prasanta K. Jana

Abstract Optimal deployment of multiple sinks has been proven to be one of the energy efficient techniques for prolonging the lifetime of wireless sensor networks (WSNs). In this paper, we propose a particle swarm optimization (PSO) based algorithm called PSO-MSPA for placement of multiple-sink in WSNs. The algorithm is developed with an efficient scheme of particle encoding and novel fitness function. For the energy efficiency of the PSO-MSPA, we consider various parameters such as Euclidian distance and hop count from the gateways to the sinks. The algorithm is tested extensively on various scenarios of WSNs by varying number of gateways and sensor nodes and the results are analyzed to show the efficacy of the proposed algorithm.

Keywords Multiple-sink placement problem · NP-hard problem · Particle swarm optimization · Network lifetime

1 Introduction

Wireless sensor networks (WSNs) are networks of spatially distributed sensor nodes which collect and process the data from a region of interest and communicate to a base station [1]. WSNs have paid enormous attention in wide range of applications such as disaster management, environment monitoring, health, and military [2]. Due to battery operated sensors, energy conservation of the sensor

C. Srinivasa Rao P. (✉) · H. Banka · P.K. Jana
Department of Computer Science and Engineering,
Indian School of Mines, Dhanbad 826004, India
e-mail: vasu.piiit@gmail.com

H. Banka
e-mail: hbanka2002@yahoo.com

P.K. Jana
e-mail: prasantajana@yahoo.com

nodes is the main challenging issue in the development of a large-scale WSN. Clustering sensor nodes is one of the most energy efficient techniques that have been studied rigorously. In a cluster based WSNs, sensor nodes are organized into number of groups called clusters and each cluster has a leader called as cluster head (CHs). The CHs collect the data from each member sensor nodes within their clusters, aggregate them and send the aggregate data to a remote base station called sink [3]. To overcome the network deficiencies researchers have introduced special nodes called gateways which are provisioned with extra energy [4]. These gateways are responsible for data collection, aggregation, and relaying the aggregated data to the sink and thus actually serve the role of CHs. However, they are also battery operated and therefore, we should take care the energy conservation of the gateways.

Position of a sink has a tremendous impact on the energy consumption of gateways in the process of data flow between the gateways and the sink [5, 6]. It may also have the influence of the other performance metric, such as delay, risk of packet loss, and throughput of the network. One possible solution is to use of multiple-sink so that the gateways can disseminated their collected data to their nearest sink with less communication cost and less network delay. However, determination of multiple sink locations is still challenging as it considerably affects the routes in some critical applications.

Multiple-sink placement problem is NP-hard under multiple objectives [7]. Traditional optimization algorithms are inefficient with the increasing size of the network. Particle Swarm optimization (PSO) is one of the swarm intelligence based algorithms, which is effective for such NP-hard problem due to ease of implementation, quick convergence, and high quality of solution [8]. Various solutions have been proposed in the literature to tackle the multiple-sink placement problems. Pan et al. [9] have proposed two heuristic algorithms to find the optimal position of sinks. One is for homogeneous application nodes (ANs) and other is for heterogeneous ANs. Homogeneous ANs have same initial energy, data rate, and data transmission rates, where as heterogeneous ANs have different initial energy, data rate, and data transmission rates. However, their algorithms are applicable for small-scale networks, as the size of the network increases the problem become NP-hard. In [10], the authors have assumed that nodes are uniformly deployed in an agriculture field. The total network is divided into clusters to ensure the scalability of the network. They have proposed three approaches for placement of sinks. They are best sink location (BSL), Finding the Minimum number of sinks while maximizing the network lifetime (MSMNL) and finally Minimizing the number of sink nodes for a predefined number of operation period (MSPOP. In [11], the authors have formulated the problem using linear programming (LP) and solved using CPLEX (an integer LP solver) tool for maximizing the network lifetime by distributing the load of among deployed sensors. Some meta-heuristic algorithms based on swarm intelligence have proposed. In [12], the authors have proposed allocation of multiple base stations based on PSO. The authors have placed the multiple-sink with respect to different ANs to increase the lifetime of ANs. Haider et al. [13], have proposed Discrete PSO (DPSO) using local search approach to

solve multiple sink placement problem for minimize the maximum worst case delay and maximize the life of the network. Md Nafees et al. [14], have proposed PSO-based approach for optimal placement of single sink nodes. In this paper, we address the following problem. For a given location of gateways, find the optimized positions of fixed number of sinks using PSO algorithm.

The rest of the paper is organized as follows. The preliminaries of PSO, energy model, network model, and terminologies are provided in Sect. 2. The proposed approach and PSO-based algorithm called as PSO-MSPA are provided in Sect. 3. The simulation results are explained in Sect. 4 followed by the conclusion in Sect. 5.

2 Preliminaries

2.1 Overview of PSO

PSO consists of a predefined number of particles say N_P , called swarm. Each particle provides a potential solution. A particle P_i , $1 \leq i \leq N_P$ has position $X_{i,d}$ and velocity $V_{i,d}$, $1 \leq d \leq D$ in the d^{th} dimension of the search space. The dimension D is same for all particles. A fitness function is used to evaluate each particle for verifying the quality of the solution. In the initialization process of PSO, each particle is assigned with a random position and velocity to move in the search space. During each iteration (generation), each particle finds its own best, i.e., personal best called $Pbest_i$ and the global best called $Gbest$. To reach the global best solution, it uses its personal and global best to update the velocity $V_{i,d}$ and position $X_{i,d}$ using the following equations.

$$V_{i,d}(t + 1) = \omega \times V_{i,d}(t) + c_1 \times \chi_1 \times (X_{Pbest_{i,d}} - X_{i,d}) + c_2 \times \chi_2 \times (X_{Gbest_d} - X_{i,d}) \tag{1}$$

$$X_{i,d}(t + 1) = X_{i,d}(t) + V_{i,d}(t + 1) \tag{2}$$

where $0 < \omega < 1$ is the inertia weight, c_1, c_2 , $0 \leq c_1, c_2 \leq 2$ are the acceleration coefficients and, χ_1, χ_2 , $0 < \chi_1, \chi_2 < 1$ are the randomly generated values. The updating process is repeated until it is reached to an acceptable value of $Gbest$. After getting new updated position, the particle evaluates the fitness function and updates $Pbest_i$ as well as $Gbest$ as follows

$$Pbest_i = \begin{cases} P_i, & \text{if } (Fitness(P_i) < Fitness(Pbest_i)) \\ Pbest_i, & \text{otherwise} \end{cases} \tag{3}$$

$$Gbest = \begin{cases} Pbest_i, & \text{if } (Fitness(Pbest_i) < Fitness(Gbest)) \\ Gbest, & \text{otherwise} \end{cases} \tag{4}$$

2.2 Energy Model

The energy model used in this paper is based on the same radio model as used in [3]. The energy consumption of the node depends on the amount of the data and distance to be sent. In this model, energy consumption of a node is proportional to d^2 when the propagation distance (d) less than the threshold distance d_0 otherwise it is proportional to d^4 [3]. The total energy consumption of each node in the network for transmitting the l -bit data packet is given by the following equations.

$$E_{TX}(l, d) = \begin{cases} l \times E_{elec} + l \times \varepsilon_{fs} \times d^2, & \text{if } d < d_0 \\ l \times E_{elec} + l \times \varepsilon_{mp} \times d^4, & \text{if } d \geq d_0 \end{cases} \quad (5)$$

where E_{elec} the energy is dissipated per bit to run the transmitter or receiver circuit, amplification energy for free space model ε_{fs} and for multipath model ε_{fs} depends on the transmitter amplifier model and d_0 is the threshold transmission distance. In the same way to receive l -bit of data the energy consumed by the receiver is given by

$$E_{RX}(l) = l \times E_{elec} \quad (6)$$

2.3 Network Model and Terminologies

2.3.1 Network Model

The sensors along with some gateway nodes are randomly deployed throughout the sensing field and a node can compute the distance to the other node based on the received signal strength. All the sensor nodes and gateway nodes are assumed to be stationary after deployment. Each node performs sensing periodically and has always data to send to its CH or BS. In the process of transfer of data from gateways to BS, gateways use multi-hop communication. The communication links are wireless, symmetric, and established between the nodes when they are within the communication range of each other.

2.3.2 Terminologies

For better understanding of the proposed algorithm, we first describe some terminologies as follows.

S	The set of sensor nodes, i.e., $S = \{s_1, s_2, s_3, \dots, s_n\}$
G	The set of gateway nodes, i.e., $G = \{g_1, g_2, g_3, \dots, g_m\}$ where, $m < n$
SN	The set of multiple sinks, i.e., $SN = \{SN_1, SN_2, SN_3, \dots, SN_1\}$

d_{\max}	The maximum communication range of a sensor node
R_{\max}	The maximum communication range of a gateway node
d_0	The threshold distance of sensor node
r_0	The threshold distance of a gateway node
E_{s_i}	The initial energy of the sensor node s_i , $1 \leq i \leq n$
E_{g_j}	The current energy of gateways g_j , $1 \leq j \leq m$
$\text{dis}(g_j, \text{SN}_k)$	The distance between a gateway node g_j and sink node SN_k
$\text{HopCount}(g_i, \text{SN}_k)$	The number of hops required to reach to the SN_k from g_i
$\text{Comm}(s_i)$	The set of nodes which are within the communication range of s_i , i.e., $\text{Comm}(s_i) = \{s_j \forall s_j \in S \wedge \text{dis}(s_i, s_j) \leq d_{\max}\}$

In our proposed algorithm we consider network lifetime as the number of rounds until the first gateway death usually referred as First gateway death (FGD) [15].

3 Proposed Approach

Initially sensor nodes can be assigned to the gateways based on distance and energy to form the clusters. After that PSO-based algorithm is run to locate the optimal positions of the multiple-sink with respect to the gateways. In order to increase the lifetime of the gateways, we have considered the Euclidian distance and hop count from the gateways to the sinks. The proposed algorithm is shown in Fig. 1.

3.1 Particle Representation and Initialization

In PSO, a particle represents a complete solution. For multiple-sink placement of the proposed algorithm, it represents optimal positions of the sinks with respect to the gateway nodes. Let $P_i = [X_{i,1}(t), X_{i,2}(t), X_{i,3}(t), \dots, X_{i,D}(t)]$ be the i th particle of the population where each component $X_{i,d}(t) = (x_{id}(t), y_{id}(t))$, $1 \leq i \leq N_p$, $1 \leq d \leq D$, denotes the coordinates of the sink nodes. Then the i th particle can be represented as follows.

$$P_i = [(x_{i1}(t), y_{i1}(t)), (x_{i2}(t), y_{i2}(t)), (x_{i3}(t), y_{i3}(t)), \dots, (x_{id}(t), y_{id}(t))] \quad (7)$$

where N_p denotes the swarm of particles and D represents the number of sinks are supposed to be placed.

Algorithm : PSO based Multiple sink placement
<p>Input: Set of Gateway nodes: $G = \{g_1, g_2, g_3, \dots, g_m\}$; Predefined swarm size: N_p. Number of dimensions of a particle: $D=l$</p> <p>Output: Optimal positions of Sink nodes $SN = \{SN_1, SN_2, SN_3, \dots, SN_l\}$.</p>
<p>Step 1: Initialize particles $P_i, \forall i, j, 1 \leq i \leq N_p, 1 \leq j \leq D=l$, number of SNs $X_{i,j}(0) = (x_{i,j}(0), y_{i,j}(0))$ /* Deployed positions of sinks*/</p> <p>Step 2: for $i = 1$ to N_p do 2.1 Calculate $Fitness(P_i)$ /* Using equation 10*/ 2.2 $Pbest_i = P_i$ end for</p> <p>Step 3: $Gbest = \{Pbest_k Fitness(Pbest_k) = \min(Fitness(Pbest_i), \forall i, 1 \leq i \leq N_p)\}$</p> <p>Step 4: for $t = 1$ to <i>Terminate</i> /*Terminate = Max. number of iterations*/ for $i=1$ to N_p do 4.1 Update velocity and position of P_i using Eqs.(1), (2) 4.2 Calculate $Fitness(P_i)$ 4.3 if $Fitness(P_i) < Fitness(Pbest_i)$ then $Pbest_i = P_i$ end if 4.4: if $Fitness(Pbest_i) < Fitness(Gbest)$ $Gbest = Pbest_i$ end if end for end for end for</p> <p>Step 5: Stop</p>

Fig. 1 PSO-based multiple-sink placement algorithm

3.2 Derivation of Fitness Function

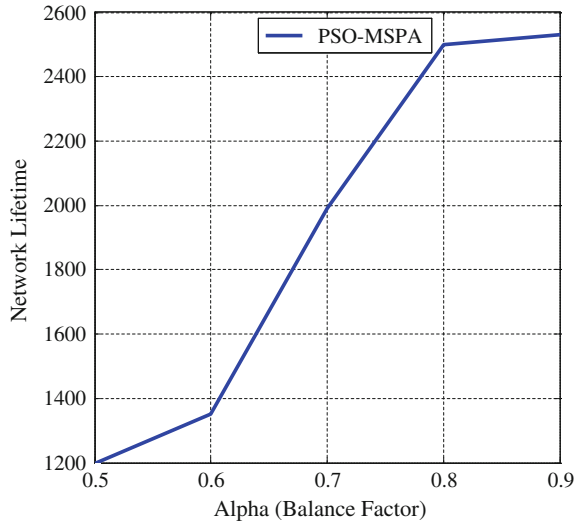
The derivation of fitness function depends on the following parameters:

- (a) *Euclidian distance* It is defined as the minimal distance from a particular gateway to the particular sink in a particle, i.e., in the gateway to the sink communication, gateways consume some energy to send data to the sinks. In order to maximize the lifetime of gateways, we need to reduce the distance between gateways to the sinks.

Objective 1: Minimize

$$f_1 = \sum_{j=1}^m \text{Min}_{k=1}^l \text{dis}(g_j, SN_k) \quad (8)$$

Fig. 2 Lifetime along with α



- (b) *Hop Count* It is the number of hops between the gateways to the sinks. During the communication between gateways to the sinks, gateways use multi-hop routes to transmit its data. In order to decrease the delay we need to minimize the number of hops between gateways and sinks.

Objective 2: Minimize

$$f_2 = \sum_{j=1}^m \text{Min}_{k=1}^l \text{HopCount}(g_j, \text{SN}_k) \tag{9}$$

In our PSO approach, this is judicious to minimize the linear combination of both the above two objective functions instead separately minimize them. Therefore, we use the following fitness function.

Fig. 3 Lifetime along with W_c

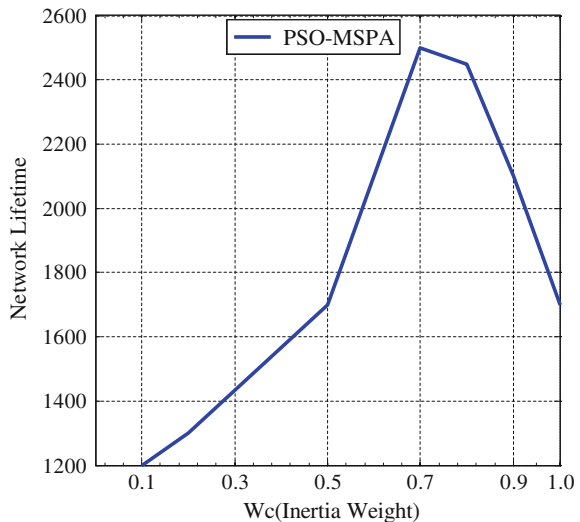
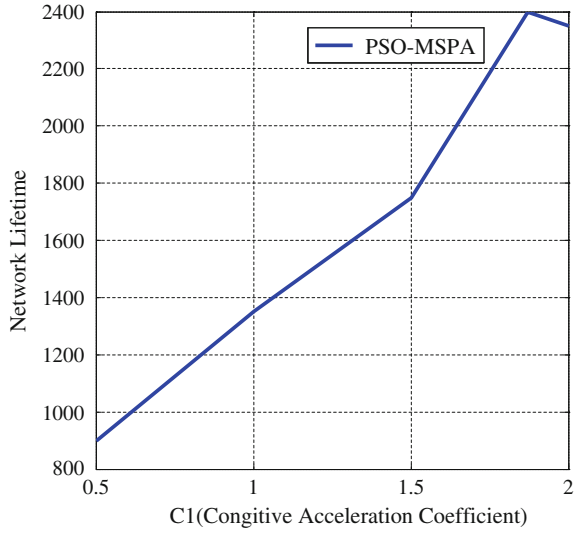


Fig. 4 Lifetime along with C_1



$$\text{Minimize Fitness} = \alpha \times f_1 + (1 - \alpha) \times f_2, \quad 0 < \alpha < 1 \text{ and } 0 < f_1, f_2 < 1 \quad (10)$$

Lifetime of Network $\propto 1/\text{Fitness}$

4 Performance Analysis

The proposed algorithm was tested using C programming and MATLAB (version 7.11) on an Intel core i7 processor with chipset 2600, 3.40 GHZ CPU 2GB RAM running on the platform Microsoft Windows 7. In the simulation, we assumed a WSN deployed in a target area of size $200 \times 200 \text{ m}^2$. There are 300 sensor nodes

Fig. 5 Lifetime along with C_2

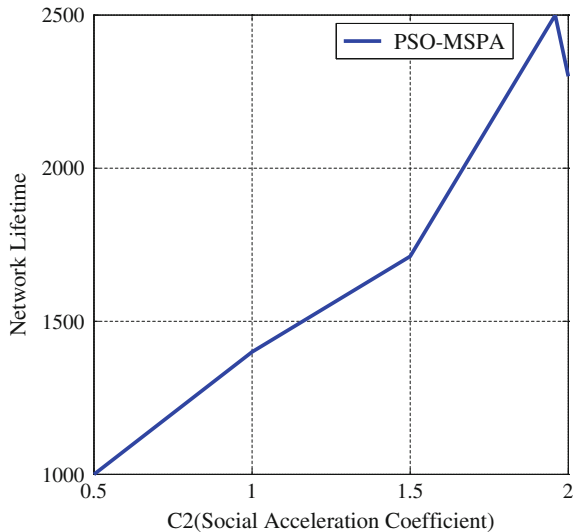
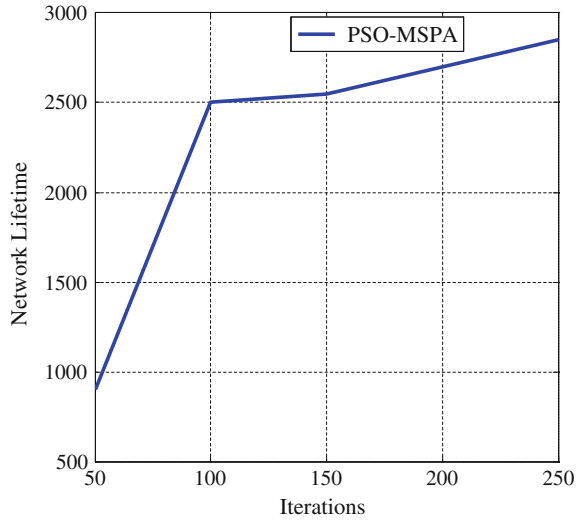


Fig. 6 Lifetime along with iterations



and energy of each sensor node was assumed to be 2J initially. 10 % of the deployed sensor nodes were considered as the total number gateways with initial energy of 10 J. The values of energy parameters were assumed as $E_{elec} = 50$ nJ/bit, $\epsilon_{fs} = 10$ pJ/bit/m², $\epsilon_{mp} = 0.0013$ pJ/bit/m⁴, $d_{max} = 100$ m, $d_0 = 30$ m, packet length = 4000 bits and message size = 500 bits. We ran the proposed algorithm for 100 iterations over various tuning parameters of PSO and found that it produced

Fig. 7 Lifetime along with gateway nodes

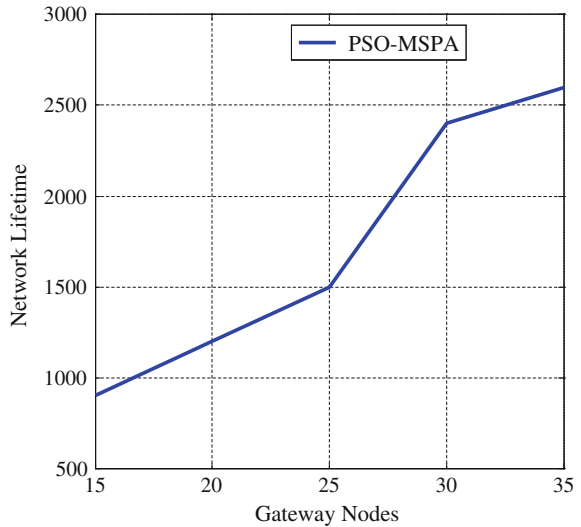
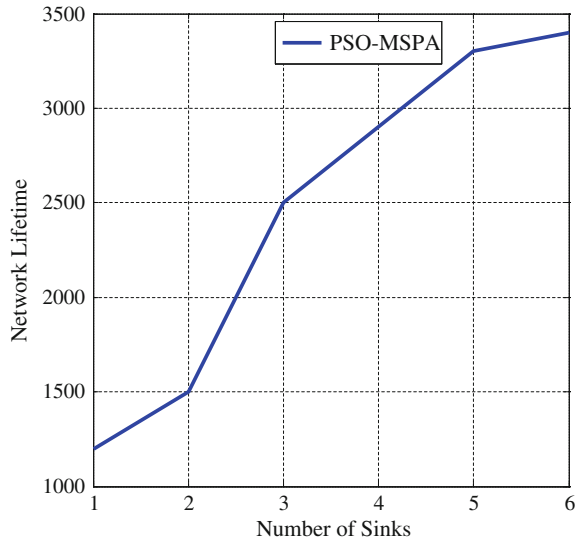


Fig. 8 Lifetime along with sinks



optimal locations of 3 sinks for 50 particles with $C_1 = 1.87$, $C_2 = 1.96$, $\alpha = 0.7$, $W_c = 0.78$, $V_{\max} = 100$, $V_{\min} = -100$.

The experimental results in Fig. 2 explain the variation of network lifetime along with control parameter between two objectives. It can be observed that at $\alpha = 0.8$, the algorithm gives comparably better network lifetime. From Fig. 3, it can be observed that at $W_c = 0.7$, the life time of network reaches highest point. From Figs. 4 and 5, it can be observed that as the value of cognitive and social acceleration coefficients increases the lifetime also increases up to $C_1 = 1.87$ and $C_2 = 1.96$ after that network lifetime fallen down. From Fig. 6, it can be observed that as the number of iterations increases the lifetime of network is significant increase up to 100.

As the number of gateways increases the lifetime of network is significant increase up to 30 gateways, after that it is less, can be observed from Fig. 7. From Fig. 8, it can be observed that as the number of base stations increases the lifetime significant increase up to 3rd SNs. As the number of particles increases the lifetime is significant increase up to 50, it is clear from Fig. 9. PSO-MSPA performs better as shown in Table 1.

Fig. 9 Lifetime along with particles

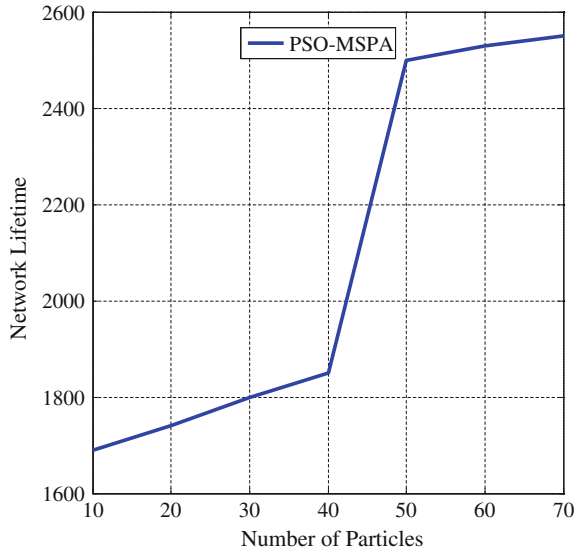


Table 1 Comparison of Network Lifetime

Algorithm	Lifetime
PSO-MSPA	3450
Exhaustive grid search algorithm by Pan [9]	3050

5 Conclusion

In this paper, we have proposed multiple-sink placement algorithm based on PSO using efficient particle representation and fitness function. For the energy efficiency of the proposed algorithm, we have considered the Euclidian distance and hop count. We have extensively tested the PSO-MSPA over various tuning parameters of PSO and WSN scenarios. To show the improvement of PSO-MSPA with an existed exhaustive grid search algorithm, we have calculated network lifetime. It can be observed that the PSO-MSPA out performs the existed algorithm. Our future work will focus on the determination of optimal number of sinks and routing with multiple-sink.

References

1. Akyildiz, I.F., Weilian, S., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Magaz.* **40**(8), 102–114 (2002)
2. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. *Comput. Networks* **52** (12), 2292–2330 (2008)

3. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* **30**(14), 2826–2841 (2007)
4. Kuila, P., Jana, P.K.: Approximation schemes for load balanced clustering in wireless sensor networks. *J. Supercomput.* **68**(1), 87–105 (2014)
5. Kemal, A., Younis, M., Youssef, W.: Positioning of base stations in wireless sensor networks. *IEEE Commun. Magaz.* **45**, 96–102 (2007)
6. Efrat, A., Har-Peled, S., Mitchell, J.S.: Approximation algorithms for two optimal location problems in sensor networks. In: *IEEE 2nd International Conference on Broadband Networks*, pp. 714–723 (2005)
7. Bogdanov, A., Maneva, E., Riesenfeld, S.: Poweraware base station positioning for sensor networks. *Proc. IEEE INFOCOM* **1**, 575–585 (2004)
8. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: *IEEE International Conference on Neural Networks, Perth, Australia*. vol. 4, pp. 1942–1948 (1995)
9. Pan, J., Cai, L., Hou, Y.T., Shi, Y., Shen, S.X.: Optimal base-station locations in two-tiered wireless sensor networks. *IEEE Trans. Mobile Comput.* **4**(5), 458–473 (2007)
10. Oyman, E.I., Ersoy, C.: Multiple sink network design problem in large scale wireless sensor networks. *IEEE Int. Conf. Commun.* **6**, 3663–3667 (2004)
11. Kim, H., Seok, Y., Choi, N., Choi, Y., Kwon, T.: Optimal multi-sink positioning and energy-efficient routing in wireless sensor networks. In: *Proceedings International Conference on Information Networking (ICOIN)*, pp. 264–274 (2005)
12. Hong, T.P., Shiu, G.N.: Allocating multiple base stations under general power consumption by the particle swarm optimization. In: *IEEE Swarm Intelligence Symposium*, pp. 23–28 (2007)
13. Safa, H., El-Hajj, W., Zoubian, H.: Particle swarm optimization based approach to solve the multiple sink placement problem in WSNs. In: *IEEE International Conference on Communications*, pp. 5445–5450 (2012)
14. Rahman, M.N., Matin, M.A.: Efficient algorithm for prolonging network lifetime of wireless sensor networks. *Tsinghua Sci. Technol.* **16**(6), 561–568 (2011)
15. Dietrich, I., Dressler, F.: On the lifetime of wireless sensor networks. *ACM Trans. Sensor Networks* **5**(1), 1–38 (2007)

Denoising of GPS Positioning Data Using Wavelet-Based Hidden Markov Tree

Ch. Mahesh, K. Ravindra and V. Kamakshi Prasad

Abstract Precise position and navigation with GPS is always required for both civil and military applications. The errors and biases associated with navigation will change the positional information from centimeters to several meters. To estimate and mitigate the errors in GPS positioning data, the wavelet transform is most significant technique and proven. The traditional wavelet threshold methods will work to a certain extent but are not useful to estimate the signal levels to the expected level due to their incapability for capturing the joint statistics of the wavelet coefficients. The wavelet-based hidden Markov tree (WHMT) is designed to capture such dependencies by modeling the statistical properties of the wavelet coefficients as well. In this paper, a WHMT is proposed to reduce positioning error of the GPS data. To establish proposed method, the position data are decomposed using wavelets. The obtained wavelet coefficients are subjected to Discrete Wavelet Transform (DWT) as well-proposed WHMT for noise removal. In this proposed methodology, an Expectation Maximization (EM) algorithm used for computing the model parameters. The root-mean square error (RMSE) of proposed method shows better performance comparatively classical DWT.

Keywords GPS · Discrete wavelet transform · HMT · Receiver independent exchange (RINEX)

Ch. Mahesh (✉)

Airports Authority of India, Hyderabad, India
e-mail: maheshaai06@gmail.com

K. Ravindra

Mallareddy Institute of Technology and Science, Hyderabad, India
e-mail: kasa_ravi@yahoo.com

V.K. Prasad

Department of Computer Science JNTUH, Hyderabad, India
e-mail: kamakshiprasad@yahoo.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_59

1 Introduction

The accuracy of positioning service for the simple stand-alone GPS system is severely affected by several types of biases and noises [1]. Those are mainly systematic and random errors. The systematic errors may be due to ionosphere, by the troposphere and clock offset. Random errors may result from satellite orbit, receiver noise, and multipath effect. The systematic errors behave like a low-frequency noise whereas the random errors are typically characterized as high-frequency noise. Filtering out these errors by applying a filter with constant length does not suit if the error behavior is not common in all the levels. If the signal is decomposed into multiscale or bands, and applying threshold gives better results than traditional filtering methods.

It is more than one decade now the wavelet transform emerged as a new tool and gained wide acceptance in the field of statistical signal and image processing. Wavelets are applied in key areas which include signal estimation, detection, classification, and filtering [2–4]. The primary properties like locality and multi-resolution made wavelet transform to become an important tool to reduce the noise and its effectiveness. Donho and Johnston, in 1998, pioneered wavelet threshold methods by grouping the wavelet coefficients as significant and insignificant and are modified by certain specific rules. The optimal threshold estimation is based on the assumption that the wavelet coefficients are sparse. This assumption is invalid in the case of coarser levels leading to error in estimating the threshold. To overcome this, several researchers have proposed Bayesian approach to capture the sparseness of the wavelet coefficients. These shrinkage methods had later been improved by inter-scale and intra-scale correlation of the coefficients. Crouse et al., developed a framework for statistical signal processing based on wavelet domain Hidden Markov Tree (HMT) models [3]. This framework has enabled them to concisely model the non-Gaussian statistics of individual wavelet coefficients and statistical dependencies between coefficients. The applications of wavelets to the GPS signal processing were quite new in 1995, and Coliin and Warrant applied wavelets for GPS cycle slip correction. The authors, Fu and Rizos, widely used the method of MRA in GPS signal processing [5–9]. Authors [10–13] continued to apply wavelets for GPS signal processing for different applications and contributed significant amount of work in the field of denoising. All these methods are used for popular wavelet thresholding/Shrinkage methods.

In this paper, a new approach based on wavelet domain hidden markov tree (WHMT) model is used to mitigate the GPS position errors. To evaluate the performance of proposed method, the WHMT is compared to DWT denoising methods. Section 2 describes problem statement of denoising in GPS positioning data. The classical wavelet denoising and WHMT model are discussed in Sect. 3. Section 4 briefly explains the data collection and analysis of GPS data. Experimental results and analysis achieved in this proposed method are discussed in Sect. 5. A conclusion of the experiment is presented in Sect. 6.

2 Problem Statement

The GPS code observable for the $L1$ single frequency $f_1 = 1575.42$ MHz is [1]

$$\rho_i = r_i^k + b_i - B^k + T_i^k + I_i^k + \varepsilon_{\rho_i} \quad (1)$$

Here ‘ r ’ is the true range between the i th receiver and k th satellite, where ‘ b ’ is the receiver clock bias, ‘ B ’ is the satellite clock bias, ‘ T ’ is the troposphere errors, ‘ I ’ is the ionosphere error, and ‘ ε_{ρ_i} ’ is the noise. The majority of the GPS receiver operates on single frequency and uses noisy code measurements for its simple positioning services. The GPS observables in Eq. (2) are subjected to systematic delays, i.e., ionospheric, tropospheric, and clock difference, and the random errors like receiver noise and multipath noise. The errors are more prominent in low-latitude region and solar days. During the period of disturbance, the receiver suffers from high noise level and the pseudo range noise distributed on a non-Gaussian tails. Traditional wavelet-based denoising methods cannot capture the non-Gaussian statistics nature of the wavelet coefficients.

In general, the denoising problem can be viewed as

$$y_i = f(t_i) + \sigma \varepsilon_i \quad (2)$$

The ‘ ε_i ’ is standard Gaussian while noise (i.i.d). ‘ σ ’ is the noise level may be known or unknown. Here, the goal is to record the under laying function ‘ f ’ from the noisy data ‘ y ’ with small error.

3 Methodology

3.1 Classical Wavelet Denoising

The DWT will decompose data, and it can be represented as pair of high- and low-pass coefficients followed by down sampling by two and iterated on the low-pass output. The outputs of the low-pass filters are the scaling coefficients, and the outputs of the high-pass filter are the wavelet coefficients. The approximated coefficients are processed successively by first down sampled and split further. The detailed coefficients $d_j(t)$ are used to estimate threshold value via median estimator to remove the high-frequency noise components. Median estimator method will estimate threshold for optimum soft threshold that minimizes the Mean Squared error (MSE). In these approaches, a prior distribution is imposed on the wavelet coefficients, which is designed to capture the sparseness of the wavelet expansions that is common to most application. The noise variance $\tilde{\sigma}_\varepsilon^2$ is estimated from the each level by the robust median estimator (MAD) [2].

$$\tilde{\sigma}_e^2 = \frac{\text{median}(|y_i|)}{0.6745} \quad (3)$$

The signal variance $\tilde{\sigma}_x$ is estimated as

$$\tilde{\sigma}_x = \sqrt{\max((\tilde{\sigma}_y^2 - \tilde{\sigma}_e^2), 0)} \quad (4)$$

Since y is modeled as zero mean, $\tilde{\sigma}_y$ can be found empirically as

$$\tilde{\sigma}_y^2 = \frac{1}{n} \sum_{i=1}^n y_i^2 \quad (5)$$

The classical wavelet denoising of signal can be obtained by the following steps.

1. Apply the DWT to the noisy data ('y') to obtain transformed noisy coefficients $w = \text{DWT}(y) = (w_j)_{j \in [1; j+1]}$ where $w_j = \mathfrak{R}^n$, and here \mathfrak{R} is set of noisy coefficients at each level 'n,'
2. Apply suitable thresholding function (I) to the transformed noisy coefficients to get $\hat{w} = I(w)$, and
3. Reconstruct the original signal by applying the inverse DWT to the wavelet coefficients.

The shrinkage technique may vary according to thresholding function and its applicability of wavelet coefficients.

3.2 HMT-Based Denoising

The primary properties of DWT assume that the wavelet coefficients are jointly Gaussian and statistically independent. In general, the actual signal wavelet coefficients have sparseness, and some residual dependency exists between the coefficients. To capture statistical dependencies between coefficients, a hidden markov model was introduced [3]. In this, the hidden state variables are introduced to match the wavelet coefficients, and the dependencies between the hidden state variables are well characterized. For estimating the model parameters of HMT, an EM algorithm is used. In HMT, the nested sets of coefficients are generated at every scale in wavelet decomposition process and represent as state variable across scale. This model connects the hidden states ' S_i ' nodes to observed wavelet coefficients W_i . Each hidden state node represents the mixture state variable (S_i) and wavelet coefficient (' W_i ') as shown in Fig. 1.

(i) Modeling of Wavelet Coefficients

The wavelet coefficient of most real-world signals is sparse: a few wavelet coefficients are large, but most are small. Therefore, we associate each wavelet coefficient

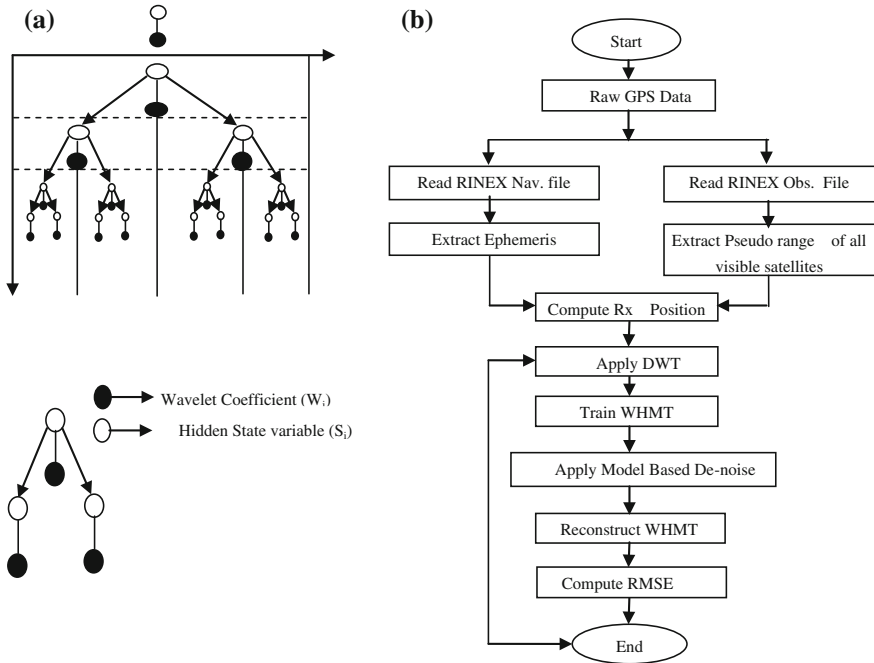


Fig. 1 **a** Signal decomposition hierarchy of HMT model. **b** Flow chart for signal denoising using DWT and WHMT

‘ w_i ’ an unobserved hidden state variable $S_i = \{S, L\}$. The state ‘ S ’ corresponds to a zero-mean, low-variance Gaussian whereas high variance, zero mean corresponds to the large state ‘ L .’ Hence, the Gaussian mixture model appears to be good fit for the distribution of the wavelet coefficient data being one of the states.

Thus, the overall pdf is given by

$$f(w_i) = p(S_i = m)f(w_i/S_i = m) \tag{6}$$

where the conditional probability $f(w_i/S_i = m)$ of the coefficient value ‘ w_i ’ given the state ‘ S_i ’ corresponds to the Gaussian distribution

$$f(w_i/S_i = m) = \frac{1}{\sqrt{2\pi\sigma_{i,m}^2}} \exp\left(-\frac{(w_i - \mu_{i,m})^2}{2\sigma_{i,m}^2}\right) \tag{7}$$

(ii) Estimation of Signal

The HMT model is completely parameterized by two-component mixture of generalized Gaussian for the wavelet coefficients at each scale. The estimation of the true signal wavelet coefficients can be obtained by using of following equation:

$$w_i^\wedge = E[w_i/\theta] = \sum p(S_i = m/w_i, \theta) \frac{\sigma_{i,m}^2}{\sigma_{i,m}^2 + \sigma_n^2} w_i \tag{8}$$

where $p(S_i = m/w_i, \theta)$ is the probability of state ‘ m ’ given the noisy wavelet coefficient ‘ w_i ’ and the model parameters ‘ Θ ’ are computed by the EM algorithm. The variance ‘ $\sigma_{i,m}^2$ ’ common to all coefficients in given scale and the noise variance ‘ σ_n^2 ’ is unknown which in turn estimated through the Median Absolute Deviation (MAD) estimator.

(iii) Implementation

The flow chart for the proposed denoising method is shown in Fig. 1 and summarized as follows:

1. Extract the raw GPS data from single frequency GPS Receiver.
2. Read the RINEX observation and navigation files.
3. Extract ephemeris and observation data of all visible satellites
4. Compute the receiver position using Least Square method.
5. Apply the DWT to the computed position.
6. Reconstruct the DWT coefficients using IDWT
7. Train the obtained wavelet coefficients from DWT using HMT.
8. Apply the model-based denoise to remove the noisy coefficients.
9. Compute the RMSE of original and estimated coefficients.

The proposed method of GPS data processing involves two steps. The first step involves calculation of receiver position at each epoch from GPS data. The second step is to compute receiver positioning data with DWT and the proposed WHMT.

4 Data Collection and Analysis

To verify the proposed method of denoising, two sets of data collected from IGS stations established at IISC Bangalore and Hyderabad as shown in Table 1, which is in RINEX format. One set of data consists of severely affected noisy data on solar eclipse day and other set consists of normal day’s data with different seasons. The collected data-sampling interval of 30 s and total 1024 epochs are taken for

Table 1 Data collection

S. No	Receiver station name	Solar day	Normal day
1	Bangalore (IISC)	15th Jan 2010	15th May and 15th Oct 2010
2	Hyderabad (HYDE)		

computation. The average receiver position of each epoch is computed and shown in Fig. 2. The original and smoothed position coordinates of WHMT method is shown in Figs. 3, 4, and 5. The error plot is shown in Fig. 6.

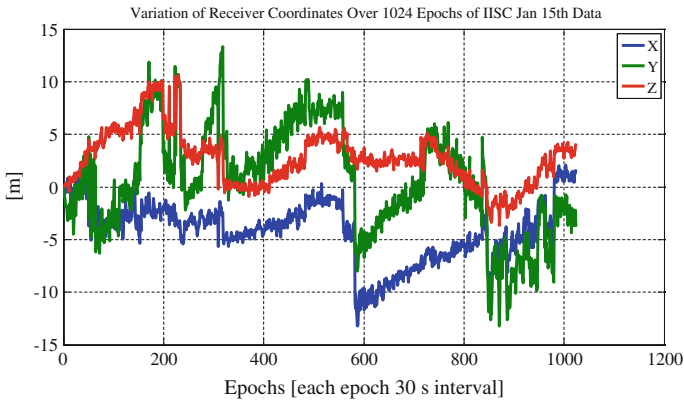


Fig. 2 The original position data signal of IISC Bangalore (15th January 2010)

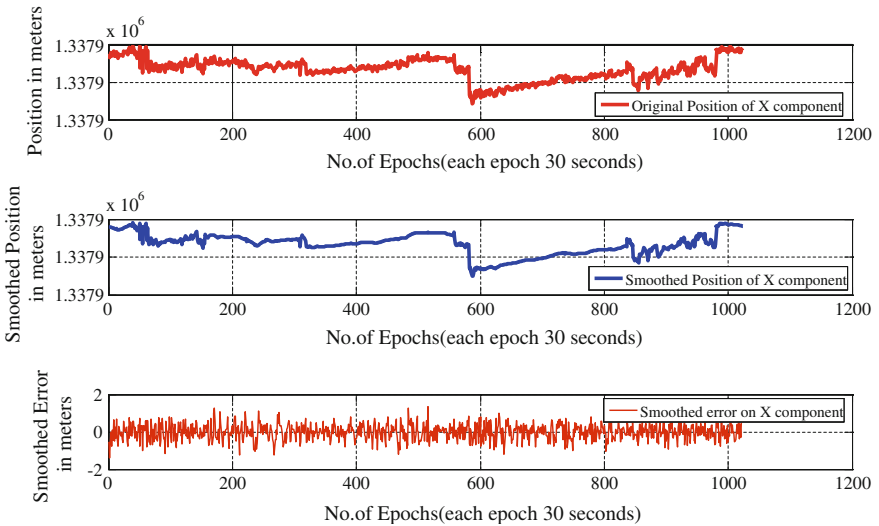


Fig. 3 The original and denoised X position using WHMT

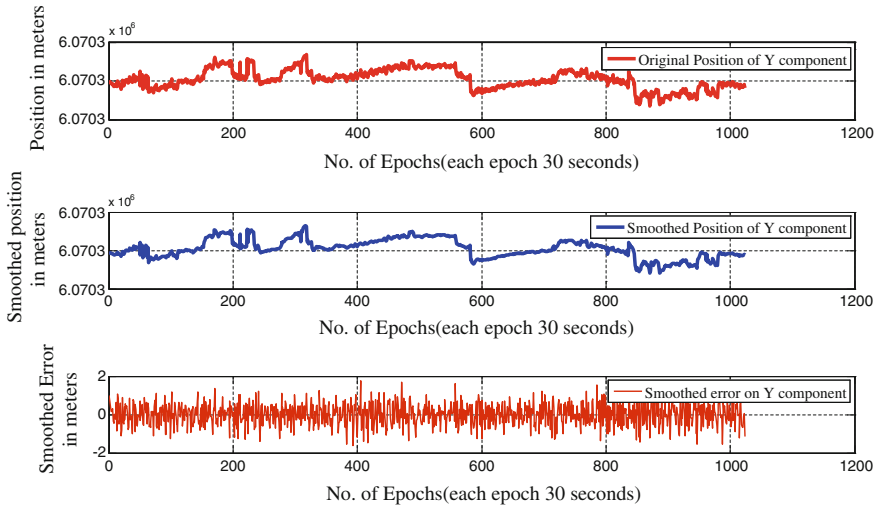


Fig. 4 The original and denoised Z position using WHMT

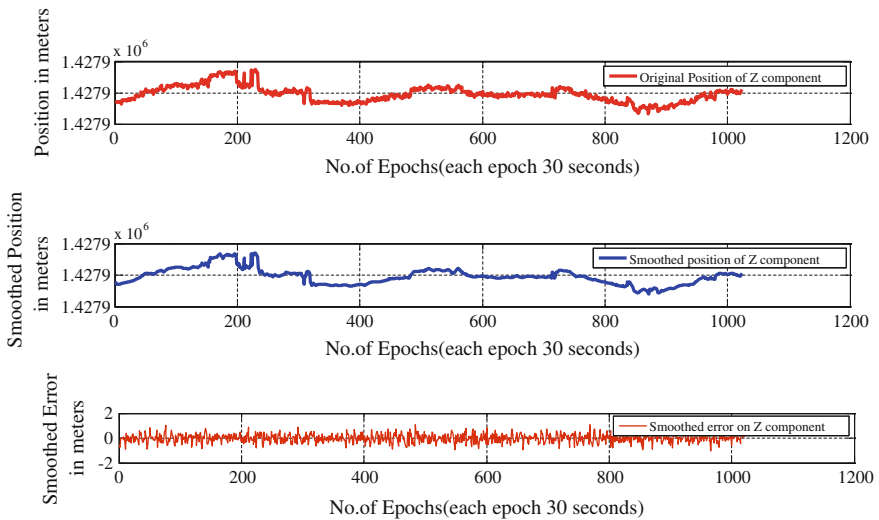


Fig. 5 The original and denoised Z position using WHMT

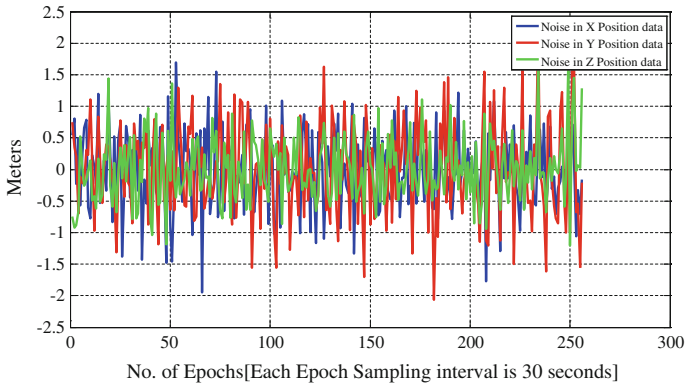


Fig. 6 The noise present in GPS positioning data

5 Results and Discussions

To evaluate the performance of the proposed method, WHMT is compared with classical DWT. In this analysis, a different seasonal data is collected and evaluated. The quality measure of this algorithm is RMSE. Different wavelet base functions considered and tested. Table 2 shows the root-mean squared error (RMSE) of the processed GPS position coordinates data using traditional denoising method DWT, as well as WHMT. As compared with the RMSE of DWT to WHMT, the WHMT shows significant improvement. The results are clearly indicating that the proposed method is best suited and promising for accurate position. In this analysis, it has

Table 2 Position smoothing using DWT and WHMT

Station name	Wavelets	Error parameter	Position smoothing using DWT			Position smoothing using WHMT		
			On X Pos	On Y Pos	On Z Pos	On X Pos	On Y Pos	On Z Pos
Bangalore (IISC) (1024 Epochs)	Db6	Min	-2.3094	4.0621	-1.8301	-1.3750	-1.6131	-1.0520
		Max	2.0885	-4.0714	2.1763	1.3707	-1.7468	1.0961
		Mean	0.4670	1.3647	0.2711	0.1808	0.3384	0.1228
		RMSE	0.6834	1.1682	0.5207	0.4252	0.5817	0.3504
	Symmlet6	Min	-2.2435	-4.0524	-1.7061	-1.4856	-1.5077	-1.1221
		Max	2.3823	4.7679	1.5965	1.3537	1.9629	1.0044
		Mean	0.4533	1.4156	0.2789	0.1846	0.3283	0.1215
		RMSE	0.6733	1.1898	0.5281	0.4297	0.5730	0.3486
	Coiflet5	Min	-2.5125	-5.1005	-1.9578	-1.1558	-1.5957	-1.1659
		Max	2.3033	4.8904	2.3462	1.2800	1.6327	1.1974
		Mean	0.4720	1.5386	0.2884	0.1722	0.3182	0.1114
		RMSE	0.6870	1.2404	0.5370	0.4150	0.5641	0.3338

been observed that the Y coordinate is noisy than other two coordinates. Among three coordinates, the coordinate Y is noisy. Further, with all wavelet basis function, the `coeflet5` gives better performance.

6 Conclusions

In this paper, the traditional DWT and proposed WHMT are used for removal of systematic and random errors of GPS positional data. The experimental result on collected data demonstrates that the proposed method can effectively remove the GPS errors and biases. This method well suited critical aviation applications like GAGAN of Indian SBAS. The median estimator (MAD) is simple and robust, and requires less computation time to estimate threshold. Therefore, here it is considered and used to denoise the position data. In future analysis, the proposed method can be tested with non-orthogonal wavelet families as well as other available threshold techniques.

Acknowledgments The author Ch Mahesh expresses sincere thanks to S. Nandulal, Sr. Manager (CNS), GAGAN, AAI, and R. Pavan Kumar Reddy for their continuous support and providing valuable suggestions for this project.

References

1. Misra, P., Enge, P.: Global Positioning system(GPS), signals, measurements and performance. Ganga-Jamun, Press, P.O. Box-633, Lincoln, MA 01773. ISBN:0-9709544-1-7
2. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81**, 425–455 (1994)
3. Crouse, M.S., Nowak, R.D., Baraniuk, R.G.: Wavelet-based statistical signal processing using hidden markov modesles. *IEEE Trans. Signal Proc.* **46**, 886–902 (1998)
4. Romberg, J.K., Choi, H., Baraniuk, R.G.: Bayesian tree –structured image modeling using wavelet—domain hidden markov models. In: *Proceedings of SPIE*, vol. 3816, pp. 31–44. Denvor, CO, July 1999
5. Collin, F., Warrant, R.: Applications of the wavelet transform for GPS cycle slip correction and comparison with kalamanfilter. *Manuscript Geodatic* (1995)
6. Fu, W.X., Rizos, C.: The applications of wavelets to GPS signal processings. In: *10th International Technical Meeting of the Satellite Division of the U.S Insitute of Navigation*, pp. 1385–1388. Kansas City, Missouri, 16–19 Sept. 1997
7. Xia, L.: Approach for multipath reduction using wavelet algorithm. In: *Proceedings of the International Technical Meeting of the Satellite Division of the Institute of Naviagation*, vol. 1, pp. 2134–2143. Kansas City, 11–14 Sept. 2001
8. Satirpod, C., Ogeja, J., Wang, R.: An approach to GPS analysis incorporating wavelet decomposition. *Artificial Satellite* **36**, 27–25 (2001)
9. Satirpod, C., Ogeja, J., Wang, R.: GPS analysis with the aids of wavelets. In: *Proceedings of the International Symposium Satellite Navigation Artificial Satellite*, vol. 36, pp. 27–25, 200ion Technology & Applications, Canberra, Austrelia (2001)

10. Mosavi, M.R., EmmamGholipour, I.: Denoising of GPS receivers positioning data using wavelet transform and bilateral filtering. *Wireless Personal communications*, pp. 2295–2312 (2013)
11. Nassar, S., El Sheimy, N.: Wavelet analysis for improving INS/DGPS navigation accuracy. *J. Navig.* **58**(1), 119134 (2005)
12. Kang, C.W., Kang, C.H., Park, C.G.: Wavelet denoising technique for improvement of the low cost MEMSGPS integrated system. In: *International Symposium on GPS/GNSS*, p. 2628 (2010)
13. Huang, D., Ding, X., Chen, Y., Gong, T., Xiong, Y.: Wavelet filters based separation of gps multipath effects and engineering structure vibrations. *Acta Geodaetica Et Carographic Sinica* **1**, 7 (2001)

Parent Selection Based on Link Quality Estimation in WSN

Siba Mitra, Sarbani Roy and Ajanta Das

Abstract Wireless sensor network (WSN) is a dynamic and infra-structure less communication system. These are used for any specific target monitoring. The meaningful data returned by the sensor nodes are used for significant decision making. So the quality of data plays a significant role in correctness of decision. Resource scarcity of WSN makes it vulnerable and prone to faults. Faults in WSN are unintended errors, which may lead to wrong decision making. Moreover, due to faults, the reliability and performance of the WSN may get affected; hence, fault tolerance is necessary over here. This paper aims to design a built-in learning-based link quality estimation approach for WSN. This will enable each sensor node to select an appropriate parent through a good-quality link for forwarding data to the base station (BS). Parameters for link quality estimation over here mainly include received signal strength (RSS), signal to interference plus noise ratio (SINR), and packet reception rate (PRR). Simulation results along with discussion are also presented here.

Keywords Wireless sensor network · Link quality estimation · Parent selection

1 Introduction

WSN is used for monitoring a target area, which can be either outdoor or indoor. The sensor nodes are deployed on the target, where they sense any ambient signal and transmit data that to the nearest BS. WSN is considered to be

S. Mitra (✉) · A. Das

Department of Computer Science & Engineering,
Birla Institute of Technology, Mesra, Kolkata Campus, India
e-mail: sibamitra@bitmesra.ac.in

A. Das

e-mail: ajantadas@bitmesra.ac.in

S. Roy

Department of Computer Science & Engineering, Jadavpur University, Kolkata, India
e-mail: sarbani.roy@cse.jdvu.ac.in

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_60

resource-constrained system due to minimal storage, low-power processing, routing, etc.; moreover, the environments, in which it is deployed, are unattended and unleashed so the network is prone to possible types of fault like sensing fault and communication fault. Faulty data communication may result in wrong decision making. So link quality, among the nodes, is to be investigated dynamically, and the routing decisions need to use those connectivity statistics to achieve reliability. Moreover, interference from neighboring nodes may disturb a node from receiving correct data from a sender. Some of the fault management schemes existent is presented vividly in [1]. It is unrealistic to have a centralized fault manager for any WSN, as it becomes very expensive; hence, the WSN should carry out self evaluation and self healing for any occurrence of fault. In our previous work [2], a fault tolerant framework for WSN was proposed, where a sensing and communication fault detection algorithm was presented. This work mainly targets to estimate the link quality hence detect and prevent transmission faults in the network, such that a sensor node can select an appropriate parent, who can carry the data through a good-quality path to the BS. The parameters considered are interference, channel capacity, RSS, SINR, PRR, and expected number of transmissions (ETX). A link quality estimation function along with a technique is proposed, which evaluates a link quality among two nodes and selects suitable parent for data forwarding to BS.

Section 2 presents related work followed by Sect. 3, where fault management module is discussed. In Sect. 4, parent selection technique is mentioned; results are discussed in Sect. 5, and Sect. 6 concludes the article.

2 Related Work

There are various link quality estimation metrics, and some of them are link quality indicator (LQI), ETX, routed number of packets (RNP), etc. A study of various link quality estimators (LQE) is presented in [3]. The study in [4] presents various LQEs like moving average, packet loss/success interval, EWMA, etc. They have introduced WMEWMA, where the ratio of actually received packets to the expected packets, at any instance, is the basis of evaluation. Qin et al. in [5] have addressed and investigated on certain issues like noise, LQI, and packet error rate. They have derived a model to describe the link quality variety in WSN at different states. In [6], researchers have analyzed packet loss and bit error and have proposed an improved LQI estimator which considers mean LQI values of the unsuccessfully received packets.

In our current work, the main aim is to design a distributed link quality estimation technique; on the basis of the same, a sensor node selects an appropriate parent for forwarding data toward BS. Transmission of data through parent is based on residual energy of the node. Moreover, a predictor is designed, which monitors energy consumption of the node. The next section explains the role of the sensor fault predictor hence proposes an approach for link quality estimation.

3 Fault Management Module

In our previous work, fault tolerant framework for distributed WSN was proposed also presented in [2]. The predictor performs a continuous self-monitoring and forecasts faults. The link quality estimator is an integral part of both sensor fault predictor and sensor fault detector. It proactively selects a parent through a good link for a sender for relaying data toward BS.

3.1 Sensor Fault Predictor

Battery depletion is a continuous process for a node. Energy conservation in WSN is implemented by using duty cycling [7]. At any point of time, it is difficult to predict sensing unit failure, but node death due to energy shortage may be forecasted and that is the role of the predictor. A packet reception failure is inferred from either non-receipt of acknowledgement (ACK) or a receipt of negative ACK (NACK), and automatic repeat request (ARQ) is also invoked on that basis. If the energy spent by a sensor node is below 70 %, then the ARQ will be ACK based; otherwise, it will be NACK based since it is energy efficient. Again if the residual energy goes below a threshold (assumed 10 %), then the in-built fault predictor of the sensor node invokes two activities; first it broadcasts a message, informing its low-energy state. Secondly, it broadcasts some probe packets for a high-residual energy sensor node to which it can offload its responsibility for the time being. If it gets a positive response, then it offloads its role to the same. And finally the fault predictor bumps up the priority of the sleep process, and the node forcefully goes to sleep.

3.2 Link Estimation for Network Fault Detection and Diagnosis

WSN is responsible for reporting to BS with meaningful data. Nodes which are direct neighbors to the BS can directly transmit data to the BS. But the remaining nodes may need some parent, which can help it in transmitting its data toward the BS. Data can be transmitted through various possible links to the BS; but depending upon the traffic and the quality of the available link, a specific link and a parent are selected to forward data. Some significant parent selection factors are *ETX* to the BS through that parent, *link quality* between the node and the parent, *energy expenditure* to send data to the parent, and *distance* between the node and the parent. In real time, there can be a scenario, where some nodes have more than one susceptible parent for relaying data to the BS, and it has to select one. In this situation,

the node should select a less economic parent to pass the data toward BS. The technique mentioned later, helps select an appropriate parent for these ambiguous nodes.

4 Proposed Approach for Parent Selection

Considering a set of sensor nodes, $S = \{S_1, S_2 \dots S_n\}$, monitoring a target area, the WSN can be represented as a graph $G(S, E)$, where E is the set of links or edges. Link exists among sensor nodes if S_i and S_j remain within each others' transmission range. Each node maintains a neighborhood table as mentioned in Table 1. If BS is a member of the table, then it can be a susceptible parent for other nodes, which can forward data from a sensor node toward the BS. The assumptions that are considered for the research are

- Link Utilization (LU) and RSS as given in Eqs. (1) and (2), respectively, are obtained from [8], where W , t_f , and t_p are number of frames delivered, time elapsed for one frame/packet transit, and data propagation time, respectively; again d is distance traversed by the signal, and λ is wavelength of signal.
- WSN is using sliding window flow control protocol. Since consolidated feedback is used, this is energy efficient. According to [9], a positive ACK confirms packet reception by a receiver, and NACK confirms the reception failure of a packet.

Table 1 Notations used

Symbol	Meaning
S_i .NBR	Neighbor table
S_i .Num_NBR	Number of neighbors
S_i .RSS	Current RSS
S_i .SINR	Current SINR
S_i .PRR	Current PRR
S_i .CC	Channel capacity
S_i .RSS_prev	Last RSS value
S_i .LU_prev	Previous LU value
S_i . E_{TOTAL}	Sum of total energy
S_i . E_{SPENT}	Already consumed energy
S_i . $E_{RESIDUAL}$	Left over energy
S_i . E_{INIT}	Initial energy
S_i .ETX	Current ETX
S_i .EEC	Current EEC
S_i .ENR	Current ENR
Th_{SLEEP}	Energy Threshold for sleep
Th_{ACTIVE}	Energy Threshold for active
Th_{PRR}	Threshold for PRR

Duty cycling, [7] in sensor nodes, is important for energy conservation. The sensors go to sleep/idle mode while not working and becomes active when required. Table 1 mainly represents the notations used in the technique explained later.

$$LU = \frac{W \times t_f}{t_f + 2t_p} \quad (1)$$

$$RSS = -\left(20 \log_{10} \frac{d}{\lambda}\right) + (20 \log(4\pi)) \quad (2)$$

$$\text{Channel Capacity} = B \log_{10}(1 + \text{SINR}). \quad (3)$$

4.1 Link Quality Estimation Technique

This section mainly presents a link quality estimation problem. The objective function for link quality estimation is multivariate and contains the parameters like energy consumed by a node, ETX, and expected retransmission count. The objective function proposed here is given by, $f(\text{LQE}) = \omega_1 \text{EEC} + \omega_2 \text{ETX} + \omega_3 \text{ENR}$, where EEC is expected energy consumption for the current data transmission, ETX is expected number of transmissions, and ENR is expected number of retransmissions. The coefficients are the weights provided to each parameter. But the sum of all weights is equal to unity, i.e., $\omega_1 + \omega_2 + \omega_3 = 1$. The above function needs to be minimized since all the inclusive parameters should preferably be minimal during a transmission. Some of the factors affecting the above functions are the Euclidean distance among the two nodes, which should be less than the communication range; the channel capacity cannot be exceeded than as mentioned in Eq. (3). Moreover, the current SINR value for each node is dependent on three factors: RSS of the current node, aggregate RSS of the signals at S_j which are transmitted by all other nodes but not S_i , and the ambient noise [10]. Another significant issue is the residual energy of each node and that should certainly be less than the difference of the initial and spent energy at any current time.

4.2 Link Quality Estimation-Based Parent Selection

This section explains the steps for parent selection. This technique is such designed that it can estimate the link quality from a sender's perspective; moreover, it estimates the energy level of the node, and on that basis, it performs all the tasks.

For each node in the WSN, E_{RESIDUAL} and E_{SPENT} are computed using the technique presented in [2]. At any instance, if E_{RESIDUAL} crosses the Th_{SLEEP} limit, then it is forcefully send to sleep mode by the predictor; otherwise, the node generates a neighbor table and finds a neighbor $S_j \in S_i.\text{NBR}$ with minimum ETX

value and then computes the subsequent S_j .RSS and S_j .SINR. If both RSS and SINR are low, then definitely the link quality is bad, and if both the values are high then the link quality is ideal. Again if S_j .SINR is low but S_j .RSS is high, then the link is suffering of huge traffic burst and sender node has good amount of radio interference, hence link quality is likely poor. Again when S_j .SINR is high or good but the RSS is not strong, then the link is likely good. Now if the neighbor with minimum ETX is not ideal then the same process is carried out for next neighbor node. In the next phase, S_j .PRR is studied for S_i being sender, and if the PRR is above Th_{PRR} , (considered as 90 %), then S_j can be a susceptible suitable parent. Otherwise, next neighbor may be checked. Once a susceptible suitable parent is obtained, the next step is to compute the $f(LQE)$ value and select the node with minimum value as parent. Moreover if S_j . $E_{RESIDUAL}$ is greater than Th_{ACTIVE} , then depending upon S_i .CC, S_i can start transmission through S_j . If S_j .ENR, for a successful packet reception, is high, then energy consumption will be more, which can be expensive for a WSN. Finally after evaluating the situation, a sender estimates a good-quality link and finds a suitable parent to forward data toward BS.

The results and corresponding discussion are presented in the next section.

5 Results and Discussion

The proposed work was simulated using MATLAB R2010b version 7.11.0.584. The evaluation process used the sensor node specification as mentioned in [11]. The simulation environment is mentioned below in Table 2. After the nodes were deployed, their RSS and interference were computed. The total number of interferers for each node was evaluated to know the traffic situation, and it is shown in Fig. 1. The results generated for RSS range from -62.64 to -82.75 dBm and the interference ranged from -84.5 to -91.57 dBm. RSS value below -75 dBm exhibits poor-quality signal. A surrounding noise of -94 dBm was considered during the simulation process.

Corresponding to SINR, different PRR values were generated, and a correlation of them is also presented in Fig. 2. The SINR-based PRR calculation was based on

Table 2 Simulation environment

Parameter	Value
Nodes deployed	30–50
Communication range	25–40 m
Interference range	50–80 m
Frequency range	2.4–2.48 GHz
Area covered	$100 \times 100 \text{ m}^2$
Initial energy of a node	21,600 J
Data packet size	250 bytes
Control packet size	20 bytes

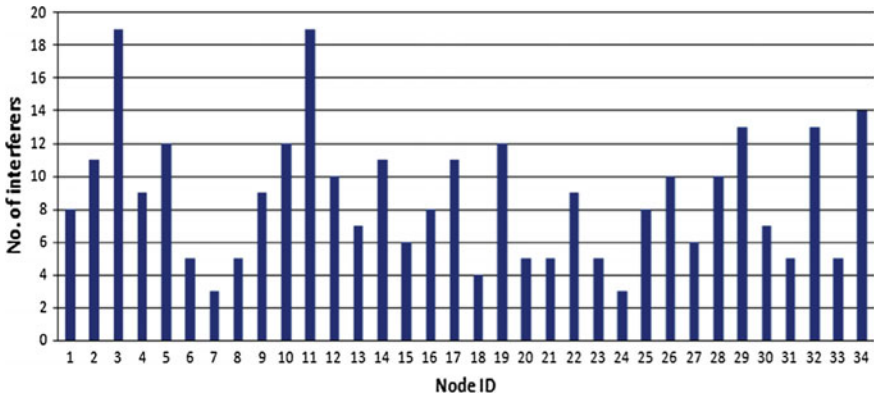


Fig. 1 Interferers to each sensor nodes

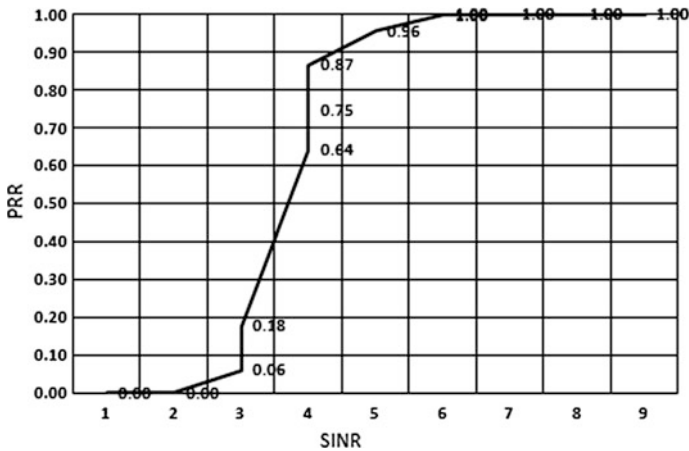


Fig. 2 PRR corresponding to various SINR values

standard SINR-PRR relation provided in [12]. The channel capacity utilization also varies as the SINR varies. Higher SINR value exhibited better channel utilization as presented in Fig. 3. Evaluation of the objective function $f(LQE)$ was done on the basis of EEC, ETX, and ENR for each node. Energy consumption for a data packet transmission is approximately 1.55 mJ, as computed in [2], and for a control, packet is around 0.1242 mJ. The retransmission was considered to be 10–25 %, and a maximum of 4-hop neighbor node to BS was taken into account. The set of values of $f(LQE) = \{4.47, 4.80, 5.13, 5.79, 6.12, 7.11, 7.77, 8.10, 8.43, 9.09, 9.75\}$, providing equal weight of 0.33 to each of the parameter. The node with minimum $f(LQE)$ was considered as suitable parent. Two schemes were used for parent selection of nodes: one was distance and RSS based, and the other one was with

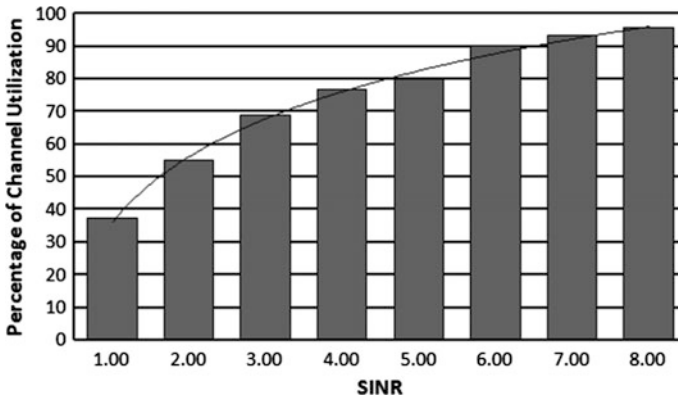


Fig. 3 Percentage of *channel utilization* based on SINR

Table 3 Comparison of RSS and $f(LQE)$ -based parent selection

Node id	Parent selection based on RSS	ENR (%)	$f(LQE)$ -based parent selection	ENR (%)
3	35	20	19	10
7	16	22	4	15
19	35	25	22	12
26	22	18	19	12

respect to the $f(LQE)$ value. Since $f(LQE)$ considers retransmission count as one of the parameter, hence, it was noticed that it is selecting a node as a parent whose ENR is less. Four random sensor nodes having more than one susceptible parent were identified, which were in ambiguity as to select which one as its parent. Both RSS and our proposed $f(LQE)$ were considered during evaluation. It was observed that the sensor node ID 3 is selecting node ID 19 as its parent on the basis of $f(LQE)$; since its RSS-based selected parent node ID 35, which is having a higher ENR. Three other cases along with the comparisons are also mentioned in Table 3. Moreover, due to extra retransmissions energy, consumption of sender was high.

6 Conclusion

This section concludes the article stating about the future work of this research. This work mainly proposes a link estimation technique, based on which a sensor node of WSN selects an appropriate parent for data forwarding toward BS. The parameters for evaluation of the link are RSS, SINR, PRR, and channel capacity. Here two approaches are used for parent selection: one is distance and RSS based and the

other is our estimator $f(\text{LQE})$ based, and a comparison is also shown. The framework already proposed by us in [2] has a module of fault recovery, and the future scope of this work directs toward the implementation of the same.

References

1. Mitra, S., De Sarkar, A., Roy, S.: A review of fault management system in wireless sensor network. In: Proceedings of International Information Technology Conference, (2012). CUBE published by ACM, Kolkata, pp. 144–148
2. Mitra, S., De Sarkar, A.: Energy aware fault tolerant framework in wireless sensor network. In: Proceedings of 2014 Applications and Innovations in Mobile Computing (AIMoC) published by IEEEExplore, pp. 139–145 (2014)
3. Baccour, N., Kouba, A., Mottola, L., Zuniga, M.A., Youssef, H., Boano, C.A., Alves, M.: Radio link quality estimation in wireless sensor networks: a survey. Published in J. ACM Trans. Sensor Networks **8**(4), 34 (2012)
4. Woo, A., Tong, T., Culler, D.: Taming the underlying challenges of reliable multihop routing in sensor networks. In: Proceedings of ACM SenSys 03, Los Angeles (2003)
5. Qin, F., Dai, X., Mitchell, J.E.: Effective-SNR estimation for wireless sensor network using Kalman filter. J. Ad Hoc Networks, Elsevier **11**(3), 944–958 (2013)
6. Chen, W., Luo, J.: An improved LQI-based link quality estimation mechanism for wireless sensor networks. Published in Proceedings International Conference on Computer Science and Service System (CSSS), pp. 74–79 (2014)
7. Anastasi, G., Conti, M., Francesco, M.D., Passarella, A.: Energy conservation in wireless sensor networks: a survey. Ad Hoc Networks J. Published by Elsevier **7**(3), 537–568 (2009)
8. Gupta, P.C.: Data Communications & Computer Networks. Prentice Hall of India Pvt. Ltd., Eastern Economy Edition (2006)
9. Karl, H., Willig, A.: Protocols and Architectures for Wireless Sensor Networks. Student Edition; John Wiley & Sons Inc., (2014). Reprint; ISBN:978-81-265-3369-5
10. Li, X.Y.: Wireless Ad Hoc and Sensor Networks. 1st edn. Cambridge University Press (2008)
11. http://www.dinesgroup.org/projects/images/pdf_files/iris_datasheet.pdf. Accessed 8th Dec 2013
12. Son, D., Krishnamachari, B., Heidemann, J.: Experimental study of concurrent transmission in wireless sensor networks. In: Proceedings of SenSys 06' (2006)

Enhanced Higher Order Orthogonal Iteration Algorithm for Student Performance Prediction

Prema Nedungadi and T.K. Smruthy

Abstract Predicting Student Performance is the process that predicts the successful completion of a task by a student. Such systems may be modeled using a three-mode tensor where the three entities are user, skill, and task. Recommendation systems have been implemented using Dimensionality reduction techniques like Higher Order Singular Value Decomposition (HOSVD) combined with Kernel smoothing techniques to bring out good results. Higher Order Orthogonal Iteration (HOOI) algorithms have also been used in recommendation systems to bring out the relationship between the three entities, but the prediction results would be largely affected by the sparseness in the tensor model. In this paper, we propose a generic enhancement to HOOI algorithm by combining it with Kernel smoothing techniques. We perform an experimental comparison of the three techniques using an ITS dataset and show that our proposed method improves the prediction for larger datasets.

Keywords Recommendation systems • HOSVD • Higher order orthogonal iteration algorithm • Kernel smoothing • Tensors

1 Introduction

Task recommendation is the process by which new tasks are recommended to users by predicting student performance. There are three main entities in modeling a task recommendation system namely user, skill, and task. Traditional recommender systems do not consider these three features together [1]. They use techniques like

P. Nedungadi (✉)

Amrita CREATE, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India

e-mail: prema@amrita.edu

T.K. Smruthy

Department of Computer Science, Amrita Vishwa Vidyapeetham, Kollam, Kerala, India

e-mail: smruthykrishnan@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_61

Collaborative Filtering (CF) applied to only two-dimensional data, users, and items. In systems modeled using all the three features, they were split into pair relations. Thus, the total interactions between the three features were missing. Later, recommendation systems were developed modeling the three entities using multi-way arrays called tensors [2]. Decomposition techniques like Higher Order Singular Value Decomposition (HOSVD) were applied on the tensors. Truncation of the factor matrices obtained as a result of HOSVD to find the low-rank approximation tensor proved sub-optimal in terms of the Frobenius norm [3]. HOSVD was combined with Kernel Smoothing Techniques, and these were shown to give more accurate results. Also, Higher Order Orthogonal Iteration (HOOI) algorithms were proposed to improvise on low-rank approximation [4] but sparseness affected their performance. Thus, Kernel smoothing was used which resulted in the mapping of the data into higher dimensions, and thereby, the data could be better separated and structured. In this paper, we aim to recommend tasks to users combining both the Higher Order Orthogonal Iterative procedure with the Kernel Smoothing Technique. The combination has proved to give out more accurate recommendations to users than the separate models. Experimental evaluation was performed using real-time datasets, and the proposed approach outperformed the previous models.

2 Recommendation Systems

2.1 Motivation

Students need to look for guidance from teachers and others when it comes to choosing tasks they are likely to solve. In order to cater to this need, many information and recommendation strategies have been developed. Task Recommendation systems are one among these that recommend tasks with high chances of successful completion to each student. There are three main entities in developing such systems: user, skill, and task. Recommendation system aims to derive a quadruplet (u, s, t, p) that provides a likeliness measure of interaction between the three features. ' p ' measures the likeliness of the user ' u ' with skill ' s ' solving task ' t .' Thus, such a system needs to be modeled considering the interaction among the three dimensions together. Also, modeling such systems will result in sparseness. So we need to develop algorithms to alleviate the effect of sparseness on the approximations.

2.2 Related Work

Breese et al. proposed a model based on collaborative filtering (CF) but the model failed to capture the relationship between the entities [1]. Later many matrix factorization model was proposed that overcame the limitations of the previous model,

but it did not address user personalization [5]. The interrelationship between entities and user personalization was together addressed using CubeSVD algorithm and division algorithm [6]. Also, multi-task learning [7] methods for collaborative filtering was proposed, which built a multi-task regression model for rating prediction. Systems were proposed to include a third attribute for user personalization to standard CF algorithms reducing three-dimensional relations to two-dimensional relations and then later applying fusion algorithms to associate correlations [8]. But this model lacked accuracy. A comparison of three classes of algorithms namely CF adaptations based on projections, page rank algorithm, and some simpler methods showed that FolkRank algorithm performs better [9]. However, we needed a model that preserves the three-dimensional relation as such and generates recommendation. Later, data was modeled using tensors—multi-way arrays. Various tensor decomposition algorithms were developed for analyzing the high-dimensional structures, and the approximations drawn from these decompositions were analyzed to reach conclusions.

There are a lot of models addressing the various tensor decomposition algorithms [10, 11]. Analysis proved that only HOSVD provided a unique solution [12]. Later, a research on the user web interest was performed by analyzing the click data and modeling it as a third-order tensor [13]. This used the concept of HOSVD to perform tensor decomposition. The data modeled for the above-mentioned researches were highly sparse due to the presence of huge number of user, item, and contexts, but the mapping between each of these combinations was very few. HOSVD with kernel functions was combined to address sparse data and the combination proved to improvise the traditional HOSVD approach especially on sparse data [14]. Truncation was done on each of the factor matrices that were obtained as a result of performing SVD to find the low-rank approximations [3]. This provided a solution that was sub-optimal in terms of the Frobenius norm. So we need a method to improvise on the rank approximations. An empirical analysis of four tensor decomposition algorithms was performed, and Higher Order Orthogonal Iteration algorithm was recommended. The Tensor Toolbox was introduced for efficient MATLAB computations [15]. This explained various functions for computations with both sparse and dense tensors. A unified model for social tagging was introduced giving out all three types of recommendation [2]. This used HOSVD with kernel functions and this approach was better than the state-of-the art methods. HOSVD is limited in terms of Frobenius norm. HOOI also provides a good orthogonal decomposition but the solution is affected by data sparseness. So we go for a combined model that alleviates the sparseness and is optimized in terms of Frobenius norm.

2.3 General Block Diagram with Different Approaches

Figure 1 represents a general block diagram of all the approaches discussed. The initial construction of the tensor and the unfolding are common for the three

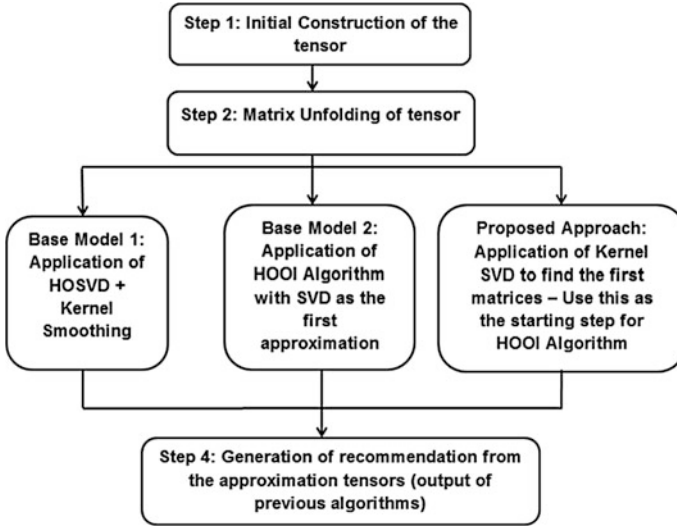


Fig. 1 General block diagram with different approaches

approaches. To these unfolding, algorithms are applied. The output of the algorithm is an approximation tensor from which recommendations can be drawn.

2.3.1 HOSVD with Kernel Smoothing Base Model 1

The Higher Order Singular Value Decomposition is a generalization of SVD computations to multidimensional data. Matrix unfolding is done to each of the modes of three-way tensor $\mathcal{A} \in R^{I_1 \times I_2 \times I_3}$, and SVD is applied to each of these modes to get the factor matrices U^1, U^2, U^3 . HOSVD is performed by n-mode product of tensor A with the factor matrices.

$$A \approx \hat{A} = S \times_1 U^{(1)} \times_2 U^{(2)} \times_2 U^{(3)}. \tag{1}$$

where U^1, U^2, U^3 are the orthogonal vectors spanning the column space of the respective matrix unfoldings. S is called the core tensor and has the property of all orthogonality. The tensors modeled are sparse tensors, and this could affect the approximations performed. Hence, SVD is applied on Kernel-defined feature space. In order to avoid the explicit mapping of the data points of each tensor unfolding into higher dimensional space and thereafter applying Kernel functions, we follow the known process called the ‘Kernel trick’ and find the factor matrices of each unfolding using Eigen value decomposition.

2.3.2 Higher Order Orthogonal Iteration—Base Model 2

Higher Order Orthogonal Iteration is an iterative algorithm to compute better low-rank approximation to tensors than HOSVD. It successively solves the restricted optimization problem:

$$\min_{U^p} \| \mathcal{A} - \mathcal{B} \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_N U^{(N)} \|_F^2. \quad (2)$$

Optimization is done over the matrix U^p with the latest available values of other U^i .

3 Proposed Approach

3.1 Outline

Modeling the data using HOSVD suffers from a sub-optimal solution in terms of the Frobenius norm due to truncation of matrices to find the low-rank approximations. Frobenius norm is defined for a three-way tensor as $\|\mathcal{A}\|_F^2 = \sum |A_{ijk}|^2$. This can be improved by a Higher Order Orthogonal Iterative procedure by repeatedly optimizing the approximations. In the traditional HOOI procedure, the first step is to find the initial factor matrices by applying simple SVD over the different modes of the tensor followed by iteratively updating factor matrices. A sparse data can largely affect the accuracy of the iterative procedure. In the proposed approach, we combine the two models. Kernel functions are applied to find the initial factor matrices, and these are then updated iteratively using the usual known procedure. Combining the two models takes care of both the sparseness problem and the Frobenius norm. Figure 2 briefs the steps followed in the proposed approach. Step 3 is the modification step done to traditional HOOI Algorithm to alleviate the sparseness and improve accuracy.

3.2 Example

Figure 3a gives the information on tasks performed successfully by each user. Figure 3b details out information on the skillset needed to perform each task. The algorithm initially constructs the three-mode tensor from the available dataset considering the three entities: user, skill, and task. A ‘1’ in a cell A_{ijk} represents that user i with skill j can do task k . Figure 3 represents the tensor constructed for the example scenario represented by the table. Matrix unfolding is done as shown in Fig. 4.

We then move on to performing Higher Order Orthogonal Iteration Algorithm. The first step is to find the initial factor matrices by applying kernel functions that can nonlinearly map the data to higher dimension and thus lead to a better

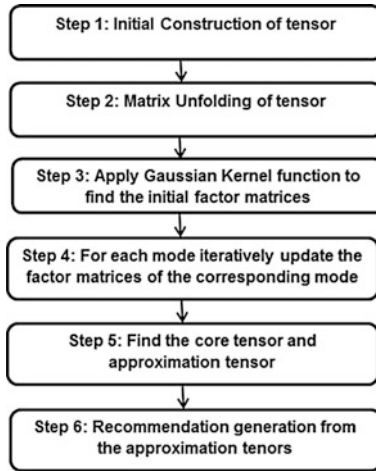


Fig. 2 Block diagram of proposed approach

(a)

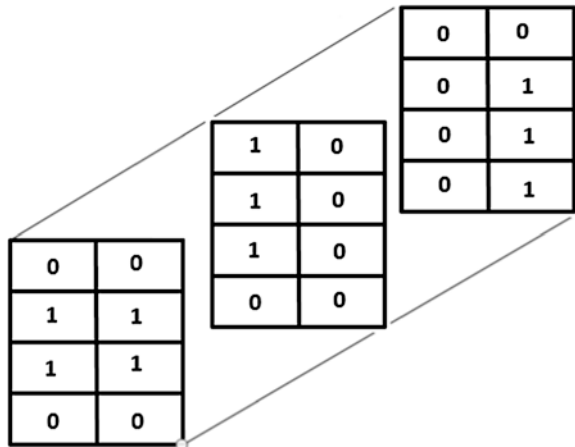
Users	Skills
U1	S1
U2	S1,S2
U3	S1,S2
U4	S2

(b)

Skill	Task
S1	T2
S2	T3
S1,S2	T1

Fig. 3 Tensor construction of the running example. a User-skill table. b Skill-task table

Fig. 4 Matrix unfolding of tensor



structuring of data. Thus, we alleviate the sparseness problem in the first step. With this as the input, we proceed to the iterative procedure to find the factor matrix of each mode using the latest available values of the other factor matrices while preserving the orthogonality as well. Gaussian Kernel functions were used taking the standard deviation of each matrix unfolding as the Gaussian parameter. A brief algorithm has been given in the Sect. 3.3.

The generation of approximation tensor marks the end of the training phase. The approximation tensor \tilde{A} measures the interaction among the three features. In the testing phase, the final approximation tensor can be interpreted as a quadruplet (u, s, t, p) where p is the likeliness measure of user u with skill s performing task t . Thus, from the existing data, for a random user, skill and task checking the value of p in the approximation tensor will give the likeliness of the successful completion of the task by user.

3.3 Proposed Approach: Algorithm

```

Initialize a 3-mode tensor t with value 0.
for each activity a, in the dataset
    (u,s,t) = Indices in tensor corresponding to a.
    t[u][s][t] = 1
for n = 1 to N do
    Compute matrix unfolding  $A_{(n)}$  for n = 1,2,3
for each matrix unfolding of tensor A
    Apply Gaussian Kernel Function
         $B_i = \exp(- \|x - y\|^2 / 2\sigma^2)$ 
         $U^i = EigenVectorOf(B_i)$ 
for k = 0 to kMax do:
    for n = 1..3 do:
         $\tilde{U} = A \times_{-n} \{U_k^T\}$ 
        Let W of size  $I_n \times R_n$  solve:
             $\max \| \tilde{U} \times_n W^T \|$  subject to  $WW^T = 1$ 
         $U_{k+1}^{(n)} = W$ 
Let  $\{U\} = \{U_K\}$  where K is the final value of previous step.
Set  $\tilde{S} = A \times \{U^T\}$ 
Set  $\tilde{A} = \tilde{S} \times \{U\}$ 
end.

```

4 Testing and Analysis

The proposed approach has been implemented in the MATLAB, and the tensor toolbox version 2.5 [15, 17] has been used for efficient computation of algorithms. To test the proposed approach, we used three sets of data from the [16] Knowledge Discovery and Data Mining Challenge 2010. In all the three sets of dataset, only records with CFA = 1 have been considered as CFA = 0 can either indicate that the student has failed in the task or indicate a not attempted task.

4.1 Effectiveness of Kernel Smoothing Technique

To measure the effectiveness of using Kernel smoothing as the starting step of the HOOI algorithm, a small sample data was extracted from the dataset, and both the traditional HOOI algorithm with SVD as the starting step and our proposed approach were applied on it. Recommendation generations were drawn from this approximation tensor, and using different samples from the dataset, we have verified and proved that the accuracy of prediction of the proposed approach outperforms the other models. A comparison of the accuracy of the base models and the proposed approach has been detailed out in the Sect. 4.2. Although the dimensions of the factor matrices obtained as a result of applying SVD and kernel functions differ, the final structure of the approximation tensor obtained using both the approaches is the same. Also, the computational complexity of using kernel functions on one unfolded tensor is $O(n^3)$ where n is the number of training vectors and that of performing SVD is $m^2n + n^3$ for a $m \times n$ unfolded tensor. The time complexity of various approaches discussed in this paper has been analyzed in the following sections.

4.2 Results

The dataset considered was partitioned into 80 % as training data and 20 % as test data. The Root-Mean Square Error (RMSE) and accuracy were calculated for the proposed approach as well as the base models. The proposed approach proved better in terms of accuracy (Fig. 6). Also, the accuracy of the proposed approach was found to increase with increase in dataset size. The proposed approach proved optimal in terms of the RMSE (Fig. 5). Although there was an increase in training time for the proposed approach, the recommendation generation can be done in real time as is the case with other approaches.

Figures 7 and 8 give a comparison of the time taken for training and recommendation generation for both the base models and the proposed approach.

Fig. 5 RMSE

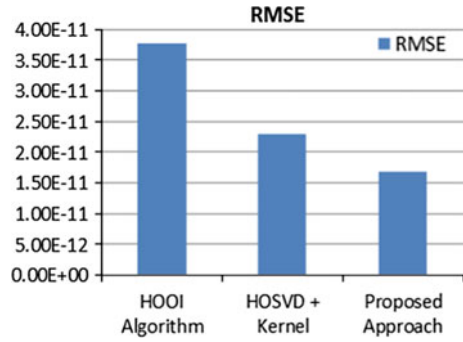


Fig. 6 Accuracy

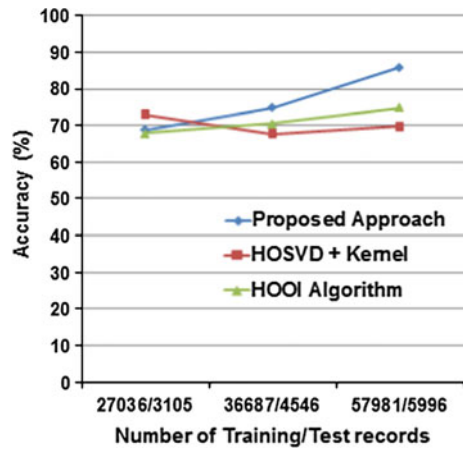


Fig. 7 Time taken for training

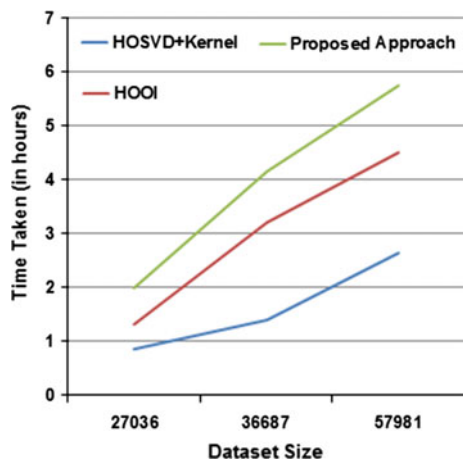
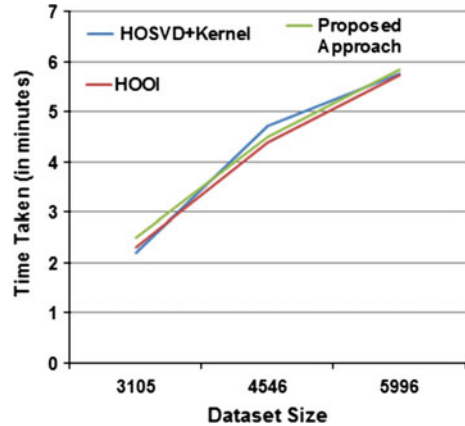


Fig. 8 Time taken for testing



5 Conclusion

The three main entities of a task recommendation system are user, skill, and task. The main idea was to consider the three entities together without splitting them. HOSVD proves sub-optimal in terms of the Frobenius norm due to the truncation of the factor matrices to find low-rank approximations. HOSVD also suffers from sparse data. Hence, it is difficult to separate the data and find a structuring of data. Thus, kernel smoothing was done to alleviate this sparseness. Higher Order Orthogonal Iterative procedure improved the accuracy of HOSVD by improvising on the low-rank approximations iteratively. But HOOI does not perform well in case of sparse tensors. Thus, we modified the starting step of the HOOI procedure in finding the first factor matrices by first projecting the data points into higher dimensional space and then iteratively finding the factor matrices. This gave an optimal solution in terms of the Frobenius norm as well as alleviated the sparseness in the original tensor constructed. Our test with over 50,000 log data records has proved that the proposed modification to HOOI algorithm is efficient than the traditional approaches.

Higher Order Orthogonal Iteration algorithm is an iterative SVD process where the resulting approximation tensors comprise negative values too. It is very difficult to interpret these negative values. Adopting a non-negative tensor factorization may lead to a better prediction.

Acknowledgment This work derives direction and inspiration from the Chancellor of Amrita University, Sri Mata Amritanandamayi Devi. We would like to thank Dr. M. Ramachandra Kaimal, Head, Department of Computer Science and Dr. Bhadrachalam Chitturi, Associate Professor, Department of Computer Science, Amrita University for their valuable feedback.

References

1. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of 14th Conference Uncertainty in Artificial Intelligence, p. 4352. Morgan Kaufmann (1998)
2. Symeonidis, P., Nanopoulos, A., Manolopoulos, Y.: A unified framework for providing recommendations in social tagging systems based on ternary semantic analysis. *IEEE Trans. Knowl. Data Eng.* **22**(2), (2010)
3. Ishteva, M., Absil, P.-A., Van Huffel, S., De Lathauwer, L.: Best Low multilinear rank approximation of higher-order tensors, based on the riemannian trust-region scheme. *SIAM J. Matrix Anal. Appl.* **32**(1), 115–135
4. Turney, P.D.: Empirical evaluation of four tensor decomposition algorithms. In: Technical Report ERB-1152, NRC-49877, November 12 2007
5. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. *Computer* **42**(8), 3037 (2009)
6. Xu, Y., Zhang, L., Liu, W.: Cubic analysis of social bookmarking for personalized recommendation. In: *Frontiers of WWW Research and Development APWeb 06*, pp. 733–738. Springer (2006)
7. Ning, X., Karypis, G.: Multi-task Learning for recommender Systems. In: *JMLR: Workshop and Conference Proceedings*, vol. 13, pp. 269–284 2nd Asian Conference on Machine Learning (ACML 2010), Tokyo, Japan, Nov. 8-10 2010
8. Tso-Sutter, K., Marinho, B., Schmidt-Thieme, L.: Tag-aware recommender systems by fusion of collaborative filtering algorithms. In: *Proceedings of ACM Symposium Applied Computing (SAC) Conference* (2008)
9. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in social bookmarking systems. In: *AI Communications*, p. 231247 (2008)
10. Kolda, T.G., Bader, B.W.: Tensor Decompositions Appl. *SIAM Rev.* **51**(3), 455500 (46 p)
11. Ishteva, M., De Lathauwer, L., Absil, P.: Dimensionality reduction for higher-order tensors: algorithms and applications. *Int. J. Pure Appl. Math.* **42**, 337–343 (2008)
12. Luo, D., Huang, H., Ding, C.: Are tensor decomposition solutions unique? In: *On the Global Convergence of HOSVD and ParaFac Algorithms*. CVPR2009
13. Sun, J.-T., Zeng, H.-J., Liu, H., Lu, Y., Chen, Z.: CubeSVD: a novel approach to personalized web search. In: *Proceedings WWW '05 Proceedings of the 14th International Conference on World Wide Web*, pp. 382–390 ACM New York, NY, USA (2005)
14. Nanopoulos, A., Krohn-Grimberghe, A.: Recommending in social tagging systems based on kernelized multiway analysis. In: *Proceedings of the 11th IFCS Biennial Conference and 33rd Annual Conference of the Gesellschaft für Klassifikation e.V.*, Dresden, March 13–18 2009
15. Bader, B.W., Kolda, T.G.: Efficient MATLAB computations with sparse and factored tensors. *SIAM J. Sci. Comput.* **30**(1), 205231
16. Stamper, J., Niculescu-Mizil, A., Ritter, S., Gordon, G.J., Koedinger, K.R.: Algebra 2005–06. In: *Challenge Dataset from KDD Cup 2010 Educational Data Mining Challenge* (2010)
17. Bader, B.W., Kolda, T.G., et al.: *MATLAB Tensor Toolbox Version 2.5*, Available online, January 2012

A One-to-One Communication Model to Facilitate Conversation Between Differently-Abled People by a Portable and Handy Machine

Rajat Sharma and Swarnima Gupta

Abstract The research study presented in this paper focuses on the process of communication between differently-abled people via technological machine model stated in the methodology section of this research paper. This challenging issue of human interaction between dissimilar impairments has sought a medium of communication between a normal person and a differently-abled person troubled by a single disability, be it blindness, be it deafness, or be it dumbness. But, apart from all such scenarios, no such solution exists in this world, full of advancements that include the word “technology” where a remedy to dual or multiple (combination of blind, deaf, dumb) defects in a person right from their births is available. So, keeping this fact alive in the woes and wits, authors have proposed a machine model by help of which the sufferers suffering with similar or dissimilar disabilities could have dialogue with one another in the easiest possible manner. The prime focus of the authors is to shed some light in the field of communication between disabled people suffering with mono, dual, or multiple defects who could meet the basic day-to-day necessities of their life by using this proposed machine model, who were using finger sensations and lip movement learning methods that were miles away from technological grounds till yet. The algorithms have been presented to convey the proper working and functioning mechanism of the machine model to ease the task of live conversation between the sufferers. The various case studies involving every possible use case diagram have been included, showing the parameters and considerations on which the recommended machine model will be best suited for working, have sought its wide discussion. The final phase of this research study will show the descriptive images highlighting the working methodology and usage of tables has also been done for providing a better knowledge of the proposed concept.

R. Sharma (✉) · S. Gupta
Computer Science and Engineering Department,
ABES Engineering College, Ghaziabad 201009, U.P, India
e-mail: march10rajat@gmail.com

S. Gupta
e-mail: swarnima.aug15@gmail.com

Keywords Differently-abled people · Technology · Machine model · Disabled · Visual impairment · Braille · Keyboard · Use case · Dual defect

1 Introduction

Science and technology have great influence in human life since their inception. Technology is considered to be the one, which is developed to reduce the efforts of people that provides countless visible advantages for human beings. But there exists a section of society comprising of differently-abled people, who experience the fury of exclusion in every aspect of their lives and this concern is, of utmost importance that could never be ignored. Creative solutions that are aimed to help differently-abled people are merely adopted on marginal grounds. So far, work done in this field and technological advancements taken place has provided medium of communication between a normal person and a differently-abled person challenged by a single disability, be it blindness, be it dumbness, or be it deafness. But the idea proposed in this paper imparts communication between persons with either of the suffering mentioned as well as the combinations. The authors have gone through a rigorous research to find out, what possible distinct cases exists or might exist (imaginary), where a person A suffering with defect X (blind, deaf, dumb, or any combination with blind (mandatory)) could easily communicate with person B suffering with defect Y (blind, deaf, dumb, or any combination with blind (mandatory)). The technology discussed below includes various algorithms that would when be applied to the proposed machine functioning, would fulfill the motive of establishing the platform for differently-abled people to have dialogue with one another. The writers have supported their research work with various different case studies along with use case diagrams showing the whole communication procedure between sufferers who are having above-mentioned defects right from their birth.

2 Literature Survey

In the following segment of literature survey done, authors have only highlighted the most relevant techniques to show communication between visually impaired persons, available so far in this world moving on the great heights of advancements. The authors of [1] have discussed the procedure of interaction between blind and deaf people where they have stated few cases like a deaf cannot speak for a blind to hear or if a blind speaks then deaf cannot hear that or if a deaf makes sign gestures (usage of sign language to communicate) would then be of no usage for blind. So,

they have presented a multimodal interface to establish a live conversation among the concerned sufferers. The researchers of [2] have suggested a dialogue model along with a general structure of communication where they have provided a unique automated assistance center for blind sufferer. On moving deeper, along the side of technological world, few researchers have used mobile as well as cloud computing and social networking [3] techniques to provide a ray of hope in the dark world of blinds where they have recommended a comprehensive framework to fulfill the communication procedure with special consideration to Arabic countries showing the usage of latest tactics like time-of-flight camera. A research study done by [4] came with a concept of finger Braille teaching system where normal person's speech is recognized perfectly and is converted to Braille code for blind sufferers to read and respond. Some proposed methodologies even proved to provide a large scope of betterment for this section of people as a concept of Bus Identification System [5] where, with usage of mechatronic system architecture, a visually impaired person could catch the bus with help of an audio device and a tactile interface by injecting the RFID and wireless sensor network technologies. The researchers of [6] have presented the usage of Body-Braille and infrared communication methodology establishing an infrared communication link where the sufferers could discover the sufferers suffering with different impairments in the nearby places. The authors of [7] have shown a procedure by which a common individual could easily converse with a disabled person where messages typed in Braille will be converted into Mandarin Phonetic symbols which will be shown on LCD [8] display for the helper to read and send the messages as a reply back on simple and will be shown on Braille [9] display for blind sufferers to read and respond.

3 The Proposed Methodology

The foremost drawback that existing technologies possess, has been discussed by the authors in their research study as, the available techniques to establish communication between differently-abled people is basically between a nondisabled person and a defected person but no such technology, that could establish a communication link between two disabled people suffering with mono, dual, or multiple defects right from their birth was possible till yet. So, by keeping this fact at utmost priority the authors have presented a machine model that will cater the service of conversation among the impaired people suffering from any kind of defect, with blind defect as a mandatory condition. The architectural machine model proposed by the authors is very unique, smaller in size, portable and is inexpensive, which provides easiness for any person to buy and use. It would be very safe and a reliable component for the user where its complexity on the grounds of software concern would be very less. Also, each proposed machine would be having a unique

identification code that can be used to trace the person having the machine, on the grounds of emotional concern to check the present condition of the person. The process of communication as per the concept discussed by authors in this research study includes the usage of two machines that would be available with the sufferers who are in need to establish the process of communication between them. After doing all the settings for the first time start-up procedure (as stated in step 2 of Algorithm 1), as soon as the two machines will come in range of 1 m, the machines will switch to active mode by using the RFID (Radio Frequency Identification) technology as both the machines will have RFID tags as well as RFID transmitters so when they will come in contact, then will sense the other machine because of the same RFID tag (all the machines will have same RFID tags so if any other RFID tag is found in the nearby area then the machine will not be activated) and thus the connecting communication link will be established. Now, the sender will send the message in the preset message format selected at the initial stage of setting phase and the message will travel via communicating media (air) and will be received by the sufferer on the other side in the preset message format. The message transference and message format conversion will occur as per the defined Algorithm 2. Once the message is sent and replied by both the sufferers, then if they want to end up with their conversation then they just have to move away from each other, i.e., more than the range of 1 m, then as per the Algorithm 2, the connecting link will breakdown and will stop the conversation furthermore the machines will come in dead mode. The algorithm that will be included for the logical working of the suggested machine making a pathway for communication between blind, deaf, and dumb is as follows:

Considerations to be made

The sufferer can be both BLIND and DUMB or could be BLIND and DEAF and DUMB (Only these two former defects are of concern for designing the algorithm).

Algorithm 1 -

BEGIN

- Step 1** : *Operate*: Function of “ON” to the power button of the machine for the first and last time (only applicable for the machine which is newly purchased from the market).
- Step 2** : There are some required settings that are to be done by the helper for the first time to give the required instructions at the time of initialization process.
- Step 3** : *Initialize*: Value of flag variable with zero as default (FLAG <- 0).
- Step 4** : *Set*: Date, time and format of the output message to be seen on the machine screen after receiving the message from sender machine.
- Step 5** : *Set*: Value of FLAG variable as one after setting the required parameters (FLAG<-1).

Step 6 : *Check:*

```

    if (FLAG==1)
    {
        Set: Default output message mode format as braille format.
    }
    else
    {
        Set : All parameters again
        Goto: Step 3.
    }

```

Step 7 : *Set:* Range for both the machines to get activated and also a connecting link is established (RANGE= 0-1metre).

END

Algorithm 2 -

BEGIN

Step 1 : *Use:* Algorithm 1 for initialising and setting the required parameters on machine start-up for the first time.

Step 2 : *Check:*

```

    if (RANGE <= 1 meter)
    {
        Activate: Two machines in range by using RFID technology.
        Send: Input is being sent from sender to the receiver.
        Receive: Message is received in Braille format.
    }
    else
        Connecting link between the machines is not established and
        no communication is performed.

```

END

Explanation of the Algorithm 1 and Algorithm 2

When the sufferer will buy a new piece of the proposed machine model from the manufacturer, then during first off operation, he/she has to “ON” the machine by switching on the power button for the first time. After operating the “ON” mode to the machine, all the required settings like date, time, and day and the most important feature—message receiving mode is to be set with the assistance of a helper (on demand). Since, the machine would come in use for the first time, so initialize the FLAG variable with zero (default value). As soon as the message format of the message that will be received and will be sent is set, i.e., Braille, then set the value of FLAG as 1, that indicates input and output message mode is set for the machine and further processing now needs to be done. Then, range for both the machines to be operated is set for 0–1 m. Now, this Algorithm 1 will be utilized to carry forward its functioning in Algorithm 2 where once the machine settings are done, now when the two machines will come in range and are “ON” mode, then they will be activated by using the RFID technology, the message from sender will

be sent in Braille format to the sufferer on the another side and in turn a replying message with same message format will be sent to the sender. Now, if the two disabled people having visual impairment defect as mandatory condition will not be in the range of 0–1 m and the “ON” is working then their machines will not start. But if they will come in the defined range then their machines will start the process to send and receive the message, i.e., communication will begin. If the power button is turned “OFF” then even if the two sufferers will be in the scope if 0–1 m then also their machines will become active and no communication process could be started.

4 Results

When both the differently-abled persons will be having the proposed machine model, then the flow of information will occur as is shown in the figure given below (Fig. 1).

The sufferer A (blind) when will come in the range of 0–1 m, will establish a connecting link with another sufferer B (blind and deaf and dumb) and will be sending the message in Braille message format by using the Braille keyboard present on the proposed machine model. The message will be sent to the machine of the other sufferer as a network link will be maintained between the machine models in contact by using the RFID technology and after receiving the message in Braille format (preset as per the defect possessed by the sufferer by birth) will send a reply in the same message format (Braille) and since the connection has been made by using RFID technology, no other interruptions will hinder the process of communication and one-to-one message transference could be done which will work perfectly possible and will be best for a visually impaired person to read and respond. Now, the authors of this research study have shown various case studies

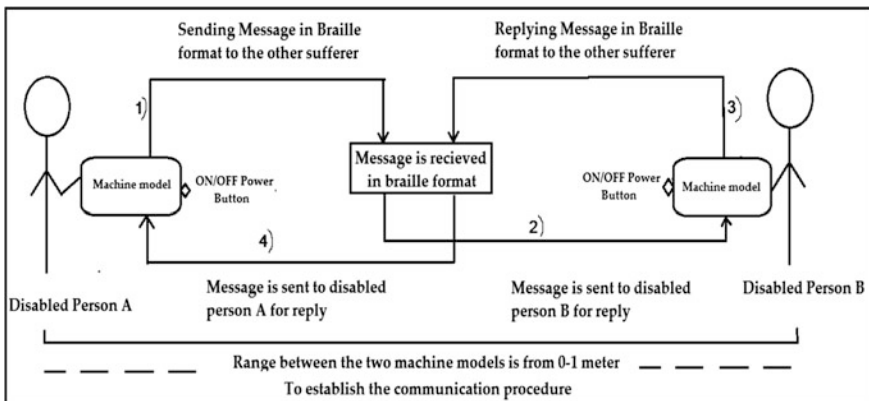


Fig. 1 The resultant snapshot showing whole working procedure of proposed concept

that are needed to be considered for showing the real usage of proposed machine model that could be summarized as—The keyboard machine that will be available to both the sufferers in contact, will be consisting of Braille keyboard where typing will occur in Braille format only, i.e., sending message format and receiving message format, both will be in Braille by default.

Assumptions to be considered before considering the case studies:

1. Blind and Dumb person can send and receive message in Braille format only.
2. Blind, Deaf, and Dumb can send and receive message in Braille format only.

Case Study: Blind, Deaf, and Dumb actor A communicating with Blind and Dumb actor B

Step 1: Activating the machines by setting up the link as:

As soon as the two actors having machines, will come in the range of 0–1 m, a connecting link will be set up between them by using RFID technology and respective machines will get activated. The communication procedure starts as

Step 2: Blind, Deaf, and Dumb actor A sends message to Blind and Dumb actor B

(a) *Sending Phase:*

The Braille messages will be sent by Blind, Deaf, and Dumb actor A to Blind and Dumb actor B. No conversion in the format of message will occur. The Braille message will be received as it is by Blind and Dumb.

(b) *Replying Phase:*

No conversion in the format of message will occur. The Braille messages will be sent by actor B (Blind and Dumb) and will be received as it is by actor A (Blind, Deaf, and Dumb) (Fig. 2).

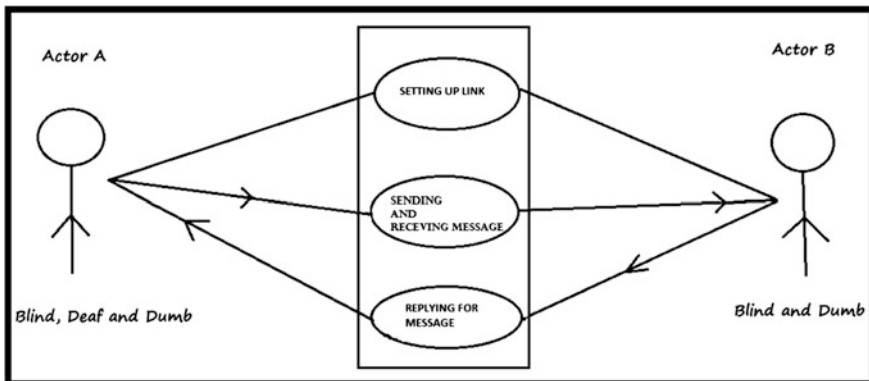


Fig. 2 Use case diagram showing the communication procedure

Table 1 Showing the other case studies possible for communication procedure

Communication case	Sending phase	Replying phase
Blind and Dumb person A to Blind and Dumb person B	Messages will be sent in the form of Braille by A	Messages in the form of Braille will be sent by B
Blind and Dumb person A to Blind actor B	Messages will be sent by A in Braille format	Messages will be sent in Braille format by B
Blind and Deaf and Dumb person A to Blind and Deaf and Dumb person B	Messages will be sent in the form of Braille by A	Braille messages will be sent by B
Blind and Deaf and Dumb person A to Blind person B	Messages will be sent in the form of Braille by A	Messages will be sent in the form of Braille by B
Blind person A to Blind person B	Messages will be sent by A in Braille format	Braille format messages will be sent by B

As discussed above few more cases where a communication link could be established between the sufferers for which Braille could prove to be an opening gate of the dark life are as follows (Table 1).

Therefore, all the discussed cases mentioned above will be the only possible scenarios where our proposed machine model will be a necessity and will help the sufferers with visual impairment and any other defect sufferer to fulfill the motive of communication among the disabled people.

5 Conclusion and Future Scope

With the above-mentioned scenarios some extracted conclusions could be discussed as, a differently-abled person with visual impairment and any other defect like deafness or dumbness or any combination would be able to communicate with another individual troubled by the similar or dissimilar disabilities. The authors have proposed a machine model with running algorithm, describing their work with various case studies of the people having mono, dual, or multiple defects to involve themselves in the procedure of dialogue in the easiest possible manner. But, the proposed concept for communication has blind impairment as mandatory condition. So, in future the suggested machine model can be upgraded to exempt this mandatory clause and can be made flexible with all kinds of defects. Furthermore, the chats of the person done could not be stored neither be recalled for future purposes, thus this machine could be given an additional touch-up to store the chats of visually impaired sufferers. Additionally, some other applications along with communication could be made possible to work concurrently on the proposed machine model.

References

1. Bourbakis, N., Exposito, A., Kobraki, D.: Multi-modal interfaces for interaction-communication between hearing and visually impaired individuals: problems and issues. In: Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on, vol. 2, pp. 522–530 (2007)
2. Cialdini, Robert, B., Borden, Richard J., Thorne, A., Walker, R.M., Freeman, S., Sloan, R.L.: Basking in reflected glory: three (football) field studies. *J. Pers. Soc. Psychol.* **34**(3), 366 (1976)
3. El-Gayyar, M., ElYamany, F.H., Gaber, T., Hassanien, E.A.: Social network framework for deaf and blind people based on cloud computing. In: Computer Science and Information Systems (FedCSIS), 2013 Federated Conference on, pp. 1313–1319 (2013)
4. Matsuda, Y., Isomura, T., Sakuma, I., Kobayashi, E., Jimbo, Y., Arafune, T.: Finger braille teaching system for people who communicate with deaf blind people. In: Mechatronics and Automation, 2007. ICMA 2007. International Conference on, pp. 3202–3207 (2007)
5. Kumar, Y., Haas, D., Nielsen, D., Mothersell, S.: RFID and GPS integrated navigation system for the visually impaired. In: Circuits and Systems (MWSCAS), 2010 53rd IEEE International Midwest Symposium on, pp. 1149–1152 (2010)
6. Ruman, S., Das, S.: Analysis of different braille devices for implementing a cost-effective and portable braille system for the visually impaired people. *Int. J. Comput. Appl.* **60**(9), (2012)
7. Su, M.-C., Chen, C.-Y., Su, S.-Y., Chou, C.-H., Hsiu, H.-F., Wang, Y.-C.: Portable communication aid for deaf blind people. *Comput. Control Eng. J.* **12**(1), 37–43 (2001)
8. Sharma, R., Jain, A., Rastogi, R.: A new face to photo security of Facebook. In: Contemporary Computing (IC3), 2013 Sixth International Conference on, pp. 415–420. IEEE (2013)
9. Sharma, R., Agarwal, A.: An innovative approach to show the hidden surface by using image inpainting technique. In: Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014, pp. 395–403. Springer International Publishing (2015)

Malicious File Hash Detection and Drive-by Download Attacks

Ibrahim Ghafir and Vaclav Prenosil

Abstract Malicious web content has become the essential tool used by cybercriminals to accomplish their attacks on the Internet. In addition, attacks that target web clients, in comparison to infrastructure components, have become prevalent. Malware drive-by downloads are a recent challenge, as their spread appears to be increasing substantially in malware distribution attacks. In this paper we present our methodology for detecting any malicious file downloaded by one of the network hosts. Our detection method is based on a blacklist of malicious file hashes. We process the network traffic, analyze all connections, and calculate MD5, SHA1, and SHA256 hash for each new file seen being transferred over a connection. Then we match the calculated hashes with the blacklist. The blacklist of malicious file hashes is automatically updated each day and the detection is in the real time.

Keywords Cyber attacks · Botnet · Malware · Malicious file hash · Intrusion detection system

1 Introduction

A drive-by download is any installing of software that occurs without the realization and permission of a user. Nowadays, drive-by downloads form a serious threat to the Internet and its users [1]. In a typical attack, the user's computer can be infected with malware by only visiting a web site that contains the malicious content. Then, the malicious code which is installed on the victim's machine can control the infected machine and perform malicious activities. Typically, sensitive

I. Ghafir (✉) · V. Prenosil

Faculty of Informatics, Masaryk University, 60200 Brno, Czech Republic
e-mail: ibrahim_ghafir@hotmail.com

V. Prenosil

e-mail: prenosil@fi.muni.cz

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_63

data is exfiltrated, passwords are stolen, and keystrokes are recorded. Also, the infected computer may become a member of a botnet [2]. Then these infected computers can be exploited for denial-of-service attacks [3] or spam campaigns [4].

In typical drive-by download attack, first, a web browser requests a web page from a remote web server. The server returns a web page as a response to the web browser request; this web page contains a malicious code that exploits a web browser's vulnerability. The malware can be delivered as part of the malicious code, or a downloader, which is a special payload, can be used to pull and then execute malware on the local workstation [5]. The whole attack occurs without the user's knowledge or permission.

In this paper we present our methodology for detecting any malicious file downloaded by one of the network hosts. Our detection method is based on a blacklist of malicious file hashes. We process the network traffic, analyze all connections, and calculate MD5, SHA1, and SHA256 hash for each new file seen being transferred over a connection. Then we match the calculated hashes with the blacklist. File hash blacklist is not generally effective at detecting new or previously unknown malicious file hash because the malicious file hash has to be known before it can be added to the blacklist. However, as part of a larger solution, performing listing of file hashes is generally worth the effort. Depending on file hash blacklist is relatively "low cost" in that using them in blocking rules or log searches does not severely impact the system performance.

The remainder of this paper is organized as follows: Section 2 presents previous related work to drive-by download attacks detection. Our methodology and implementation of our detection method are explained in Sect. 3. Section 4 shows the results and Sect. 5 concludes the paper.

2 Related Work

In [6], Hsu et al. proposed BrowserGuard, a runtime and behavior-based system. BrowserGuard protects the browser against drive-by download attacks by recording the download scenario of every file that is loaded into the host through the browser. Then based on this scenario, it can provide the user with alerts about suspicious downloaded files. Another drive-by download detection method was proposed by Zhang et al. [7]. They do not rely on the malicious content, instead of that, their algorithm depends on the URLs of the MDN's (malware distribution network) central servers. Based on those URLs, they generate a set of regular expression-based signatures.

Google conducted a study in which they investigated malware pushed onto the client machine as a result of drive-by download attacks [8]. The Google researchers found that the malware pushed by a URL does not change, while some URLs become different over time. Cujo, a system for automatic detection and prevention of drive-by download attacks, was presented by Rieck et al. [9]. It uses efficient

mechanisms of machine learning for extracting and analyzing static and dynamic code features.

Many drive-by download attacks use heap sprays [11] to achieve the attacks. Nozzle [10] detects heap spray attacks based on the monitoring that shellcode used in a heap spray attack is often prepended with a long NOP sled. Gadaleta et al. [12] inserts interrupt orders into each JavaScript string variable, and then stores it storing it in the heap and gets back the modified string to the original one before using it.

In [1], Provos et al. managed large-scale, long-duration study, in which billions of URLs were analyzed. From these URLs, more than 3 million launched drive-by download attacks. They found that drive-by downloads are triggered by visiting web sites which are not necessarily of malicious intent, but the content of these sites attracts users into the malware distribution network. Song et al. [13] proposed an inter-module communication monitoring-based mechanism to detect malicious exploitation of vulnerable components.

3 Methodology

In this section we propose our methodology for detecting any malicious file downloaded by one of the network hosts. Our detection method is based on a blacklist of malicious file hashes. As it is shown in Fig. 1, we process the network traffic, analyze all connections, and calculate MD5, SHA1, and SHA256 hash for each new file seen being transferred over a connection. Then we match the calculated hashes with the blacklist. The blacklist of malicious file hashes is automatically updated each day and the detection is in the real time. We have

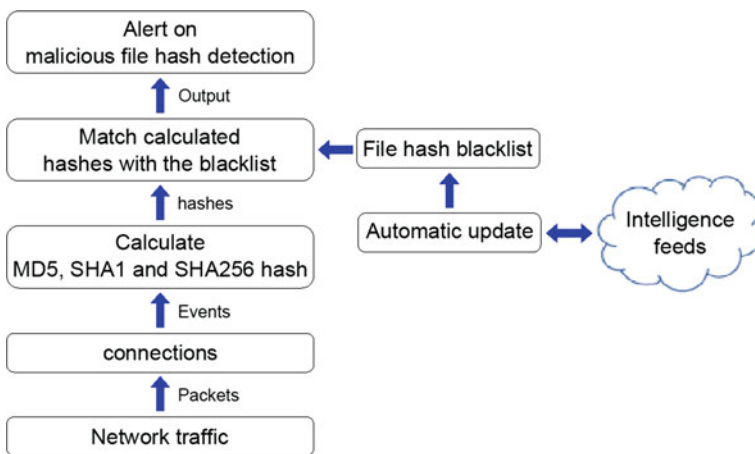


Fig. 1 Methodology of malicious file hash detection

implemented our detection method on top of Bro [14] which is a passive, open-source network traffic analyzer.

We have made use of *Bro Intelligence Framework* [15]; this framework enables you to consume data from different data sources and make it available for matching. In our detection method we configured the intelligence framework to monitor all file hashes which are seen being transferred over the network traffic. We connected this framework to *blacklist.intel* text file which contains the file hash blacklist.

Figure 2 explains the implementation of our detection method in Bro, first we start processing the network traffic. Bro is able to reduce the incoming packet stream into a series of higher level *events*, so we can get *file_new* event. This event indicates that an analysis of a new file, seen being transferred over a connection, has begun. At this time, we calculate MD5, SHA1, and SHA256 hashes for the current file; these calculations are performed by three functions in Bro, which are

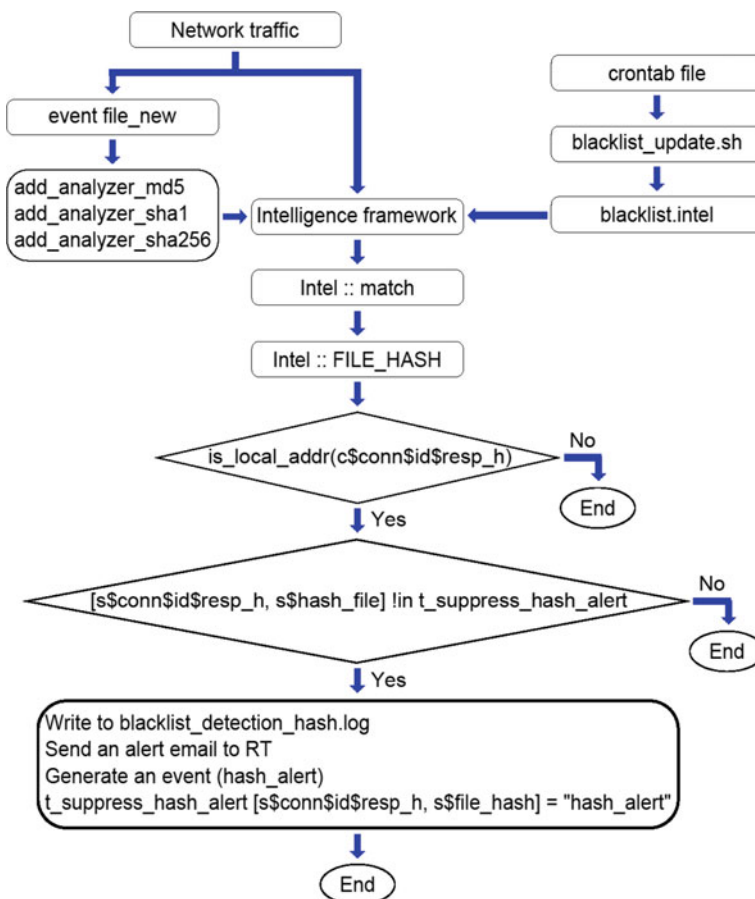


Fig. 2 Implementation of malicious file hash detection method

add_analyzer_md5, *add_analyzer_sha1*, and *add_analyzer_sha256*. All calculated hashes will be sent to the intelligence framework where its presence should be checked within the intelligence data set (*blacklist.intel* text file).

When a piece of intelligence data is detected, the intelligence framework will generate *Intel::match* event. This event is generated when any *indicator_type* of intelligence data is detected, because the intelligence data set may contain many of *indicator_types* for intelligence data (like *ADDR*, *DOMAIN*, *CERT_HASH*) and not only *FILE_HASH indicator_type*, therefore if the *indicator_type* is *FILE_HASH*, that means this connection is interesting for us.

This detection method is able to detect the malicious file in both cases, uploaded and downloaded, but we are interested only in downloaded malicious file. Thus, we check if this connection is oriented to one host from our network by checking the connection destination IP address through *is_local_addr* function; this function returns true if an address corresponds to one of the defined local networks, false if not. And here we should define the subnet of our network. After that and before we raise an alert, we check if we got an alert from the same host and for the same file hash during the last day because we do not want to send many alerts about the same set (host and hash) during 1 day, therefore we check if the current set is existent in *t_suppress_hash_alert* table, this table contains all detected sets during the last day.

We send an alert email about malicious file hash detection to RT (Request Tracker) where the network security team can perform additional forensics and response to it. We generate an event, *hash_alert*, about malicious file hash detection; this event can be used for alert correlation [16]. We should add the current detected set (host and alert) into *t_suppress_hash_alert* table where it will stay for 1 day to be sure that we will not get another alert about the same set during 1 day.

For automatic update of *blacklist.intel* text file which is used in our methodology, Fig. 3 shows how it is done.

We start from user *crontab file* which is configured to run *blacklist_update.sh* each day at 3:00 am, this shell script will connect through Internet to the data source server [17] and download updated blacklist of malicious file hashes into a new *blacklist.intel* text file. This text file is connected to the *Intelligence Framework* which consumes it as it is explained above. This automatic update is done without stopping network traffic live monitor.



Fig. 3 Automatic update of the intelligence data

4 Evaluation and Results

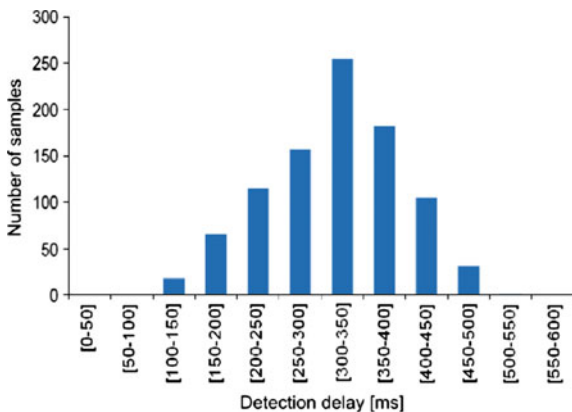
We used three scenarios to evaluate our methodology. In the first one, we ran a script, downloading malicious files, on a computer connected to the monitored network. In the second scenario, we applied our method on a pcap file which contained malicious files transferred over the connections. In the third scenario, we monitored the campus network for hosts downloading malicious files.

In the first scenario, a script downloading malicious files was installed on a computer connected to the Internet through a network monitored by our detection method. In fact, those downloaded files were not malicious but, only for this test, we added their hashes into the blacklist. In this scenario, we focused on the real-time detection capabilities of our method. The method was set up to send a report to RT (Request Tracker) as soon as a malicious file was detected. Request Tracker is often used to help incident handlers to deal with events that need an automated action or human attention.

The test consisted of the following steps. First, a script downloaded a malicious file and noted the connection time with millisecond precision. Second, the detection method detected the downloaded malicious file after the first step and automatically created an RT ticket. Third, we received the RT ticket and noted the time of arrival with millisecond precision. We compared the start-up time with the time of RT ticket arrival and noted the detection delay. Average detection delay was 340 ms with a standard deviation of 60 ms. Figure 4 shows the results.

In the second scenario, we applied our methodology on a pcap file contained traffic infected by *Nuclear EK* malware, which has an MD5 file hash *dc5c71ae-f24a5899f63c3f9c15993697* [18]. This pcap file was analyzed by the provider, so we used this fact to set the ground truth. The infection was delivered by drive-by download attack and five malicious IPs were involved.

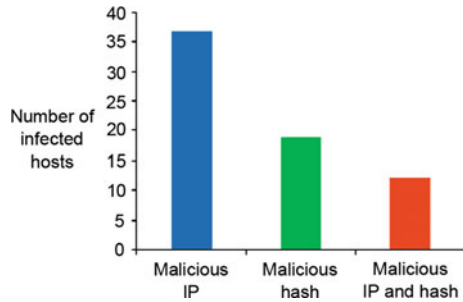
Fig. 4 Detection delay of our method



```
#fields timestamp alert_type infected_host malicious_hash
#types time string addr string
1411999086.158059 hash_alert 192.168.204.148 dc5c71aef24a5899f63c3f9c15993697
#close 2015-01-25-15-56-24
```

Fig. 5 Part of the log produced by our detection method

Fig. 6 Detected hosts by malicious file hash and IP detection methods



We set up our method that it consumed the pcap file and produced a log file. We applied our detection method on the pcap file and it was successfully able to detect the malicious file (*Nuclear EK* malware) and determine the connection over which the malware had been downloaded. Note that we did not provide the ground truth blacklist to our detection method. Figure 5 shows part of *blacklist_detection_hash.log* produced by our detection method. The log file contains more information about the malicious connection than in the figure (like source IP address, source and destination ports) but the figure shows only the interesting part of the log.

In the third scenario, we monitored the campus live traffic for detecting the hosts involved in downloading malicious files. The method was set up to create a log file of detected malicious file hashes. We set up a server hosting our detection method and passively analyzing the campus live traffic. The monitoring was performed for 1 month. We correlated the list of hosts involved in downloading malicious files with results of a malicious IP address detection method. As it is shown in Fig. 6, 19 hosts were detected involved in downloading malicious files and 37 hosts were detected involved in malicious IP address connections. Of these, 12 hosts were detected involved in both malicious IP connection and downloading malicious file, indicating that there was a malware infection.

Our detection method also sent an alert email about each malicious file hash detection to RT, where the network security team can perform additional forensics and response to it. Figure 7 shows an example of *Bro_Malicious_Hash* ticket which was sent by email to RT.

Fig. 7 Bro_Malicious_Hash ticket

```

Greetings,

the security team CSIRT-MU detected involvement of the IP address
[REDACTED] into the following incident:

Incident type: Bro_Malicious_Hash
Time of detection: 2014-12-10 14:43:18 +0100
IP address: [REDACTED]
Domain name: ---

Details of this incident can be found at this address:
https://reports.csirt.muni.cz/A4FE72DC-65A8-1C74-8627-5664BE78D651

Best regards,
CSIRT-MU, the security team of Masaryk University
http://www.muni.cz/csirt
Date: Wed, 10 Dec 2014 14:43:36 +0100

```

5 Conclusion and Future Work

Drive-by download attacks are one of the most serious security threats to computer and network systems nowadays. In this paper we have presented our methodology for detecting any malicious file downloaded by one of the network hosts. Our detection method is based on a blacklist of malicious file hashes. The blacklist is automatically updated each day and the detection is in the real time.

For future work, the output of this detection method will be correlated with the outputs of other detection methods to raise an alert on APT attack detection.

Acknowledgments This work has been supported by the project CYBER-2 funded by the Ministry of Defence of the Czech Republic under contract No. 1201 4 7110.

References

1. Mavrommatis, N.P.P., Monrose, M.A.R.F.: All your iframes point to us. In: USENIX Security Symposium, pp. 1–16 (2008)
2. Silva, S.S., Silva, R.M., Pinto, R.C., Salles, R.M.: Botnets: a survey. *Comput. Netw.* **57**(2), 378–403 (2013)
3. Moore, D., Shannon, C., Brown, D.J., Voelker, G.M., Savage, S.: Inferring internet denial-of-service activity. *ACM Trans. Comput. Syst. (TOCS)* **24**(2), 115–139 (2006)
4. Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G.M., Paxson, V., Savage, S.: Spamalytics: an empirical analysis of spam marketing conversion. In: Proceedings of the 15th ACM Conference on Computer and Communications Security, pp. 3–14 . ACM (2008)
5. Seifert, C.: Cost-effective detection of drive-by-download attacks with hybrid client honeypots. (2010)
6. Hsu, F.-H., Tso, C.-K., Yeh, Y.-C., Wang, W.-J., Chen, L.-H.: Browserguard: a behavior-based solution to drive-by-download attacks. *IEEE J. Sel. Areas Commun.* **29**(7), 1461–1468 (2011)
7. Zhang, J., Seifert, C., Stokes, J.W., Lee, W.: Arrow: generating signatures to detect drive-by downloads. In: Proceedings of the 20th International Conference on World Wide Web, pp. 187–196. ACM (2011)

8. Provos, N., McNamee, D., Mavrommatis, P., Wang, K., Modadugu, N., et al.: The ghost in the browser analysis of web-based malware. In: Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets, pp. 4–4 (2007)
9. Rieck, K., Krueger, T., Dewald, A.: Cujo: efficient detection and prevention of drive-by-download attacks. In: Proceedings of the 26th Annual Computer Security Applications Conference, pp. 31–39. ACM (2010)
10. Sotirov, A.: Heap feng shui in javascript. Black Hat Europe (2007)
11. Ratanaworabhan, P., Livshits, V.B., Zorn, B.G.: Nozzle: a defense against heap-spraying code injection attacks. In: USENIX Security Symposium, pp. 169–186 (2009)
12. Gadaleta, F., Younan, Y., Joosen, W.: Bubble: a javascript engine level countermeasure against heap-spraying attacks. In: Engineering Secure Software and Systems, pp. 1–17. Springer (2010)
13. Song, C., Zhuge, J., Han, X., Ye, Z.: Preventing drive-by download via inter-module communication monitoring. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, pp. 124–134. ACM (2010)
14. The-Bro-Project: The bro network security monitor. <https://www.bro.org/> (2015). Accessed 15 Feb 2015
15. Bro-Project: Intelligence framework. <https://www.bro.org/sphinx/frameworks/intel.html> (2015). Accessed 15 Feb 2015
16. Leau, Y.B., Tan, S.F., Manickam, S., et al.: A comparative study of alert correlations for intrusion detection. In: Proceedings-2013 International Conference on Advanced Computer Science Applications and Technologies, ACSAT 2013, pp. 85–88. IEEE (2014)
17. Computer-Incident-Response-Center-Luxembourg: Md5, sha1 and sha256 blocklist. <http://misp.circl.lu/> (2015). Accessed 15 Feb 2015
18. Network-Traffic-Analysis: Nuclear EK delivers digitally-signed cryptowall malware. <http://malware-traffic-analysis.net/2014/09/29/index.html> (2015). Accessed 15 Feb 2015

Parameter Optimization of Support Vector Machine by Improved Ant Colony Optimization

Srujana Rongali and Radhika Yalavarthi

Abstract Support vector machine (SVM) is one of the significant classification technique and it can be applied in various areas like meteorology, financial data analysis, etc. The performance of SVM is influenced by parameters like C, which is cost constant and kernel parameter. In this paper, an improved ant colony optimization (IACO) technique is proposed to optimize the parameters of the SVM. To evaluate the proposed approach, the experiment adopts two benchmark datasets. The developed approach was compared with the ACO–SVM algorithm proposed by Zhang et al. The experimental results of the simulation show that performance of the proposed method is encouraging.

Keywords Support vector machines · Ant colony optimization · Parameter optimization

1 Introduction

Support vector machines (SVMs) are administered learning approaches and were initially proposed by Vapnik [1] in the mid 1990s. SVM is based on structural risk minimization principle [2]. SVMs can be used to solve several problems such as fault detection [2, 3], fault prediction [4], forecasting [5, 6], aquaculture water quality prediction [7], predicting bankruptcy [8], predicting wave transmission [9], fault diagnosis [10], etc.

The training data is mapped to a high-dimensional feature space [11] through a mapping function, thereby constructing a separating hyperplane that maximizes the

S. Rongali (✉) · R. Yalavarthi
Computer Science & Engineering Department, GIT, GITAM University,
Visakhapatnam, India
e-mail: rsrujana8@gmail.com

R. Yalavarthi
e-mail: yradhikacse@gmail.com

margin. SVM can efficiently be used for building classification model using the training data and the model is tested using test data [11]. Proper parameter setting improves the SVM classification accuracy [12]. The parameters to be optimized are C and γ [12].

In the recent past researchers have suggested several methods to select the parameters of the classification model. Huang et al. [12] employed grid algorithm- and genetic algorithm-based approach for parameter searching on 11 real-world datasets [12] from the UCI database [12]. Huang et al. compared their proposed genetic algorithm-based approach with grid algorithm and observed that genetic algorithm-based approach significantly improved the classification accuracy. Aydin et al. [13] suggested a multi-objective artificial immune algorithm to optimize the SVM parameters and used the algorithm for diagnosing faults of induction motors. Wang et al. [14] proposed a genetic algorithm-based smooth twin parametric-margin SVM (STPMSVM) [14], which selects the parameters efficiently in addition to discriminative feature selection [14]. Zhao et al. employed genetic algorithm with feature chromosomes to optimize the feature subset and SVM parameters [15].

Reif et al. [16] presented the ideas from meta-learning for classifier parameter optimization to provide initial points to the genetic algorithm, such that acceptable accuracy is obtained. Genetic algorithm-based SVM parameter optimization was used by Yu et al. [17] where the parameters are automatically optimized. Yu et al. compared their proposals with gradient descending method and they observed that their method yields promising results. Particle swarm optimization (PSO) has been used along with SVM by Lin et al. [18] for optimal parameter search and to obtain a feature subset that has been used to train and test the model. The authors compared the proposed method with grid search method on public, UCI datasets like Australian, bioinformatics, breast cancer, iris, etc., and they observed that the classification accuracy improved noticeably.

ACO algorithm was introduced by Macro Dorigo and team in early 1990s for the solution of combinational optimization [19]. ACO deals with discrete optimization problems [20]. ACO involves basic mathematical operations. Due to the characteristics like parallelism and distribution [21], ACO processes massive data. Researches have indicated that ACO has the capability of solving global optimization problems.

Zhang et al. [19] developed an ant colony optimization (ACO) based technique for optimizing SVM parameters, i.e., ACO-SVM model. This model has been tested on five UCI benchmark datasets, namely breast cancer, diabetes, heart, thyroid, and titanic [19]. The authors compared the results with grid-based SVM algorithm, fivefold cross-validation, etc., Zhang et al. [19] stated that their model resulted in more accurate results and hence they suggest that ACO algorithm is feasible for optimizing SVM parameters. An enhanced ant colony optimization (IACO) was proposed by Li et al. [2] for selection of SVM parameters. Here meshing is applied to adjust the parameters optimized. Li et al. used rolling bearing vibration signal investigated data and compared with genetic algorithm, cross-validation and standard ACO. It has been observed that IACO-SVM performs better in optimizing the parameters of SVM [2].

2 Support Vector Machine

Support vector machine (SVM) is one of the emerging data classification techniques introduced by Vapnik [1] and is based on structural risk minimization principle [2]. SVM can be used to solve both linear and nonlinear classification problems [22–24]. A classifier model is built by considering an input vector X_i . The model is used to classify an unknown input vector.

Given a set of training data $T = \{X_i, Y_i\}_{i=1}^n$, where X_i , Y_i , and n represents the input data, class label, and number of samples respectively. The linear discriminate function (LDF) for this input vector can be written as:

$$g(X_i) = W^T \cdot \lambda(X_i) + b. \tag{1}$$

Where $W^T \cdot \lambda(X_i) + b$ is a plane, i.e., a straight line in a two-dimensional space, W is n -dimensional weight vector, and b is bias value.

A parameter γ is set in such a way that it will separate the input feature vector and the hyperplane with maximum separation such that any disturbance or noise may not result in erroneous classification. Hence, the linear function can also be written as:

$$g(X) = Y_i(W^T \cdot \lambda(X_i) + b) > \gamma. \tag{2}$$

So for designing a SVM the value of $W^T \cdot \lambda(X_i) + b$ has to be maximum and can be obtained by lowering the $\|W^2\|$ and maximizing b in above equation. This can be done by Lagrange function $L(W, b, \alpha)$ [10]

$$L(W, b, \alpha) = \frac{1}{2} \|W^2\| - \sum_{i=1}^n \alpha_i [Y_i(W^T \cdot \lambda(X_i) + b) - 1]. \tag{3}$$

In order to obtain optimal generalization, slack variable is introduced

$$\min \lambda(W, Y) = \frac{1}{2} \|W^2\| + C \sum_{i=1}^n Y_i. \tag{4}$$

Where C is the cost constant. The Lagrange function $L(W, b, \alpha)$ [10] can be written as

$$L(W, b, \alpha, Y) = \frac{1}{2} \|W^2\| + C \sum_{i=1}^n Y_i - \sum_{i=1}^n \alpha_i [Y_i(W^T \cdot \lambda(X_i) + b) - 1] \sum_{i=1}^n \Pi_i Y_i. \tag{5}$$

Taking derivative of L pertaining to W , b and Y results in

$$w = \sum_{i=1}^m \alpha_i Y_i \lambda(X_i). \tag{6}$$

$$\alpha_i [Y_i (W^T \cdot \lambda(X_i) + b) - 1] = 0. \quad (7)$$

By using Karush–Kuhn–Tucker (KKT) conditions [1], Eq. (7) is used to determine b and it can be written as

$$b = \frac{1}{N_s} \sum_{0 < \alpha_j < C} Y_j - W^T \cdot \lambda(X_j). \quad (8)$$

For unknown sample X , the classification function is

$$f(X) = \text{sign}(W^T \cdot \lambda(X) + b). \quad (9)$$

Substituting Eqs. (6) and (8) in Eq. (9), nonlinear classification function [1] is obtained as

$$f(X) = \text{sign} \left(\sum_{i=1}^N \alpha_i Y_i K(x_i, x_j) + \frac{1}{N_s} \sum_{0 < \alpha_j < C} \left(Y_j - \sum_{i=1}^N \alpha_i Y_i K(x_i, x_j) \right) \right). \quad (10)$$

Various kernel functions, namely radial kernel/ Gaussian, sigmoid kernel/multilayer perceptron, polynomial kernels, etc., can be used.

Usually SVM accuracy is influenced by certain user-defined parameters, namely C and kernel parameter [25]. Careful selection of SVM parameters affects the accuracy of the SVM model. A comprehensive examination can be performed over the parameter space but it is computationally complex. Trial-and-error method consumes more time and the obtained results may not be trustworthy [26].

Generalization error estimation and gradient descent [19] are the most elaborated techniques for parameter optimization. In 2001, Olivier et al. [27] proposed a method for minimizing the error along with gradient descent algorithm where the methods are updated iteratively.

Zhang et al. [19] presented an ACO algorithm to optimize the SVM classifier parameters. Regularization constant C and kernel function parameters have an effect on the ability of SVM. The authors [19] used ant's solution to represent C and sigma, where sigma is the RBF kernel parameter. RBF is used as it can effectively deal with high-dimensional data.

The overall process of ACO as mentioned by Zhang et al. [19] is as follows:

Primarily, initialization of parameters and variables is performed followed by calculation of grid size.

The pheromone levels of the parameter combination are same at all nodes [19], the ants are placed randomly and state transition rule is applied, to construct solution for c and σ

$$P_{ij} = \frac{\tau_{ij}}{\sum_{i=1}^N \tau_{ij}}. \tag{12}$$

P_{ij} = Problem model, τ_{ij} = Pheromone values.

$$\text{Minimize } T = \frac{1}{l} \sum_{i=1}^l \psi(-y_i^1 \cdot f(x_i^1)). \tag{13}$$

Ψ = Step function, f = Decision function of SVM, T = Pheromone trail parameters, x_i, y_i = Input and output pairs of a training set.

Then state updating rule is applied to update the pheromone level of the finest parameter set.

$$\tau_{ij}^{new} = (1 - P) \cdot \tau_{ij}^{old} + \frac{Q}{e^T}. \tag{14}$$

τ = Pheromone value, $P = (m_j - \Delta) * h_j$, Q = Pheromone intensity.

Choice of next node depends on pheromone quality, where the pheromone is maximum and the rest of pheromone is reduced.

$$v_j^{lower} \leftarrow v_j^{lower} + (m_j - \Delta) * h_j; \Delta = \text{coefficient}. \tag{15}$$

$$v_j^{upper} \leftarrow v_j^{upper} + (m_j - \Delta) * h_j. \tag{16}$$

The above process iterates, until the predefined accuracy is greater than the grid interval h_j ; and finally optimal parameter values are obtained.

Reducing the search space after each iteration increases the overhead and this needs the selection of a pair of C and sigma values which may generate local maxima for the SVM classification. As an ant proceeds from one position to another, there is no assurance that the next node is not occupied. Therefore it leads to high consumption of ants. The next section discusses an improved ACO algorithm that overcomes the disadvantages in the technique proposed by Zhang et al.

3 Proposed Improved ACO Algorithm

In the initial step, an ant establishes a best path. Initial values of the cost constant C and RBF kernel parameter gamma are fixed. Values of number of ants, iterations, records in test data, grids, ranges for C and γ are also fixed. Then C and γ are meshed according to the grid size as $P_i = (P_{max} - P_{min})/N$, where P_{max} and P_{min} specify the upper and lower bounds of the parameter, N = Grid size and P_i = Parameter interval. There is a uniform distribution of the pheromone levels over the grid and the ants are placed stochastically all over the grid.

The second step/phase involves training the SVM with the parameters represented by the ants' locations. An SVM model is generated for each parameter pair and is tested using the test data.

The third step relates to comparing the accuracy of the model associated with each ant with the best accuracy of SVM found until now. If the accuracy of a new SVM model is better than the previous one, update the latest accuracy and record the corresponding C and γ values.

The fourth step involves, updation of pheromone to the node with a parameter combination, to attract the forthcoming ants. Search the vicinity of each ant to identify a node which has a higher pheromone level than the one it is currently occupying. If such node is found, move the ant to that node. If not, move the ant to a random node in its neighborhood ensuring that the node that an ant moves to is not being occupied by another ant currently.

Finally, the process continues until the prespecified numbers of iterations are met. Classification accuracy of the best SVM model, along with its corresponding C and γ values are observed.

4 Experimental Results

In this section, the performance of the proposed improved ACO algorithm for SVM parameter optimization is assessed. The proposed improved ant optimization algorithm and the technique proposed by Zhang [19] were implemented in computing environment of LIBSVM and C++ compiler.

Diabetes and thyroid datasets from UCI benchmark and IDA benchmark are used to validate the developed model. The number of attributes for training sample and testing sample is 468 and 300, 140 and 75 in diabetes and thyroid datasets [19] respectively. The number of dimensions in diabetes and thyroid datasets is 8 and 5 respectively.

Search was performed between 2^{-10} and 2^{10} for both the parameters. The key parameters here are number of ants, iterations, and grids. The number of ants considered here are 10, 20, 30, and 40. The search space for number of iterations and grids ranged from 10 to 100. Initially, the evaporation coefficient is predefined at 5 and a timer to 0. In the experiment, the timer gets incremented for every

Table 1 Results of the proposed Improved ACO algorithm contrasting with Zhang et al. technique

Data	Zhang et al. technique				Improved ACO algorithm			
	Optimal C	Optimal γ	Accuracy (%)	Time taken (s)	Optimal C	Optimal γ	Accuracy (%)	Time taken (s)
Diabetes	176.23	151.58	77	1929.8	0.0009	136.534	79	15.24
Thyroid	37.38	2.22	97.33	666.2	65.368	2.55	97.99	13.218

pheromone level at grids and if the timer is equal to evaporation coefficient then both are again reset to starting values.

The results obtained are presented in Table 1. After experimenting with different number of ants and search space, the proposed algorithm has converged at $C = 0.0009$ and $\gamma = 136.534$ for diabetes dataset and $C = 65.368$ and $\gamma = 2.55$ for thyroid dataset. The obtained results are presented below in Table 1:

The accuracy in the above table is calculated based on the number of perfectly classified samples. Comparing the results of two methods given in Table 1, it can be noticeably observed that the accuracy of the proposed algorithm is high in the case of diabetes dataset, whereas the accuracy is almost similar in the case of thyroid dataset. These obtained results prove the effectiveness of the improved ACO algorithm.

5 Conclusions

This paper presents an improved ACO–SVM technique to obtain optimal model parameters. The proposed improved ant colony optimization technique overcomes the disadvantages of the ACO–SVM technique proposed by Zhang et al. The proposed improved model has been tested on standard UCI and IDA benchmark databases. The obtained results demonstrate that proposed method is promising.

References

1. Vapnik, V.: The nature of statistical learning theory. Springer (1995)
2. Li, X., Zheng, A., Zhang, X., Li, C., Zhang, L.: Rolling element bearing fault detection using support vector machine with improved ant colony optimization. *Measurement* **46**, 2726–2734 (2013)
3. Samanta, B., Al-Balushi, K.R., Al-Araimi, S.A.: Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection. *Eng. Appl. Artif. Intell.* **16**, 657–665 (2003)
4. Hou, S., Li, Y.: Short-term fault prediction based on support vector machines with parameter optimization by evolution strategy. *Expert Syst. Appl.* **36**, 12383–12391 (2009)
5. Wu, Q.: A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization. *Expert Syst. Appl.* **37**, 2388–2394 (2010)
6. Wu, Q., Law, R., Wu, E., Lin, J.: A hybrid-forecasting model reducing Gaussian noise based on the Gaussian support vector regression machine and chaotic particle swarm optimization. *Inf. Sci.* **238**, 96–110 (2013)
7. Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., Wei, Y.: A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Math. Comput. Model.* **58**, 458–465 (2013)
8. Wu, C., Tzeng, G.-H., Goo, Y.-J., Fang, W.-C.: A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *ScienceDirect* **32**, 397–408 (2007)

9. Patil, S.G., Mandal, S., Hegde, A.V.: Genetic algorithm based support vector machine regression in predicting wave transmission of horizontally interlaced multi-layer moored floating pipe breakwater. *Adv. Eng. Softw.* **45**, 203–212 (2012)
10. Chen, F., Tang, B., Chen, R.: A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm. *Measurement* **46**, 220–232 (2013)
11. Lin, S.-W., Lee, Z.-J., Chen, S.-C., Tseng, T.-Y.: Parameter determination of support vector machine and feature selection using simulated annealing approach. *Sciencedirect* **8**, 1505–1512 (2008)
12. Huang, C.-L., Wang, C.-J.: A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst. Appl.* **31**, 231–240 (2006)
13. Aydin, I., Karakose, M., Akin, E.: A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Appl. Soft Comput.* **11**, 120–129 (2011)
14. Wang, Z., Shao, Y.-H., Wu, T.-R.: A GA-based model selection for smooth twin parametric-margin support vector machine. *Pattern Recogn.* **46**, 2267–2277 (2013)
15. Zhao, M., Fu, C., Ji, L., Tang, K., Zhou, M.: Feature Selection and parameter optimization for support vector machines; A new approach based on genetic algorithm with feature chromosomes. *Expert Syst. Appl.* **38**, 5197–5204 (2011)
16. Reif, M., Shafait, F., Dengel, A.: Meta-learning for evolutionary parameter optimization of classifiers. *Mach. Learn.* **87**, 357–380 (2012)
17. Yu, Q., Zhang, B., Wang, J.: Parameter Optimization of E-support Vector Machine by Genetic Algorithm. In : International Conference on Natural Computation (2009)
18. Lin, S.-W., Ying, K.-C., Chen, S.-C., Lee, Z.-J.: Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst. Appl.* **35**, 1817–1824 (2008)
19. Zhang, X., Chen, X., He, Z.: An ACO-based algorithm for parameter optimization of support vector machines. *Expert Syst. Appl.* **37**, 6618–6628 (2010)
20. Alwan, H., Ruhana, K., Mahamud, K.: Optimizing Support Vector Machine Parameters Using Continuous Ant Colony Optimization. In: Computing and Convergence Technology (2012)
21. Zhang, X., Chen, X., Zhang, Z., He, Z.: A Grid-based ACO Algorithm for Parameters Optimization in Support Vector Machines. In: IEEE International Conference (2008)
22. Vapnik, V.: An Overview of Statistical Learning Theory. In: IEEE Transactions on Neural Networks, New York, vol. 10, p. 12 (1999)
23. Cristianini, N., Taylor, J.: An introduction to support Vector Machines: and other kernel-based learning methods. (2000)
24. Wang, L.: Support vector Machine Theory and Applications. Springer (2005)
25. Duan, K., Keerthi, S., Poo, A.: Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing* **51**, 41–59 (2003)
26. Imbault, F., Lebart, K.: A stochastic optimization approach for parameter tuning of support vector machines. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, pp. 1051–4651 (2004)
27. Olivier, C., Vladimir, V., Olivier, B., Sayan, M.: Choosing Multiple Parameters for Support Vector Machines. *Mach. Learn.* **46**, 131–159 (2002)

KT3F: A Key-Based Two-Tier Trust Management Filtering Scheme for Intrusion Detection in Wireless Sensor Network

Moutushi Singh, Rupayan Das, Mrinal Kanti Sarkar,
Koushik Majumder and Subir Kumar Sarkar

Abstract The air and ambience of Wireless Sensor Network (WSN) are always suitable for intruders and they always gear up to hit this network. So the safety of WSN is very much essential and extremely challenging. Lots of intrusion detection schemes have been proposed and trust management is a significant part of them. The traditional trust management schemes were developed on the basis of calculation of trust value of nodes and provide one or few levels of intrusion detection within the network. This paper has a new Key-based Two-Tier Trust Management Filtering Scheme (KT3F), which not only allows us to spot intruder several times but also strictly filters the network by using filtering and key management scheme. By serving strict filtering, this scheme boosts the security factor much higher than the previous techniques. This approach is simple and strict filtering of network helps to mark the intruder more accurately and provides high level of security.

Keywords WSN · Security · Intrusion · Trust · Motes

M. Singh (✉)

Department of IT, Institute of Engineering & Management, Kolkata, India
e-mail: moutushisingh01@gmail.com

R. Das · M.K. Sarkar

Department of CSE, University of Engineering & Management, Jaipur,
Rajasthan, India
e-mail: rupayan.das@iemcal.com

M.K. Sarkar

e-mail: mrinalkanti.sarkar@iemcal.com

K. Majumder

Department of CSE, West Bengal University of Technology, Kolkata, India
e-mail: koushikzone@yahoo.com

S.K. Sarkar

Department of ETCE, Jadavpur University, Kolkata, India
e-mail: su_sircar@yahoo.co.in

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_65

1 Introduction

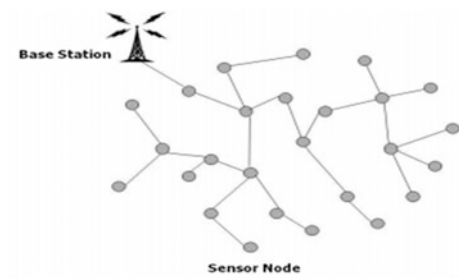
WSNs, a self-configured or self-organized network are now considered to be the ultimate frontier in modern communications where all the nodes attached with smart sensors known as motes communicate with each other using wireless transceiver without any fixed infrastructure [1–5]. Here IP flow takes place between mobile nodes through radio wave. Nowadays, *WSN* is emerging as the most promising and important platform for real-time applications. But due to the open nature of the communication medium these networks are prone to several attacks. The attackers can eavesdrop on wireless radio transmission, inject bits in the communication, drops the packets, jams the traffic of the network, replays the data packets, and many more [6]. So, *WSN* is insecure and less immune to the intruder. Some well-known emerging security attacks like wormhole black hole, jamming, hello flood are very harmful for *WSN* [6–11]. So security threats must be detected to provide complete protection of the network. Key management scheme is also an important concept to secure the system by assigning unique key to the user through a group leader. It also helps to authenticate the system [12]. Besides with key management, managing trust is also important in order to detect the intruder accurately and to make the system more secure. In our daily life the level of confidence of a person or a thing can be measured by the term trust [13]. The same concept is generally used in case of intruder detection in a network (Fig. 1).

This paper has been arranged as follows: Literature survey is done in Sect. 2. Section 3 describes the problem statement and structural blueprint of the proposed scheme. Test cases are presented in Sect. 4. Section 5 compares the scheme with existing schemes. Finally, Sect. 6 has the conclusion.

2 Literature Survey

The authors of [14] were made use of a trust model for the credentials of wormhole inside the wireless sensor network. In this trust-based system, all source nodes utilize their own trust information to estimate the most dependable pathway to a particular end by evading intermediary malicious nodes. It is a combination of both

Fig. 1 Wireless sensor network



time and trust model which identifies the bad nodes for false detection. These two models work jointly. F. Bao proposed [15] a hierarchical trust management protocol for a large quantity of sensor nodes to contract with selfish and malicious nodes and intrusion tolerance. This protocol can rigorously learn from the previous experience and can be accustomed with changeable environmental conditions. A Stochastic Petri net probability model is used to characterize the assorted WSN to discover the ground truth character. In [16] the author offers a trust management scheme to protect nodes in a cluster-based sensor network (CBSN). In order to observe the activities of every cluster node, an observer node is used which uses a watchdog mechanism to assess the credibility of nodes. The authors of [17] used efficient trust awareness in network aggregation approach for flexible WSN. Trust assessment method is applied to identify the bad nodes and categorize the trust worthiness of each sensor nodes. The concept of “all nodes in a network cooperating with each other” is not always true for the compromised node. The limitation of this approach is that it does not provide full security as it undergoes with the security issues of partial aggregation result forging. In Trust Reputation System Approach [18] the authors presented an algorithm where the trust level of nodes have been calculated on the basis of Gaussian probability function in association with Byzantine decision-making. In this scheme the trust values have been categorized as communication trust and data trust. The reputation of the neighbor node depends on the trust value (both communication trust and data trust).

3 Problem Statement

In our proposed scheme we have introduced a group leader in the network to assign unique keys to the participating motes (Fig. 2).

A. Key Management:

In this scheme the new motes must register themselves to the group leader in order to participate in the data communication. Group leader also known as Cluster Head (CH) checks the authentication of the motes and assigns a unique key. CH generates the key by random functions. After registration the motes can participate in the network for data communication. This phase is known as Key Management phase or Tier-1 phase. The diagram of key issuing phase is given below (Fig. 3).

A.1. Assigning of keys for new mote:

CH assigns a unique key to the new user randomly based on the authentication. For this it follows random function to generate unique key for each and every user. To get the unique key from the CH, motes must register themselves.

A.2. Assigning of keys for mature or old mote:

In this scheme old motes must also renew their key which is previously assigned by the CH. The algorithm for renewing a key is:

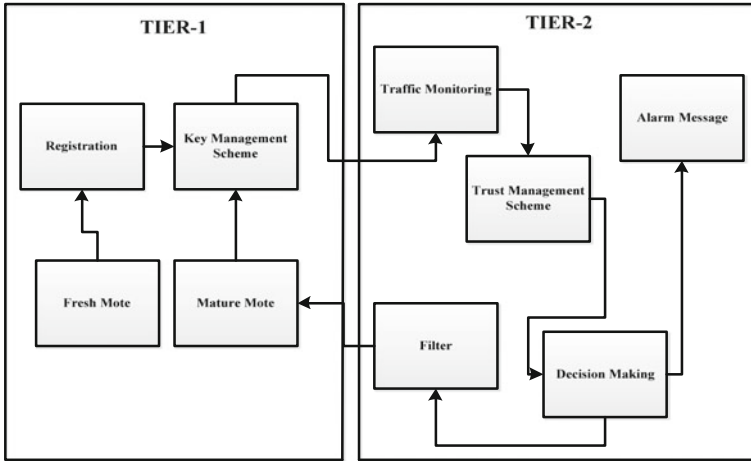


Fig. 2 Proposed two-tier architecture

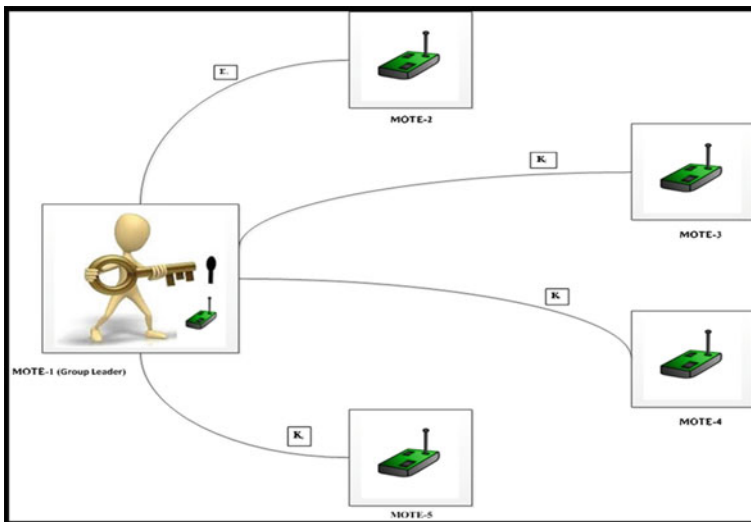


Fig. 3 Issuing key for registered mote

*If Trust value of each mote \geq Threshold level
then renew key
Else motes can not take part in the communication further.*

It is mandatory for the mature motes, having satisfactory trust value, to register again in order to renew their key. The key renewal scheme keeps the record of the trust level and helps to increase the trust rank of the motes. So, repeated renewal of keys makes the registered motes to be more trustworthy to the network.

A.3. Filtering of notes:

The key management scheme acts as a filter. It filters the good notes based on their trust value and again deploys them within the network with the updated trust value for further communication. Then follows all the steps in two-tier trust management filtering scheme. Here good notes, which are not interested for further communication can remain silent and gradually loses their trust value and deleted from the network. Thus filtering scheme makes the network much secure than the other existing trust-based scheme. The block diagram of filtering model is shown below in Fig. 4.

B. Trust Management:

After the key management phase, CH starts monitoring the network traffic. This is performed in tier-2. Traffic monitoring is based on the following criteria (Fig. 5):

1. If the notes communicate with their neighbor notes based on “request-reply” or “receive-acknowledgement” scheme then we can characterize those notes as good notes. Awful notes generally do not obey this scheme of “request-reply” or “receive acknowledgement.” The dataflow diagram for a particular network in the presence of both good mote and bad notes is shown below.
2. In the above diagram good notes follow two different schemes. First, they request to its immediate neighbor for data. Second, after receiving data from its immediate neighbor, the receiver mote must acknowledge it by sending greeting message (*ACK signal*). On the other hand bad notes do not maintain the above two criteria and gradually drop their trust level in the network.

B.1. Trust Value:

In our scheme, the trust value is defined as an integer and in every occurrence it is either incremented or decremented based on the circumstances. The life of the notes is decided on the basis of threshold trust value of the network.

B.2. Trust Value Calculation:

The calculation of trust value is based on traffic monitoring. Here the significant part is that the trust value is dynamic in nature and for each and every valid data communication it is incremented. The mathematical expression for trust value is derived below.

B.2.1: Notation and Preliminaries:

- T_{FSA} : Time taken for first successful attempt
- $T_{REQUEST}$: Time taken for sending the ‘request’ signal
- T_{REPLY} : Time taken for receiving the ‘reply’ signal



Fig. 4 Filtering model

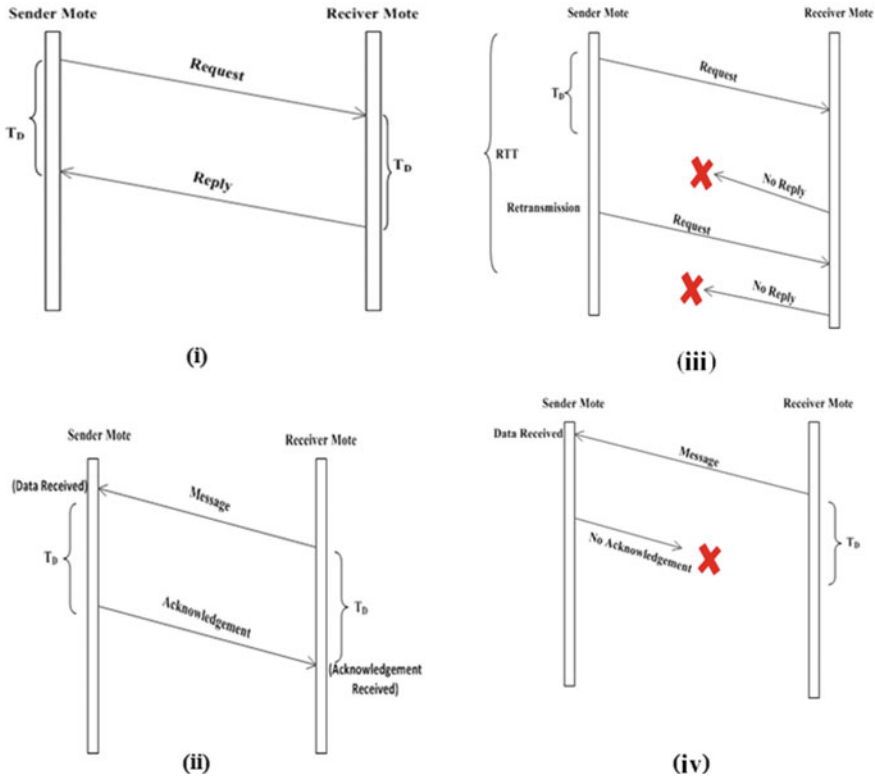


Fig. 5 (i) "Request-reply" in 'good' mote (ii) "receive-acknowledgement" in 'good' mote. (iii) "Request-reply" in 'bad' mote (iv) "receive-acknowledgement" in 'bad' mote

- T_D : Time gap between 'request' signal and 'reply' signal
- RTT : Retransmission Time
- TV_{RM} : Trust value (Changeable) for Registered Mote
- F : Finite value
- T_{RA} : Time taken for receiving the data and sending the greeting message (ACK)
- T_{RCV} : Time taken for receiving the data
- T_{ACK} : Time taken for sending the acknowledgement
- T_{STATUS} : Status of the trust value after first iteration
- $T_{THRESHOLD}$: Threshold value of trust
- K_{PREV} : Initial key value assigned by the group leader
- K_{UPDT} : Updated key value after the renewal of key
- $\Omega()$: Function which generates any integer value randomly when it calls.

B.2.2: Mathematical Expression of trust value calculation:

B.2.2.1: Definition-1: For the valid communication ("receive-reply"):

$$T_{FSA} = T_{REQUEST} + T_{REPLY} \text{ Where } RTT = 0 \quad (1)$$

$$RTT = T_D + T_{REQUEST} + T_{REPLY} \text{ (For 2nd attempt)} \quad (2)$$

Then the trust value is:

$$TV_{RM} = TV_{RM} + 1 \text{ Iff } (T_{FSA} \cup RTT) \in F \quad (3)$$

B.2.2.2: Definition-2: For the valid communication (“receive-acknowledgement”):

$$T_{RA} = T_{RCV} + T_{ACK} \quad (4)$$

$$TV_{RM} = TV_{RM} + 1 \text{ Iff } T_{RA} \in F \quad (5)$$

B.2.2.3. Definition-3: After first iteration, based on the trust value, the key of mote is updated. And for every key update event the trust value is also increased by 1.

$$\text{If } T_{STATUS} \geq T_{THRESHOLD}$$

$$\text{Then } T_{UPDT} = T_{PREV} + \Omega() \quad (6)$$

$$TV_{RM} = TV_{RM} + 1 \text{ Iff } T_{UPDT} \in F \quad (7)$$

B.2.3: Mathematical Example of trust value calculation:

In presence of “Bad mote”:

Suppose $T_{REQUEST} = 5$ ms, $T_{REPLY} = \infty$ (for no reply),

Initial Trust Value = 1

$T_{FSA} = 5 + \infty = \infty$. So RTT value is: $RTT = 10 + 5 + \text{Reply}$

Now suppose again “no reply” then in that case

$$RTT = 10 + 5 + \infty = \infty$$

So $T_{FSA} \cup RTT = \infty$ and $TV_{RM} = \text{No increment (i.e. } TV_{RM} = 1)$

In presence of “Good mote”:

Suppose $T_{REQUEST} = 5$ ms, $T_{REPLY} = 2$ ms. Initial

Trust Value = 1. $T_{FSA} = 5 + 2 = 7$ ms

So RTT value is: $RTT = 0$ (Communication is valid and successful)

So $T_{FSA} \cup RTT = \text{Finite value}$ and $TV_{RM} = TV_{RM} + 1 = 1 + 1 = 2$. So the trust value of Good mote is now 2.

Trust value after renew of keys (Reward Phase):

Suppose $T_{STATUS} = 6$ and $T_{THRESHOLD} = 5$ so $T_{STATUS} > T_{THRESHOLD}$. So the key will be updated and after that the trust value will be updated before deploying in the network further this phase is known as reward phase for the good mote (Based on the first impression or first iteration). So, $TV_{RM} = 6 + 1 = 7$.

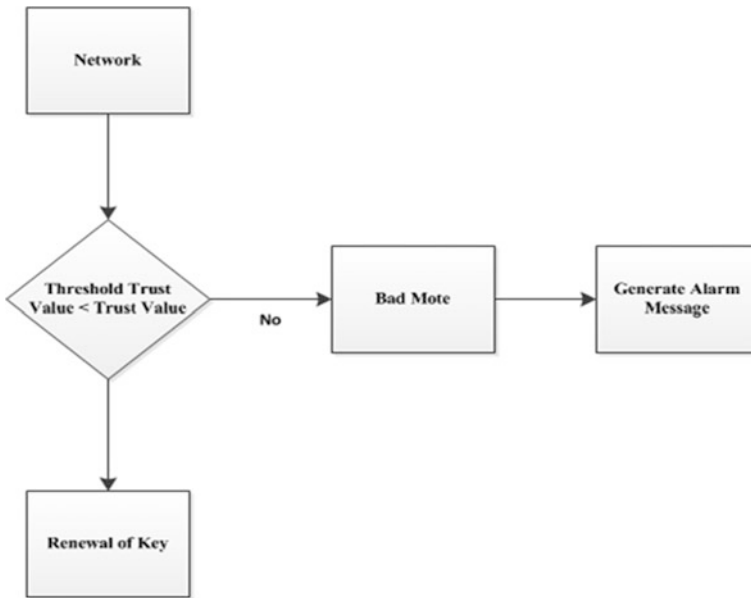


Fig. 6 Flowchart of decision module

C. Decision-Making Block:

CH monitors the whole system by rigorously checking the current trust value of each and every mote. Then *CH* takes the following decisions (Fig. 6):

1. If the trust value of each mote is satisfactory then they renew their key. Motes with acceptable trust value are extracted from the network and this technique is known as filtering.

On the other hand for at least one bad mote (Mote having trust value less than the threshold level) the *CH* generates an alarm message.

4 Security Analysis with Test Case

Test case-1: Let the number of participating motes are: 10 (Table 1).

TC-1.1: After First iteration:

From the above data set we can filter the motes having trust value less than threshold level (Table 2).

TC-1.2: Renewal Status (After filtering):

In this phase the motes having satisfactory trust value are filtered and become mature nodes. Their trust values are also incremented further by key modification (Renewal of keys). The status tables for renewal of key are given below (Tables 3 and 4).

Table 1 Table of keys and trust value

	MR ₁	MR ₂	MR ₃	MR ₄	MR ₅	MR ₆	MR ₇	MR ₈	MR ₉	MR ₁₀
Trust value	1	5	7	10	3	2	6	8	9	11
Keys	K ₁	K ₂	K ₃	K ₄	K ₅	K ₆	K ₇	K ₈	K ₉	K ₁₀

Threshold trust value = 5

We assume random trust values for every mote participating in the network. The table of keys and trust values are given below

Table 2 Decision table

Measurement parameters	Value
Max trust value	11
Min trust value	1
Threshold trust value	5
Good motes	MR ₂ , MR ₃ , MR ₄ , MR ₇ , MR ₈ , MR ₉ , MR ₁₀
Bad motes in hit list	MR ₁ , MR ₅ , MR ₆

Table 3 Status table before renewal of key

Mature motes/good motes	Old keys	Trust value
MR ₂	K ₂	5
MR ₃	K ₃	7
MR ₄	K ₄	10
MR ₇	K ₇	6
MR ₈	K ₈	8
MR ₉	K ₉	9
MR ₁₀	K ₁₀	11

Table 4 Status table after the renewal of key

Mature motes/good motes	Old keys	Trust value	New keys	New trust value
MR ₂	K ₂	5	K ₂₁	6
MR ₃	K ₃	7	K ₃₁	8
MR ₄	K ₄	10	K ₄₁	11
MR ₇	K ₇	6	K ₇₁	7
MR ₈	K ₈	8	K ₈₁	9
MR ₉	K ₉	9	K ₉₁	10
MR ₁₀	K ₁₀	11	K ₁₀₁	12

After the renewal of keys, the filtered motes are again deployed in the network with updated trust value. This renewal of key is very important as the CHs can trace the good motes here. After every filtering, the motes update their trust value by

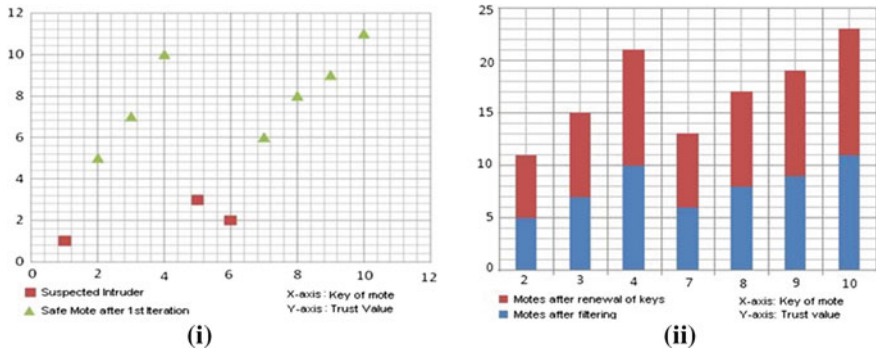


Fig. 7 (i) Trust level after first iteration. (ii) Trust level after filtering and key renewal

incrementing it. So in our scheme, the increment of trust value is not only dependent on the valid data communication between neighbor motes but also on the renewal of keys. The simulation results of above test case are shown below (Fig. 7).

5 Comparison with Existing Trust Management Schemes

Various intrusion detection techniques based on trust management are already proposed. In this article a key based trust management scheme is proposed which deals with the detection process as well as the security protection of the network. The comparisons are shown in the table below (Table 5).

Table 5 Comparison Table

Parameters	Scheme [14]	KT3F	Scheme [16]	Scheme [17]	Scheme [18]	Scheme [15]
Trust value	Not clearly defined	Threshold value is n (dynamic) range: 1 to n	0-1	-1 to 1	-n to n	0-1
Key concept	No	Yes	No	No	No	No
Authentication	No	Yes	No	No	No	No
Renewal of keys	Not done	Done	Not done	Not done	Not done	Not done
Reward scheme	Not present	Present	Not present	Not present	Not present	Not present
Trust metric	Past information	Past experience plus renewal of key (keys of mote)	Watch dog mechanism	Inconsistency checking	Past reported trust value	Quality of trust as well as social trust

6 Conclusion

The main endeavor of this paper is to present a key-based two-tier model for security challenges in *WSN*. We have introduced a new scheme to prevent an unauthorized user to access the network. The theoretical approach and results are serving as to proof the scheme as fault tolerant. In future this approach will be utilized to maintain minimum computational complexity.

References

1. Rajabhushanam, C., Kathirvel, A.: Survey of wireless MANET application in battlefield operations. *Int. J. Adv. Comput. Sci. Appl. (IJACSA)* **2**(1), (2011)
2. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. *Comput. Netw.* **38**, 393–422 (2002)
3. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. *Comput. Netw.* **52**, 2292–2330 (2008)
4. Dwivedi, A.K., Tiwari, M.K., Vyas, O.P.: Operating systems for tiny networked sensors: a survey. *Int. J. Recent Trends Eng.* **1**(2), (2009)
5. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* 102–114 (2002)
6. Singh, M., Das, R.: Survey of different techniques to detect wormhole attack in wireless sensor network. *Publ. Int. J. Sci. Eng. Res. (IJSER)* **3**(10), (2012). ISSN:2229-5518
7. Wu, B., Chen, J., Wu, J., Cardei, M.: A survey on attacks and countermeasures in mobile ad hoc networks. In: Xiao, Y., Shen, X., Du, D.-Z. (eds.) *Wireless/Mobile Network Security*, pp. – –c. Springer (2006)
8. Al-Shurman, M., Yoo, S.-M., Park, S.: Black hole attack in mobile ad-hoc networks. In: *ACM Southeast Regional Conference* (2004)
9. Aristides, M., et al.: A survey on jamming attacks and countermeasures in WSNs. *IEEE Commun. Surv. Tutorials* **II**(4), (2009). Fourth Quarter
10. Karlof, C., Wagner, D.: Secure routing in wireless sensor networks: attacks and counter measures. *Ad Hoc Netw.* **1**(2–3), 293–315 (2003). *Sens. Netw. Protoc. Appl.*
11. Godwin Ponsam, J., Srinivasan, R., A survey on MANET security challenges, attacks and its countermeasures. *Int. J. Emerg. Trends Technol. Comput. Sci. (IJETTCS)* **3**(1), (2014). ISSN:2278-6856
12. Yeh, C.-H.: A secure shared group model of cloud storage. In: *2013 27th International Conference on Advanced Information Networking and Applications Workshops*
13. Shaikh, R.A., Jameel, H., d’Auriol, B.J., Member, IEEE Computer Society, Lee, H., Member, IEEE, Lee, S., Member, IEEE, Song, Y.-J.: Group-based trust management scheme for clustered wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **20**(11), (2009)
14. Özdemir, S., Meghdadi, M., Güler, Y.: A time and trust based wormhole detection algorithm for wireless sensor networks (manuscript in Turkish). In: *3rd Information Security and Cryptology Conference (ISC’08)*, pp. 139–4, (2008)
15. Bao, F., Chen, I.-R., Chang, M.J., Cho, J.-H.: Hierarchical trust management for wireless sensor networks and its applications to trust-based routing and intrusion detection. *IEEE Trans. Netw. Serv. Manage.* (2012)

16. Zhou, Y., Huang, T., Wang, W.: A trust establishment scheme for cluster based sensor networks. *Int. Conf. Wireless Commun. Netw. Mob. Comput.* (2009)
17. Deng, H., Jin, G., Sun, K., Xu, R., Lyell, M., Luk, J.A.: Trust-Aware In-network Aggregation for Wireless Sensor Networks. *Intelligent Automation Inc.* (2009)
18. Stelte, B., Matheus, A.: Secure trust reputation with multi-criteria decision making for wireless sensor networks data aggregation. *IEEE Trans. Netw. Serv. Manage.* (2011)

Cascading De-noising Algorithm for Improving GPS Positional Accuracy

Ch. Mahesh, R. Pavan Kumar Reddy, K. Ravindra
and V. Kamakshi Prasad

Abstract The temporal and spectral variations of GPS errors severely degrade the precise positioning. These errors may change the positional information from centimeters to several meters and may result catastrophic errors in navigational information of civil aviation. Applying of de-noising methods by targeting both temporal and spectral features of GPS errors is a new strategy. With conventional code carrier smoothing algorithms, the maximum amount of temporal variations is eliminated, and by novel wavelet multiresolution analysis, the other systematic and random errors of GPS are eliminated. In present research, a combination of code carrier smoothing algorithm and wavelet multiresolution analysis is used to improve the GPS receiver position smoothing. The proposed method is well suited for critical application like GAGAN of Indian SBAS augmentation for improving positional accuracy. In this analysis, the raw GPS data is firstly processed with traditional code carrier smoothing to remove the temporal variations before position is estimated. Subsequently, the Wavelet Hidden Markov Tree (WHMT) is used to reduce the systematic and random biases of GPS.

Keywords Global positioning system (GPS) · Hatch filter · Multiresolution analysis · Wavelet hidden markov model (WHMT)

Ch. Mahesh (✉)
Airports Authority of India, Hyderabad, India
e-mail: maheshaai06@gmail.com

R.P.K. Reddy
VBIT, JNTU, Hyderabad, India
e-mail: rakashi2pavan@gmail.com

K. Ravindra
Mallareddy Institute of Technology and Science, Hyderabad, India
e-mail: kasa_ravi@yahoo.com

V.K. Prasad
Department of Computer Science, JNTUH, Hyderabad, India
e-mail: kamakshiprasad@yahoo.com

1 Introduction

In Global Navigation Satellite System, accurate position and its navigation information are the key for civil and military aviation applications. A simple GPS system is widely used because of its instantaneous position deriving capability at anywhere globally in the world. It uses the standard positioning service signal L1 (1575.42 MHz) band and the lowest resolution coarse/acquisition code [1].

The accuracy of GPS position is affected by several errors and biases. Mainly GPS errors are categorized based on their temporal and spectral behavior. The temporal variations like broadcast satellite clock models, orbit parameters, and the ionospheric variation can change significantly over the time and contribute to GPS errors. GPS errors are, based on spectral behavior, further classified into systematic and random errors. Systematic errors may be due to ionosphere, by the troposphere and clock offset. Random errors may result due to satellite orbit, receiver noise, multipath effect, etc. The systematic errors behave like a low-frequency noise, whereas the random errors are typically characterized as high-frequency noise [2, 3]. Some of these errors and biases can be eliminated or reduced through appropriate combination of the GPS observables. Instead of this method, one can deal these errors by filter. Until now, the GPS research community concentrated either on temporal or on spectral behavior errors but not on both errors combined. In view of this, it has been proposed as a cascading de-noising method which handles combined temporal and spectral characteristic features of GPS errors to improve the positional accuracy.

To take the advantage of the precision of the carrier signal, aviation applications frequently combine the code and carrier measurements. With code and carrier combination, the maximum amount of temporal variation of GPS errors is removed. The standard code carrier smoothing used in GPS data processing is the Hatch filter. The idea is to smooth less precision and more distortion data by less noisy data with more precision data [4].

Wavelets are applied in key areas which include signal estimation, detection, classification, and filtering. The wavelet transform decomposes the data as nested set of scales, called multiresolution approach (MRA), and distinguishes data and noise at each level. The primary properties like locality and multiresolution made wavelet transform to become an important tool in signal de-noising. The basic procedure of wavelet de-noising is to transform the noisy wavelet coefficients in the wavelet domain, threshold the wavelet coefficients, and then perform the inverse wavelet transform to obtain the denoised data [5].

In this paper, it has been considered the temporal behavior aside with strong correlation of spectral characteristics of GPS errors, to improve the GPS position accuracy. The required data were collected from International GNSS Service (IGS) stations of India. The paper organized as follows. Section 2 discusses the importance of the problem. The methodology and the proposed method described in Sects. 3 and 4. Data collection is discussed in Sect. 5. A performance analysis and comparison made with the proposed and traditional method are discussed in Sect. 6. Conclusions and future scope are discussed in Sect. 7.

2 Problem Statement

The basic measurements in a GPS receiver are code and carrier phase measurements. The code and carrier phase measurement at each epoch ' k ' can be written as [1]

$$\rho_1 = r_i^k + b_i - B^k + T_i^k + I_i^k + t_m + \varepsilon_{\rho_1} \quad (1)$$

$$\phi_1 = r_i^k + b_i - B^k + T_i^k - I_i^k + \lambda_1 N_1 + t_m + \varepsilon_{\phi_1} \quad (2)$$

where ' r ' is the true range between the i th receiver and k th satellite. The receiver clock bias ' b ,' satellite bias ' B ,' and tropospheric error T are non-dispersive, and they do not vary with signal frequency. The carrier phase contains an ambiguous integer number ' N ' along with other noise and biases. The GPS observables in Eqs. (2) and (3) are subject to systematic delays caused by the ionosphere (I^k), by the troposphere (T^k) and the clock difference (δ_{i_k}), ε_{ϕ} , t_m are the errors due to receiver noise and multipath. The systematic errors behave like a low-frequency noise where as the random errors are typically characterized as high-frequency noise. To apply existing de-noising algorithm to reduce the impact of noise and disturbances, the knowledge of these characteristics in the spectrum domain is essential [2]. Therefore, the proposed method is based on temporal as well as spectral variation of GPS errors and applied to cascade conventional Hatch filter with novel WHMT for improve the GPS positional accuracy.

3 Methodology

3.1 Code Carrier Smoothing

The code carrier smoothing can be implemented by hatch filter. Hatch filter is a recursive filter which linearly combines the raw pseudo range measurement at an each epoch with smoothed pseudo ranges of previous epoch and updated to the current epoch using carrier measurement which uses the code and carrier measurements to obtain more reliable and accurate position solution and also used to smooth pseudo range data using carrier phase observables for accurate position. The single frequency carrier smoothing filter uses the code measurement. The process can be expressed as [6–8]

$$\overline{\rho(t_i)} = \frac{1}{K} \rho(t_i) + \frac{(K-1)}{K} \left[\overline{\rho}(t_{i-1}) + (\phi(t_i) - \phi(t_{i-1})) \right] \quad (3)$$

$$\overline{\rho}(t_1) = \rho(t_1) \quad (4)$$

to a zero mean, high variance. The high variance, zero mean corresponds to the large state and low variance, zero mean for small state. Hence ,the Gaussian mixture model appears to be good fit for the distribution of the wavelet coefficient data being one of the states [9, 10].

Thus the overall *pdf* is given by

$$f(w_i) = p(S_i = m)f(w_i/S_i = m) \tag{5}$$

where the conditional probability $f(w_i/S_i = m)$ of the coefficient value ‘ w_i ’ given the state ‘ S_i ’ corresponds to the Gaussian distribution

$$f\left(w_i/S_i = m\right) = \frac{1}{\sqrt{2\pi\sigma_{i,m}^2}} \exp\left(-\frac{(w_i - \mu_{i,m})^2}{2\sigma_{i,m}^2}\right) \tag{6}$$

(ii) *Estimation of Signal*

The HMT model is completely parameterized by two component mixture of generalized Gaussian for the wavelet coefficients at each scale. The estimation of the true signal wavelet coefficients can be obtained by using of following equation [10].

$$w_i^\wedge = E[w_i/\theta] = \sum p(S_i = m/w_i, \theta) \frac{\sigma_{i,m}^2}{\sigma_{i,m}^2 + \sigma_n^2} w_i \tag{7}$$

where $p(S_i = m/w_i, \theta)$ is the probability of state ‘ m ’ given the noisy wavelet coefficient ‘ w_i ’ and the model parameters ‘ Θ ’ are computed by the EM algorithm.

The variance ‘ $\sigma_{i,m}^2$ ’ common to all coefficients in given scale and the noise variance ‘ σ_n^2 ’ are unknown which in turn estimated through the Median Absolute Deviation (MAD) estimator. The robust median estimator MAD is given [11].

$$\sigma_n^2 = \frac{\text{median}(|y_i|)}{0.6745} \tag{8}$$

(iii) *Wavelet basis selection*

The most challenging task in de-noising of GPS position data is the selection of wavelet basis function as well as the level of signal decomposition. In general, the orthogonal wavelet family, i.e., Daubechies (Db), Coeiflet, and Symelet, compactly supports and well captures the transient like behavior of the signal. The level of data decomposition depends on length of the data. Therefore, the orthogonal functions are used in this analysis to obtain better results.

4 Proposed Method

The flow chart for the proposed cascading De-noising method is shown in Fig. 2 and steps summarized as follows.

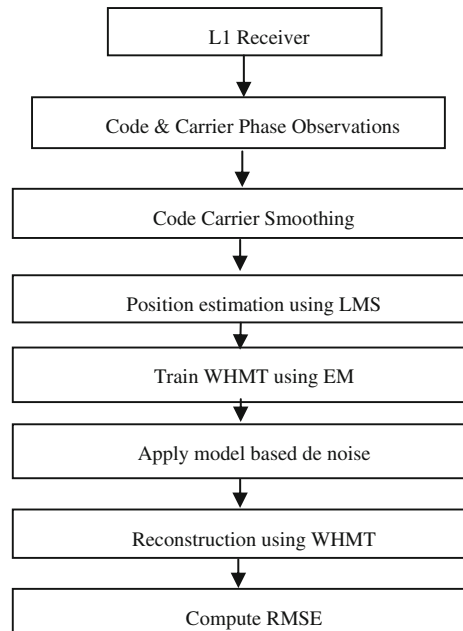
1. Extract the code and carrier frequency information from single frequency (L1 = 1575.42 MHz) GPS receiver.
2. Apply the hatch filter on extracted code and carrier smoothing purpose.
3. Estimate the average user position.
4. Apply the WHMT to the estimated position data.
5. Train the obtained wavelet coefficients using WHMT.
6. Apply the model based de-noising to remove the noisy coefficients.
7. Compute the RMSE of WHMT and Cascade WHMT.

In this proposed method, the temporal and spectral variations of GPS errors are the key to apply the cascade de-noising algorithm. The temporal variations of GPS data are well smoothed using prominent hatch filter. The spectral behavior of the GPS errors is well modeled with wavelet multiresolution analysis and applied filtering at each decomposed level.

Implementation

The estimation of true user position involves the following steps as shown in Fig. 2. Firstly, the raw pseudo range data is smoothed using code and carrier measurement with hatch filter. This is followed by computing the user position using least-square method. The unmodeled biases and left-out noise in the process of code and carrier smoothing were removed with WHMT [11] (Fig. 3).

Fig. 2 Flow chart for signal de-noising using cascade WHMT



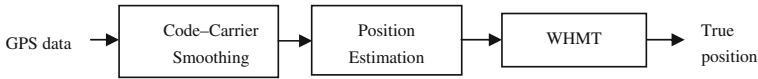


Fig. 3 Block diagram signal de-noising using cascade de-noising

5 Data Collection and Analysis

The GPS data collected from IGS stations established at IISC Bangalore (13.0210 N 77.50E) $X = 1345351.6606$, $Y = 068489.1760$, $Z = 1427815.3149$, and NGRI, Hyderabad (17.4170 N, 78.5880 E) $X = 1204618.5736$, $Y = 5967791.3377$, $Z = 1897112.0434$ on 15th January, 2010, 15th May 2010, and 15th October of 2010 which is in RINEX (Receiver Independent Exchange) format contains Navigation and Observation data files. The Navigation data contains the orbital information about time of transmission, whereas observation data contains pseudo range information of all visible satellites. The collected data contaminated with noisy on (15th Jan and 15th Oct. 2010), and solar eclipse data (15th Jan 2010) are shown in Table 1. The navigation data file contains 38 parameters; among them, 21 parameters are used to calculate the satellite position. The satellite positions have estimated at each epoch. The duration of each epoch is 30 s sampling interval and contains pseudo range information (C1), carrier phase information (L1, L2) along with other parameters. The Coarse Acquisition (C/A) code of L1 is used to compute the satellite position.

6 Results and Discussions

To evaluate the performance of the proposed method, experiments were carried out with two locations’ data acquired from IGS network (Bangalore and Hyderabad) which consists of solar eclipse data and severely affected noisy data. The data is processed as shown in Fig. 4. The least-square method is used in both cases to compute the average user position. Total 1024 (approximately 8.5 h) epochs are used for computing the coordinates of each position of each coordinates, which are shown in Figs. 5, 6, and 7.

The root-mean square error (RMSE) is calculated for each epoch, and comparison is made with the proposed and traditional WHMT. The RMSE of WHMT is enumerated in Tables 2 and 3. Table 4 shows the percentage improvement with

Table 1 Data collection

S. No	Receiver Station Name	Day, month and year
1	Bangalore (IISC)	15th Jan, 15th May and 15th Oct 2010
2	Hyderabad (HYDE)	

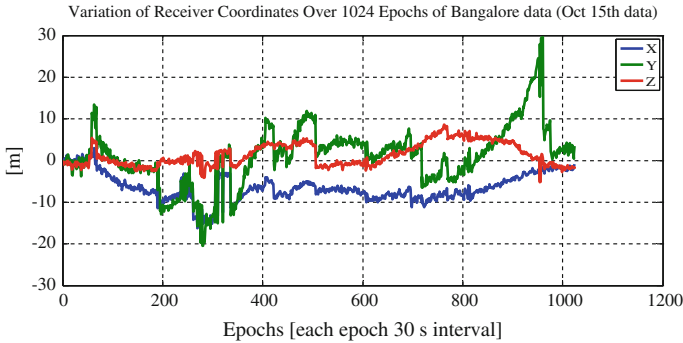


Fig. 4 Average position calculated with raw pseudo range (IISC Bangalore 15th October, 2010)

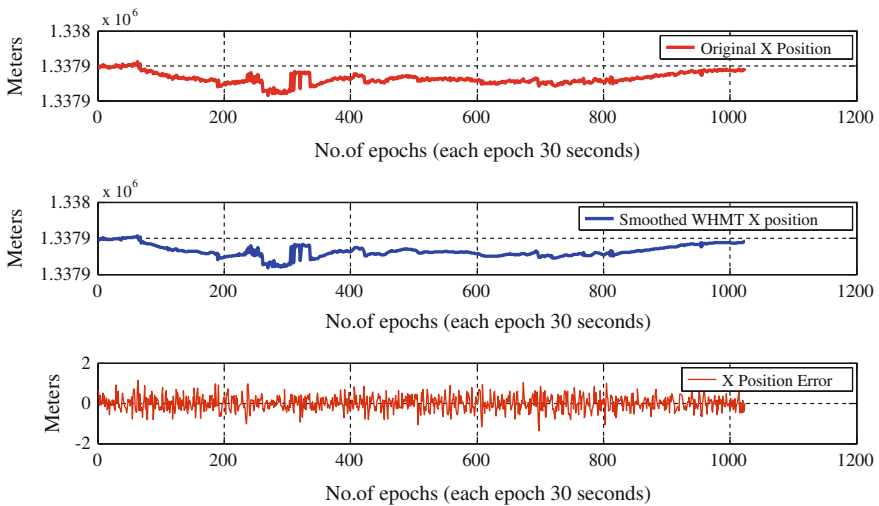


Fig. 5 Original and Smoothed X position

proposed one compared to WHMT. The three wavelet base functions, i.e., daubechies (DB6), coiflet 5, and symelet 6, are used in this analysis. In WHMT, Wavelet multiresolution approach is used to decompose the position coordinates data and applied median threshold estimator with soft thresholding. In this analysis, eight levels of decomposition are performed, and each level is estimated with median estimator. The result shows that the average error reduction is better with DB6 when compared with the Coiflet 5 and Symmlet 6 in WHMT based de-noising. The denoised data derived from traditional and WHMT is shown in Figs. 5, 6, and 7.

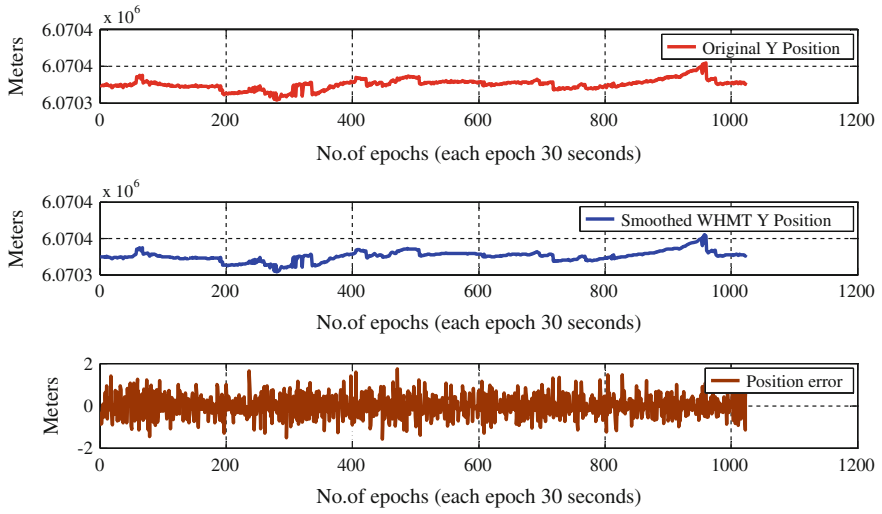


Fig. 6 Original and Smoothed Y position

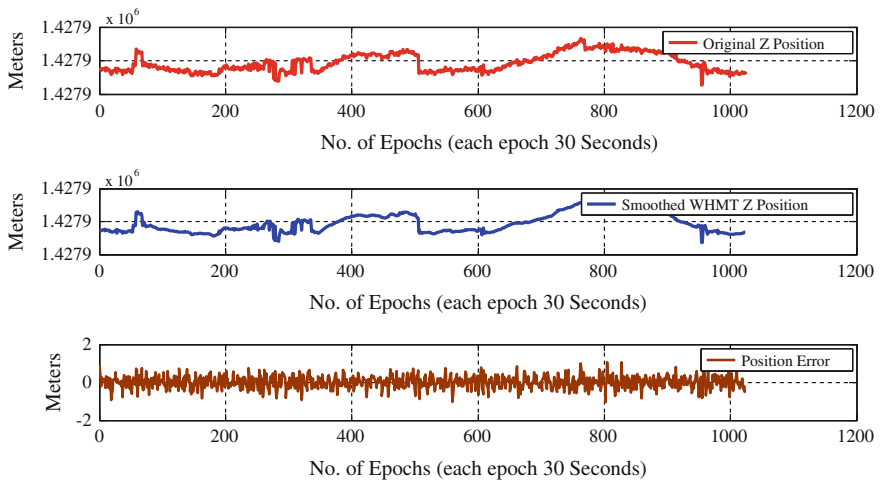


Fig. 7 Original and smoothed Z position

Table 2 WHMT De-noising

Station name	Wavelets	Month and day	RMSE		
			X	Y	Z
Bangalore (IISC)	Db6	Jan 15th	0.8217	1.2540	0.5684
		May 15th	0.3578	0.6367	0.3297
		Oct 15th	0.4259	0.4923	0.3194
	Symmlet 6	Jan 15th	0.5316	0.4977	0.3133
		May 15th	0.3924	0.5622	0.4232
		Oct 15th	0.4079	0.4532	0.3887
	Coiflet 5	Jan 15th	0.4998	0.5759	0.3413
		May 15th	0.3733	0.7076	0.3701
		Oct 15th	0.4113	0.4120	0.4262

Table 3 Cascade WHMT

Station name	Wavelets	Month and day	RMSE		
			X	Y	Z
Bangalore (IISC)	Db6	Jan 15th	0.4295	0.4650	0.2256
		May 15th	0.2827	0.5747	0.2747
		Oct 15th	0.2876	0.4393	0.2573
	Symmlet 6	Jan 15th	0.3064	0.4144	0.2108
		May 15th	0.2264	0.5096	0.2182
		Oct 15th	0.2737	0.4390	0.2217
	Coiflet 5	Jan 15th	0.1126	0.1101	0.1875
		May 15th	0.1253	0.1528	0.0863
		Oct 15th	0.3064	0.4144	0.2108

Table 4 Percentage of improvement

Method	X	Y	Z	Total	Improvement
WHMT	2.6441	1.3242	1.2376	5.2019	62 %
Cascade WHMT	1.1201	1.1321	0.9842	3.2364	

7 Conclusions

In this analysis, the solution for de-noising of GPS position data by combining the Code Carrier smoothing with WHMT is presented. The results of the proposed method are compared with traditional WHMT de-noising method. To establish advantage the proposed method, the RMSE is computed for both methods. In this analysis, the spectral errors treated as noise and removed effectively with the proposed method aside removing temporal errors of GPS. The test results have shown that the proposed algorithm is helpful where the precise accuracy is

important. Because of its simplicity and easy implementation, the basic hatch filter is used for code carrier smoothing. In future research, this methodology can take an advanced code carrier smoothing algorithm instead of hatch to obtain better positional accuracy.

Acknowledgments The author Ch Mahesh, expresses sincere thanks to the team of Nav-Aid, Hyderabad International Airport limited, Hyderabad and S. Nandulal, Asst. GM (CNS), GAGAN, AAI, for their continuous support and valuable suggestions of this project.

References

1. Misra, P., Enge, P.: *Global Positioning System (GPS), Signals, Measurements and Performance*. Ganga-Jamuna Press, P.O.Box-633, Lincoln, MA 01773. ISBN:0-9709544-1-7
2. Fu, W.X., Rizos, C.: The applications of wavelets to GPS signal processings. In: 10th Intl. Tech. Meeting of the Satellite Division of the U.S Inst. of Navigation, Kansas City, Missouri, pp. 1385-1388, 16–19 Sept 1997
3. Satirpod, C., Ogeja, C., Wang, J., Rizos, C.: An approach to GPS analysis incorporating wavelet decomposition. *Artif. Satell.* **36**, 27–25 (2001)
4. Rife, J., Tufts University, Sen, S., Tufts University: Design of a Single-Frequency Filter that Minimizes Ionosphere Divergence Error
5. Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation via wavelet shrinkage. *Biometric* **81**, 425–455 (1994)
6. Adhika Pradipta Lie, A.: Hatch Filter: Single and Dual—Frequency Carrier Smoothing. (2006)
7. Sunehra, Dhiraj: Validation of GPS receiver instrumental biases results for precise navigation. *Indian J. Radio Space Phys.* **42**, 175–181 (2013)
8. Hwang, P., McGraw, G., Bader, J.: Enhanced differential GPS carrier-smoothed code processing using dual-frequency measurements. *Navigation* **46**(2), 127–137 (1999)
9. Crouse, M.S., Nowak, R.D., Baraniuk, R.G.: Wavelet-based statistical signal processing using hidden markov modes. *IEEE Trans. Signal Proc.* **46**, 886–902 (1998)
10. Romberg, J.K., Choi, H., Baraniuk, R.G.: Bayesian tree—structured image modeling using wavelet–domain hidden markov models. In: *Proceedings of SPIE, Denver, CO*, vol. 3816, pp. 31–44, July 1999
11. Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**(432), 1200–1224 (1995)

A Corpus of Word-Level Offline Handwritten Numeral Images from Official Indic Scripts

Sk Md Obaidullah, Chayan Halder, Nibaran Das and Kaushik Roy

Abstract Dataset development is one of the most imperative tasks in document image processing research. The problem becomes more challenging when it comes about Numeral Image Database (*Nldb*) for official Indic scripts. Few efforts are made so far but they were restricted on single script which is basically a local script of the fellow researcher who prepared the database. In this paper, a technique for development of a handwritten *Nldb* of four popular Indic scripts namely Bangla, Devanagari, Roman and Urdu is proposed. Initially data were collected in unconstrained manner at Word-level from different writers with varying age, sex and educational qualification. All the images are stored in grey-level at .jpg format so that the data can be used in various ways as per need. A benchmark result on the present dataset is proposed using a novel hybrid approach with respect to Handwritten Numeral Script Identification (*HNSI*) problem.

Keywords Document image analysis · Numeral image database · Handwritten numeral script identification · Wavelet radon transform · Benchmarking

S.M. Obaidullah (✉)

Department of Computer Science & Engineering, Aliah University, Kolkata, W.B, India

e-mail: sk.obaidullah@gmail.com

C. Halder · K. Roy

Department of Computer Science, West Bengal State University, Kolkata, W.B, India

e-mail: chayan.halder@gmail.com

K. Roy

e-mail: kaushik.mrg@gmail.com

N. Das

Department of Computer Science & Engineering, Jadavpur University, Kolkata, W.B, India

e-mail: nibaran@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International*

Conference on Computer and Communication Technologies, Advances

in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_67

1 Introduction

Script identification from document images is an essential task before choosing script-specific Optical Character Recognizer (*OCR*) for a country like India, where 23 (including English) different languages are present and 13 (including Roman) different scripts are used to write them [1]. Whenever the term *OCR* comes to our mind normally it is about the alphabetic characters. But numerals have also its importance in *OCR* related work. In our day to day life we come across various important documents where numeral handwritten characters play a major role. Bank check containing numeral account number, postal document having numeral PIN code etc. are very popular examples of such type of documents. To process these numeral strings by a script specific *OCR*, first we need to know by which script the string is written. That is why prior script identification is an essential step. Again the whole script identification work can be categorized into two types namely Printed Numeral Script Identification (*PNSI*) and Handwritten Numeral Script Identification (*HNSI*) depending on the input acquired. Developing *HNSI* technique are more challenging than *PNSI* due to dynamic nature of handwritten data compared to printed one. Again each of *PNSI* and *HNSI* techniques can be categorized into Document-level, Block-level, Line-level, Word-level and Character-level based on presence of the multi-script texts in real life document images.

To facilitate the work of *OCR*, different research groups across the world are deeply engaged towards developing standard databases for document images. But most of available databases are on document images having alphabetic characters. All these databases can be further categorized into Printed Image Database (*PIdb*) and Handwritten Image Database (*HIdb*) based on type of the document acquired (machine generated or human written). Detail about different handwritten image database can be found here [2].

But the above mentioned databases are mainly on alphabetic characters. The efforts on numeral database development are not so popular among researchers till date. Among the available ones, Chaudhuri [3] developed a Bangla numeral script database. A volume of 23392 offline isolated strings and numerals are present in this database. There are few efforts on different international scripts which are namely *NIST*, *MNIST* [4], *CEDAR* [5], *CENPARMI* [6–8] etc.

Form the quick survey it is indeed very clear that there is a necessity of Handwritten Numeral Image database (*HNIdb*) on official Indic scripts. In this paper we are trying to bridge this gap by providing a *HNIdb* at Word-level of four popular Indic scripts namely *Bangla*, *Devanagari*, *Roman* and *Urdu*. The benchmarking is also done with respect to the *HNSI* problem which is also not addressed so far by the researchers.

Bangla numeral system is followed by Bangla, Assamese and Manipuri languages. Different eastern Indian states and county like Bangladesh use this numeral script. Devanagari numeral system is followed by popular North Indian languages like Hindi, Sanskrit, Marathi etc. Four numeral digits of Devanagari script are almost structurally and visually similar with Bangla [9]. Roman is a popular script throughout the Indian

Sub-continent which is used by English and Santhali languages. Four numeral digits of Roman are almost structurally and visually similar with Bangla and Devanagari scripts [9]. This similar nature of Bangla, Devanagari and Roman numerals make the task very challenging when it comes about Handwritten Numeral Script Identification or Writer Identification problem considering these scripts. Urdu is a variant of Eastern-Arabic family of numerals. Urdu numeral script is very much structurally and visually distinct in comparison with Bangla, Devanagari and Roman. Only one Urdu numeral digit looks similar with other three.

Rest of the paper is organized as follows: Sect. 2 proposes the scheme for dataset generation and representation. Section 3 provides experimental details with benchmark results. Conclusion and future scopes are discussed in Sect. 4. Finally references of literature are mentioned in the last section.

2 Proposed Scheme for Database Generation and Representation

2.1 Format of Data Collection Form

A standard A4 size format using word processor software has been prepared for handwritten numeral data collection. It is basically a grid with 20 rows and 7 columns distributed over 140 boxes. At the below author identity information fields were provided. To understand the variability of the contributors four standard parameters namely name of the person, age, sex and educational qualification were provided. The writing environment was totally unconstrained with respect to pen type, ink color, writing time etc. Each writer was asked to write a numeral string of having minimum three or maximum five digits. Following Fig. 6 shows a standard data collection form having Bangla handwritten numerals.

2.2 Preprocessing

2.2.1 Segmentation

Different handwritten numeral words from the pre formatted data collection form have been collected by applying an automated segmentation technique. Special care has been taken for line touching words. The numeral words are segmented from the original page by applying a Line Structuring Element (*LSE*) whose dimension is fixed experimentally. Then morphological dilation operation is applied using *LSE* on the complemented version of the threshold image. It will create single block for each of the word image. Then component labeling is done and numeral word blocks are extracted applying bounding box technique on the original image file. Experimentally a threshold block area size (min 30000, max 60000) is provided to

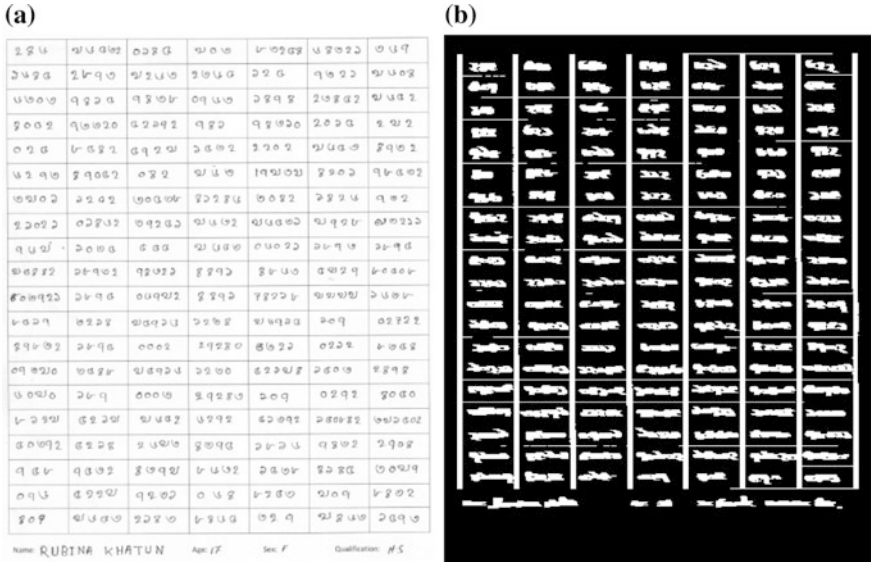


Fig. 1 a Original numeral data collection form with samples collected from a Bangla writer, b numeral word block extraction technique

avoid broken/undesired components. Measures are also taken to remove unwanted pixels at boundary region. Finally using a bicubic interpolation technique all numeral images are resized at a dimension of (140 × 340) pixels and stored in .jpg format. Figure 1 shows an original data collection form and an intermediate phase of segmentation where numeral word blocks are created.

2.3 Data Preparation

The collected forms have been digitized using HP flatbed scanner at 300 dpi. Word-level data are prepared from the original document pages using automated segmentation technique mentioned above. Each of the word blocks is stored as.jpg format to preserve the maximum image quality. Segmented word images are resized to maintain equal aspect ratio for each of the images. For present work the pixel ratio is maintained as (140 × 340). The present dataset is provided in grey-scale format so that other user can play with them as per their need. Following Fig. 2 shows sample Word-level numeral strings from our present dataset.

(a)	(b)	(c)	(d)
Bangla	Devnagari	Roman	Urdu
২৯৪৪	१५२२	91107	۲۱۵۴
৯০৪৫	२००९	67423	۶۸۹۱
৯০০৯	२४३২	1972	۶۶۲۲
৬৯০৬	२७৪	72621	۱۲۰۰
২৬৯৬	२६६	36336	۹۱۳
৬০০২	६९২৭	4315	۹۱۳۷২
৭৫৬২	৬২৭	12341	۳۳۳۳۳
৬৬২২	३३৩০	59761	۹২৫৫
২৭০২	२৫৬	9652	۱۳۴۱۸
৩৪৩৬	९২৩৩	23246	۱۶۸۹০

Fig. 2 Sample numeral words from our present dataset. **a** Bangla, **b** Devanagari, **c** Roman, **d** Urdu (left to right)

2.4 Statistical Distribution

Total 5659 Word-level handwritten numeral images have been collected for the present work. The whole dataset is distributed over four scripts in a quantity of 1602 words for Bangla, 1139 words for Devanagari, 1602 words for Roman and rest 1316 for Urdu. Total 43 different writers contributed to build the entire set of data. Out of these 43 persons number of writers for Bangla, Devanagari, Roman and Urdu was 12, 9, 12 and 10 respectively. Following Table 1 show the statistical

Table 1 Statistics distribution of the present dataset

Sl. no.	Script	Total words	No. of writers
1	Bangla	1602	12
2	Devanagari	1139	9
3	Roman	1602	12
4	Urdu	1316	10
5	Total	5659	43

distribution of our present dataset mentioning script name, word count and number of writers involved.

3 Experimental Details

3.1 Numeral Script Identification Problem

The requirement of *HNSI* techniques is already discussed in the introductory section. In the light of that discussion a benchmark results has been proposed here which is based on a novel hybrid technique named as Wavelet Radon Transform (*WRT*). The following section discuss briefly about the benchmarking technique and experimental results.

3.1.1 Feature Extraction Based on WRT

During experimentation grey-scale numeral word images have been binarized using our existing two-stage based binarization algorithm [10] which can be termed in short as global method on local blocks. The advantage of using two-stage based approach is that the result will be at least as good as if only global binarization algorithm would have been applied. Afterwards, these two tone images are analyzed in frequency domain. Wavelets [14] are successfully used so far in many image processing applications [11–13] for frequency domain analysis by dividing the original images into different sub-bands. Radon Transform computes projection of a 2D image matrix along a specified direction θ [14]. For feature extraction these two frequency domain techniques (Discrete Wavelet Transform and Radon Transform) has been combined to form a new hybrid technique named as *WRT*. The *WRT* features are computed as follows: Firstly, *DWT* decomposition of the input binary images is done using Daubechies *db4* decomposition and four sub-band images are produced at first level. Then *RT* is computed on each of them (*cA*, *cH*, *cV*, *cD*) at seven different rotation angles starting from 0° and ending at 180° varying with a distance of 30° . So the set $\theta = \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ \text{ and } 180^\circ\}$ for present experiment. Finally some local features namely entropy, mean, standard deviation are computed on each of the *WRT* spectrum to generate the final feature set. The classification process is carried out using MLP classifier where number of neurons in the input and output layers are forty two and four respectively. The number of hidden units, back Propagation learning rate and acceleration factor are set to suitable values, based on trial runs [15].

3.2 Experimental Result and Analysis

The present problem of *HNSI* from document images is a challenging task as the number of samples in each script is only ten (zero to nine). Among those ten samples few are also common to each other. As an example, the digit ‘zero’ is almost similar for Bangla, Devanagari and Roman scripts. This visual and structural similarity can make an adverse effect on the overall script identification performance. As per our observation, four numeral digits of Bangla and Devanagari and Roman scripts are visually and structurally almost similar. For present work extensive experimentation is carried out by evaluating the performances of Bi-scripts, Tri-scripts, Four-scripts combinations. Initially the whole dataset is divided into training and testing samples at a ratio of 2:1. The confusion matrix of Four-scripts combination on the test dataset is shown in Table 2. The average identification rate of Four-scripts combination has been found to be 65.5 % on the test dataset. Maximum misclassification is found among Bangla, Devanagari and Roman scripts. This is due to the structural similarity of the digits of those scripts as mentioned earlier. As per our observation performance of Urdu script is encouraging among the all. This is because except the numeral ‘one’ in Urdu which is very much similar with numeral ‘one’ in Roman, all other digits are visually and structurally distinct in comparison with other three. The evidence of our observation is supported by the misclassification rate of Urdu script, which is almost equal for Bangla, Devanagari and a bit higher for Roman (Bangla: 2.28 %, Devanagari: 2.97 % and Roman: 9.83 %). The graphical representation of Table 1 is shown by Fig. 8. The performances of Bi-scripts and Tri-scripts combinations can be found from Tables 3 and 4 respectively. The average Tri-scripts and Bi-scripts identification rate has been found to be 69.5 and 82.1 % respectively. Among the 4C_3 or four Tri-scripts combinations highest accuracy is reported for {Devanagari, Roman, Urdu} (73.1 %) combination, immediately followed by {Bangla, Roman, Urdu} (72.1 %) and {Roman, Urdu} (82.1 %). Lowest Tri-script accuracy rate is found for

Table 2 Confusion matrix using MLP classifier on the test dataset for four-scripts combination

Script name	Bangla	Devanagari	Roman	Urdu
Bangla	333	100	101	38
Devanagari	103	190	59	36
Roman	62	48	365	52
Urdu	10	13	43	371

Table 3 Tri-script identification rate using MLP classifier on the test dataset of 4C_3 combinations

Sl. no.	Script combination	Acc. rate (%)
1	{Devanagari, Roman, Urdu}	73.1
2	{Bangla, Roman, Urdu}	72.1
3	{Bangla, Devanagari, Urdu}	69.8
4	{Bangla, Devanagari, Roman}	62.9
5	Average tri-script	69.5

Table 4 Bi-script identification rate using MLP classifier on the test dataset of 4C_2 combinations

Sl. no.	Script combination	Acc. rate (%)
1	{Bangla, Urdu}	92.2
2	{Devanagari, Urdu}	90.7
3	{Roman, Urdu}	87.8
4	{Devanagari, Roman}	78.3
5	{Bangla, Roman}	73.5
6	{Bangla, Devanagari}	69.9
7	Average bi-script	82.1

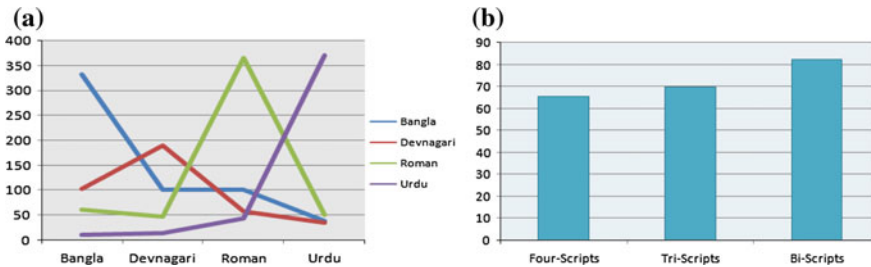


Fig. 3 **a** Confusion matrix for Four-scripts combination on the test dataset, **b** average accuracy rate of different script combination patterns i.e. Four-scripts, Tri-scripts and Bi-scripts using MLP classifier

{Bangla, Devanagari, Roman} script combination which is 6.6 % below the average Tri-scripts accuracy rate. It is to be noted that the Urdu script is not present in this combination. Table 3 shows the average Tri-scripts accuracy rate of 82.1 % for 4C_2 or six combinations. The pattern of the result is also similar here in comparison with Tables 2 and 3. Top three Bi-scripts identification rate has been found whenever there is a presence of Urdu script in the combination, which are {Bangla, Urdu} (92.2 %), {Devanagari, Urdu} (90.7 %), {Roman, Urdu} (87.8 %) correspondingly. For other three combinations the identification rate is below the average Bi-scripts identification rate. A graphical representation of the comparison of the average identification rate of Four-scripts, Tri-scripts and Bi-scripts combinations are shown in Fig. 3.

4 Conclusion and Future Scope

This paper presents a corpus of handwritten numeral document images of four popular Indic scripts. The necessity of *NSI* problem has been discussed here in the light of several real life examples. A benchmark result for *HNSI* problem on four Indic scripts using a novel hybrid approach has been proposed. The experimental results for different script combinations are critically analyzed and the reasons for

misclassification are pointed out. Future scope of the work includes enriching the corpus with more number of scripts and samples involving more writers. Scopes can be further extended to develop new techniques for *HNSI* and improving the present result to some possible extent. Finally it is worth to mention here that, the authors will be very happy to contribute the developed dataset to the document image processing researchers for non commercial use.

Acknowledgement The authors are very much thankful to Mr. Tousif Jaman and Mr. Sahaniaj Dhukra, students of Aliah University for their immense help during data collection process.

References

1. Obaidullah, S.M., Das, S.K., Roy, K.: A System for Handwritten Script Identification from Indian Document. *J. Pattern Recogn. Res.* **8**(1), 1–12 (2013)
2. Obaidullah, S.M., Rahaman, Z., Das, N., Roy, K.: Development of document image database for handwritten indic scripts. *Int. J. Appl. Eng. Res.* **9**(20), 4625–4630 (2014)
3. Chaudhury, B.B.: A complete handwritten numeral database of Bangla—A Major Indic Script. In: 10th International Workshop on Frontiers in Handwriting Recognition, France (2006)
4. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradientbased learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
5. Hull, J.J.: A database for handwritten text recognition research. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 550–554 (1994)
6. Saito, T., Yamada, H., Yamamoto, K.: On the database ELT9 of handprinted characters in JIS Chinese characters and its analysis (in Japanese). *Trans. IECEJ, J.* **68-D**(4), 757–764 (1985)
7. Al-Ohali, Y., Cheriet, M., Suen, C.: Databases for recognition of handwritten Arabic cheques. *Pattern Recogn.* **36**, 111–121 (2003)
8. Nomi, T., Matsui, T., Yamashita, I., Wakahara, T., Tsutsumida, T.: Tegaki Suji Database ‘IPTP CD-ROM1’ no Ichi Bunseki (in Japanese). In: 1994 Autumn Meeting of IEICE, D-309, September, 1994
9. Vajda, S., Roy, K., Pal, U., Chaudhuri, B.B., Belaid, A.: Automation of Indian postal documents written in Bangla and English. *Int. J. Pattern Recogn. Artif. Intell.* **23**(8), 1599–1632 (2009)
10. Roy, K., Banerjee, A., Pal, U.: A system for word-wise handwritten script identification for indian postal automation. In: Proceedings of IEEE India Annual Conference 2004, pp. 266–271 (2004)
11. Mandal, J.K., Sengupta, M.: Authentication/secret message transformation through wavelet transform based subband image coding (WTSIC). In: International Symposium on Electronic System Design 2010, pp 225–229, ISBN 978-0-7695-4294-2, Bhubaneswar, India, doi:10.1109/ISED.2010.50.2010
12. Bhateja, V., Urooj, S., Mehrotra, R., Verma, R., Ekuakille, A.L., Verma, V.D.: A composite wavelets and morphology approach for ECG noise filtering. *PREMI* 2013, pp. 361–366
13. Dey, N., Das, A., Chaudhuri, S.S.: Wavelet based normal and abnormal heart sound identification using spectrogram analysis. *Int. J. Comput. Sci. Eng. Technol. (IJCSSET)*, **3**(6) (2012). ISSN: 2229–3345
14. Matlab Documentation: http://www.mathworks.in/help/pdf_doc/images/images_tb.pdf. Accessed Mar 01 2015
15. Obaidullah, S.M., Mondal, A., Das, N., Roy, K.: Script identification from printed indian document images and performance evaluation using different classifiers. *Appl. Comput. Intell. Soft Comput.* 2014, Article ID 896128, 12 (2014). doi:10.1155/2014/896128

A Planar Ultra-Wide Band Antenna Design Using Circularly Truncated Corners and Notches

H.S. Mewara, M.M. Sharma, Mayank Sharma,
Mukesh Gupta and Ajay Dadhich

Abstract This paper describes the design and parametric study of a planar rectangular ultra-wide band antenna (UWB). The proposed design exploits U -shaped notch and T -shaped slot with circularly truncated corners on the radiating patch along with stepped microstrip feed. Circular truncation is also used at the corners of the partial ground plane with a rectangular notch along with an electromagnetically coupled truncated rectangular strip. This antenna operates in a frequency band ranging from 2.975 to 10.684 GHz with an impedance bandwidth nearly 7.709 GHz. This antenna finds its applications in 5.2/5.8 GHz WLAN, 3.5/5.5 GHz WiMAX, 4 GHz C band, and lower frequencies of X band. The effect of the truncated ground plane and the dimensions of U , T and rectangular shaped notches on the optimization of the return loss is parameterized and discussed in detail.

Keywords Rectangular planar ultra-wide band antenna · Stepped microstrip feed · Circular truncated partial ground plane · Coupled truncated rectangular strip

H.S. Mewara (✉) · M.M. Sharma · M. Sharma · M. Gupta · A. Dadhich
Government Engineering College Ajmer, Near Narelli Jain Temple,
NH-8, Badliya Circle, Ajmer, India
e-mail: hsmewara@gmail.com

M.M. Sharma
e-mail: mmsjpr@gmail.com

M. Sharma
e-mail: msecajmer@gmail.com

M. Gupta
e-mail: mg.gupta@gmail.com

A. Dadhich
e-mail: ajaydadhich13@gmail.com

1 Introduction

As the Federal Communications Commission (FCC) has approved the license free use (for short range and low power) of ultra-wide band (3.1–10.6 GHz) on 14th February, 2002, the UWB applications has shown the myriad of challenges and opportunities to researchers for designing of antennas for UWB applications [1]. Various techniques have been proposed for designing an UWB patch antenna like slotting, notching, partial ground, defected ground, stepping, trident fed, parasitic strips, protruded strips, etc. but most of the proposed designs do not operate over entire specified ultra-wide band ranging from 3.1 to 10.6 GHz, which is the key feature of the proposed design [2–7]. There has been a large increase in the demand of multiband antennas for wireless applications. An UWB antenna must operate over entire frequency band allocated by FCC with low power characteristics and $VSWR \leq 2$ for entire band. The size and structure of radiating patch and ground plane is the major concern in designing an UWB antenna [8].

The geometry of proposed UWB antenna is shown in Fig. 1. In this paper, a rectangular planar UWB antenna is proposed using one *U*-shaped notch and one *T*-shaped slot with two circularly truncated corners on the radiating patch along with stepped microstrip feed. The design also includes circularly truncated cornered partial ground plane with a rectangular notch and an electromagnetically coupled truncated rectangular strip. The *U* and *T*-shaped notches are used to provide additional bandwidth so as to make the proposed antenna desirable for UWB applications. As the ground plane beneath the radiating patch is partially removed

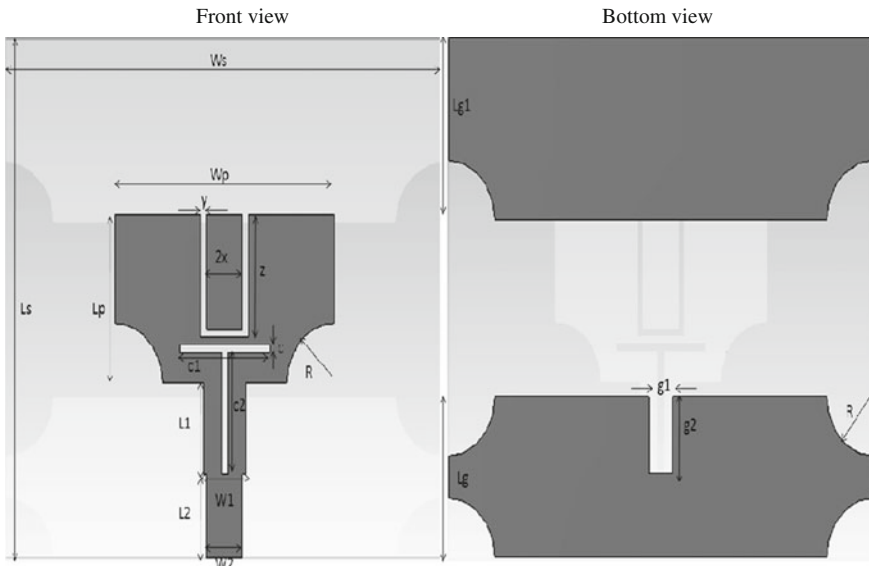


Fig. 1 Geometry of the proposed antenna with parameters

the bandwidth of the planar microstrip patch antenna increases tremendously. Moreover, the impedance matching is achieved by the use of stepped microstrip feed and by optimizing the length of the ground plane [3].

2 Antenna Design

The configuration and geometry of proposed ultra-wide band antenna is shown in Fig. 1. The proposed antenna is constructed on FR-4 substrate of 1.6 mm thickness with dielectric constant of 4.8 and loss tangent ($\tan \delta$) of 0.02. The stepped microstrip feed of this antenna is divided in two parts of length L_1 and L_2 with W_1 and W_2 width, respectively. The rectangular patch with $W_p \times L_p$ dimensions consist of two circularly truncated corners on the feeding edge. The radiating patch also includes one U -shaped notch with x , y , z and one T -shaped slot with c , c_1 , c_2 dimensions, respectively. In addition to this, for appropriate matching of impedance the distance between the radiating patch and ground plane is optimized using $d = L_f - L_g$, where ' d ' is the distance between patch and ground plane, ' L_f ' is the length of feed, and ' L_g ' is the length of ground plane [2]. Furthermore, circularly cornered ground plane with a rectangular notch of dimension $g_1 \times g_2$ is used along with two circularly truncated corners of rectangular parasitic strip. In the proposed design, U , T and rectangular notches with circular corners in radiating patch and ground plane is used which increases the surface current path resulting in additional resonance along with increase in impedance bandwidth. Circular truncation in the ground plane results gives smooth transition from one resonant mode to another ensuring good impedance matching and stable gain. The optimal designed parameters of the proposed antenna are as follows: $W_s = 36$ mm, $L_s = 34$ mm, $W_p = 18$ mm, $L_p = 11$ mm, $W_1 = 3.5$ mm, $L_1 = 6$ mm, $W_2 = 3$ mm, $L_2 = 5.5$ mm, $x = 1.5$ mm, $y = 0.5$ mm, $z = 8$ mm, $L_g = 10.55$ mm, $g_1 = 2$ mm, $g_2 = 5$ mm, $R = 3.9$ mm, $L_{g1} = 11.95$ mm, $c = 0.5$ mm, $c_1 = 7.5$ mm, $c_2 = 8$ mm, $h = 1.6$ mm, and $m_t = 0.035$ mm.

3 Results and Discussions

The proposed UWB antenna is designed and parameterized by selecting the optimal and best values of the various above said parameters. In this parametric study, the optimal value of each parameter was chosen and remaining parameters were optimized by fixing it. The Computer Simulation Technology Microwave Studio (CST-MWS 11) is used for this parametric study and simulation results were obtained.

At first, the effect of stepped microstrip feed was investigated and the lengths (L_1 and L_2) and widths (W_1 and W_2) of this asymmetric structure were optimized for best results. Fairly good impedance matching is achieved by using stepped microstrip feed. As the values of L_1 , L_2 , W_1 and W_2 are varied the return loss is

improved tremendously until the optimal value is reached. The lower frequency of the return loss increases as the lengths of the above parameters is increased.

Secondly, the dimensions of U -shaped notch is optimized which adds to the bandwidth of proposed antenna. As the values of x and y is increased the peak value of return loss is improved at lower and the upper frequencies of the return loss curve but this leads to decrease in peak value of return loss at mid frequencies. Length of the U notch, i.e., z is controlling parameter of both bandwidth and peak S_{11} . As the value of z is increased, the upper frequency peak of return loss curve shifts to lower frequency and the peak S_{11} at lower frequency decreases with bandwidth which can be easily seen in Fig. 2.

Then after the dimensions of T -shaped notch is varied to obtain the additional bandwidth for ultra-wide band operation. The upper frequencies are also sensitive to variation of c and c_1 , the peak S_{11} shifts to upper frequencies on increasing the values of these parameters. In addition to this, the peak value of the return loss curve is improved by varying the parameter c_2 . The dimensions of the rectangular notch on the ground plane, i.e., g_1 and g_2 are varied thereafter. As the ground plane beneath the radiating patch is removed partially the bandwidth increases. This notch provides additional impedance matching between ground plane and the radiating patch. By virtue of the dimensions of this notch the S_{11} improves. The bandwidth of the proposed antenna increases with increasing g_1 and decreases with increasing g_2 .

The dimensions of the ground plane and the parasitic rectangular strip, i.e., L_g and L_{g1} are varied to achieve good matching of impedance. As the value of L_g increases the peak S_{11} at lower frequency improves. In the last, the radius ' R ' of the circular truncation at the corners of the structure is optimized by parameter variation method. The circularly truncated corners are used to make proposed antenna fit for ultra-wide band operations. The impedance bandwidth is sensitive to the variation of the

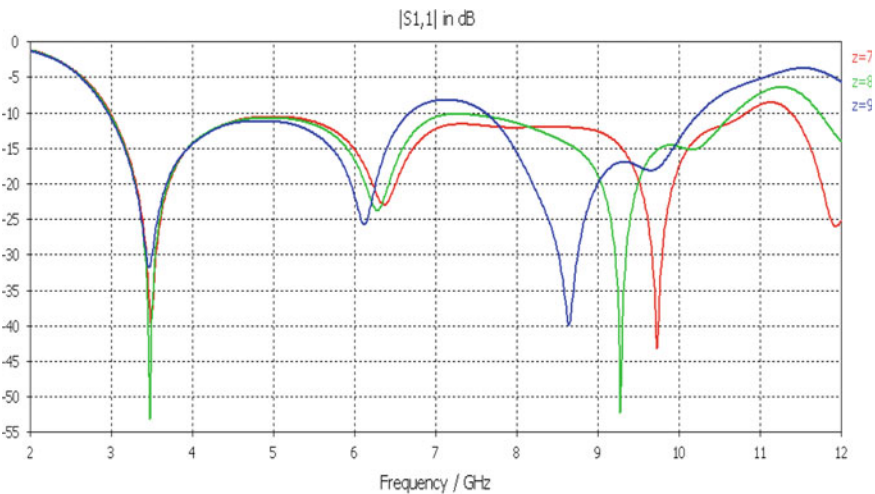


Fig. 2 Variation of parameter z of U -shaped notch

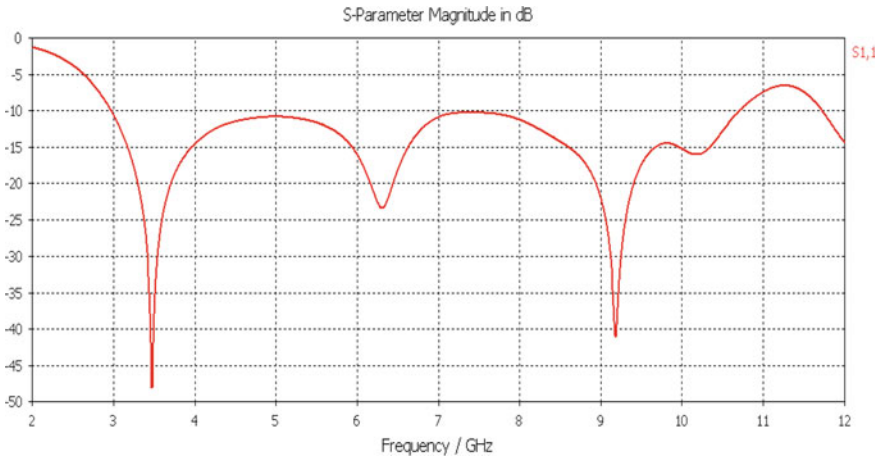


Fig. 3 Return loss (S_{11}) curve of the proposed antenna

parameter 'R'. The return loss curve for the proposed antenna is shown in Fig. 3. The peak value of return loss curve for the proposed antenna is nearly -48 dB at 3.47 GHz and for entire operating bandwidth the VSWR is less than 2.

One parasitic rectangular strip with two circularly truncated corners is used in the ground plane to provide additional current path so that the electromagnetic coupling between patch and ground plane is improved. The polar plot (Gain) and three-dimensional (Directivity) radiation patterns of the proposed antenna in shown in Figs. 4 and 5. The proposed antenna has 6.489 dBi at 9.18 GHz of directivity.

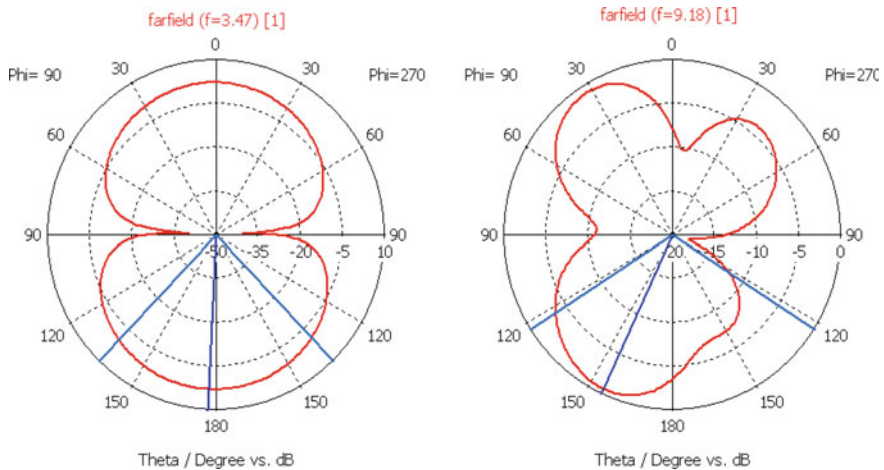


Fig. 4 The polar plot (gain) radiation patterns

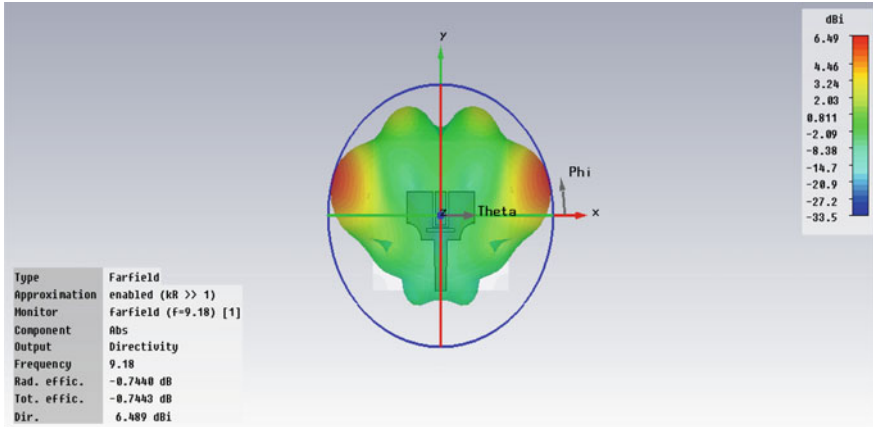


Fig. 5 The three-dimensional (directivity) radiation pattern

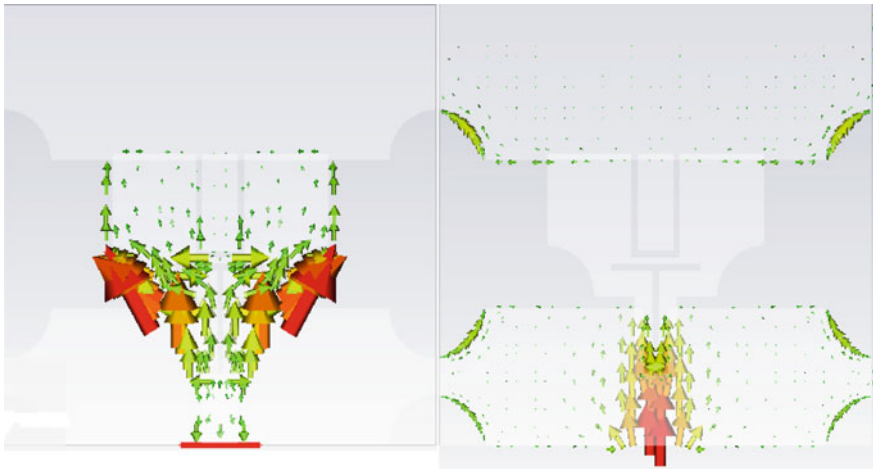
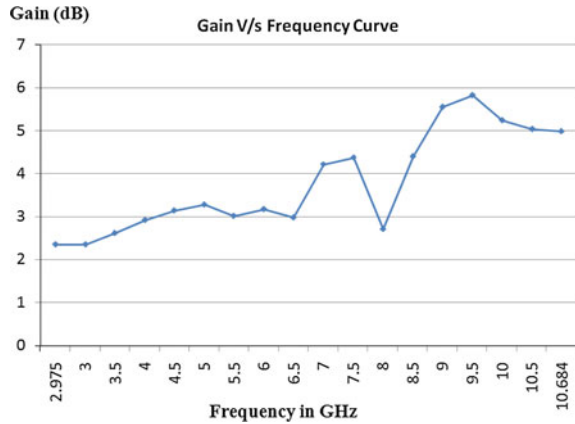


Fig. 6 Current distribution at 3.47 GHz

The current is distributed all through the radiating element and the reflecting plane which is shown in Fig. 6 which is simulated at 3.47 GHz.

The gain is almost constant between 2.348 and 5.835 dB for the entire ultra-wide band. The gain-frequency curve of the proposed antenna is shown in Fig. 7.

Fig. 7 Gain-frequency curve of the proposed antenna



4 Conclusions

In this paper, a rectangular planar ultra-wide band antenna is proposed. This antenna has been designed with notches, stepped microstrip feed and circularly truncated corners on radiating patch and ground plane. The dimension of each parameter used for designing the proposed antenna was optimized for improved return loss curve and ultra-wide band operations. This paper demonstrates a detailed parametric study of U , T and rectangular notches along with stepped microstrip feed and circular corners to make the proposed antenna well suited for UWB applications.

References

1. ET Docket 98–153: First report and order in the matter of revision of part 15 of the commission’s rules regarding ultra-wide band transmission systems. Federal Communications Commission, Washington, DC, FCC 02–48, 22 Apr 2002
2. Moosazadeh, M., Ghobadi, C., Dousti, M.: Small monopole antenna with checkered—shaped patch for UWB application. *IEEE Antennas and Wireless Propagation Letters*, vol. 9 (2010)
3. Lim, K.S., Nagalingam, M., Tan, C.P.: Design and construction of microstrip UWB antenna with time domain analysis. *Progr. Electromagn. Res. M* **3**, 153–164 (2008)
4. Rafiqul, I.Md., Zahirul, A.A., Khan, M.F.A.J., Alkaraki, S.: Design of microstrip patch antenna using slotted partial ground and addition of stairs and stubs for UWB application. *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT)*, May Edition (2012)
5. Lee, C.P., Chakrabarty, C.K.: Ultra wideband microstrip diamond slotted patch antenna with enhanced bandwidth. *Int. J. Commun. Network Syst. Sci.* **4**, 468–474 (2011)
6. Ojaroudi, M., Yazdanifard, Sh., Ojaroudi, N., Naser-Moghaddasi, M.: Small square monopole antenna with enhanced bandwidth by using inverted t-shaped slot and conductor-backed plane. *IEEE Trans. Antennas Propag.* **59**(2) Feb 2011

7. Zitouni, A., Boukli-Hacene, N.: A new four truncated corners ultra-wideband antenna with two crossed slits in the path. *Int. J. Microwave Opt. Technol.* **6**(4) (2011)
8. Mewara, H.S., Sharma, M., Sharma, M.M., Dadhich, A.: A novel ultra-wide band antenna design using notches and stairs. In: 2014 International Conference on Signal Propagation and Computer Technology (ICSPCT), pp. 425, 429, 12–13 July 2014

Genetic Algorithm for k -Connected Relay Node Placement in Wireless Sensor Networks

Suneet K. Gupta, Pratyay Kuila and Prasanta K. Jana

Abstract Wireless Sensor Networks (WSNs) are widely used for many applications including health care, environment monitoring, underground mines, and so on. In WSN, deployment of relay nodes to cover specific region or target is an important issue. In a target-based WSN, it is important that all the targets must be covered by sensor nodes, and the sensor nodes are connected with the backbone network. In this paper, we propose two algorithms for relay node placement which provide k -connectivity of the sensor nodes. The first algorithm is based on Genetic Algorithm (GA), and the second one is based on greedy approach. We have also to extensively simulate both the algorithms to study their performance.

Keywords Potential positions · Relay node · k -connected · Wireless sensor networks · Genetic algorithm

1 Introduction

The design of wireless sensor networks (WSNs) is a complicated task with substantial impact on its efficiency design and cost in real-life applications. Sensor nodes are tiny electronic devices with limited memory, energy, and transmitting capabilities [1]. An important goal of this type of network is to prolong the life time and make highly fault tolerant so that the sensor nodes cover a region of interest and forward the important data to the remote base station (BS) directly or by multi-hop

S.K. Gupta (✉)

Department of Computer Science and Engineering, OPJIT, Raigarh 496001, India
e-mail: suneet.banda@gmail.com

P. Kuila

School of Computer Engineering, KIIT University, Bhubaneswar 751024, India
e-mail: pratyay_kuila@yahoo.com

P.K. Jana

Department of Computer Science and Engineering, ISM, Dhanbad 826004, India
e-mail: prasantajana@yahoo.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_69

communication. Connectivity among the sensor nodes is crucial for most of the applications because the possible partition of the network may cause undesirable effects on the amount of information forwarded to the interested users [2]. Connectivity defines that individual sensor nodes are connected with different relay nodes to forward its data to the BS, and k -connectivity indicates that each sensor node is connected with at least k different relay nodes. Here, our objective is to design an efficient algorithm to place minimum number of relay nodes to provide desired k -connectivity. Note that the above k -connectivity problem is NP-hard [2]. Therefore, GA-based algorithm or a greedy-based heuristic is suitable for finding the approximate solution.

In [3], authors have surveyed various strategies of node deployment to achieve coverage and connectivity. Meguerdichian et al. [4] have pointed out that the coverage objective is a measure of the Quality of Service (QoS) that is provided by a particular network design. Another solution to maximize network lifetime is to deploy redundant sensor nodes in an area and periodically use active and inactive nodes. Martins et al. [5] suggest to periodically re-arrange active and inactive nodes to prolong the network lifetime. The most popular proactive strategy for preserving the network connectivity in the presence of a faulty node is to carefully place redundant sensor nodes during or after the initial deployment of a WSN. The sensor placement can be grouped into two categories. The first tries to just establish connectivity between end points, i.e., $k = 1$ [6]. In the second category, higher degrees of connectivity are to be achieved, i.e., $k > 1$ [7]. Given the scope of this section, we focus on the second category.

In this paper, we have presented LPP formulation of k -connected relay node placement problem and two algorithms: one is based on GA, and another is greedy approach. We describe all the basic steps of GA, i.e., initial population, selection, crossover, and mutation for the solution of the problem. We perform the extensive simulation of our proposed algorithm and compare the results.

The rest of the paper is organized as follows. The terminologies and proposed work are described in Sect. 2. Experimental results are presented in Sect. 3 followed by conclusion in Sect. 4.

2 Terminologies and Proposed Work

The problem is given x number of potential positions and n number of targets, we have to select minimum number of potential positions to place relay nodes, so that the targets will be k -connected (for some value of k). Here, it is assumed that a sensor node is placed adjacent to a target. Therefore, we can say that target and sensor node is a single entity. Some pre-defined potential positions are there where we can place relay node to provide connectivity to the sensor nodes so that the sensor nodes can easily send their data to the base station (BS). It is also assume that the relay nodes have enough communication range to directly or indirectly communicate with the BS. We use the following terminologies in the proposed algorithms:

- (1) The set of target (sensor node) is denoted by $\tau = \{\tau_1, \tau_2, \dots, \tau_n\}$.
- (2) The set of relay nodes is denoted by $\zeta = \{\zeta_1, \zeta_2, \dots, \zeta_m\}$.
- (3) The set of pre-defined potential positions is denoted by $P = \{p_1, p_2, \dots, p_x\}$.
- (4) R_C denotes the communication range of the sensor nodes.
- (5) $\text{dist}(\tau_i, \zeta_j)$ denotes the Euclidian distance between τ_i and ζ_j .
- (6) $\text{ComSet}(\tau_i)$ denotes the set of potential positions those are within the range of τ_i . In other words,

$$\text{ComSet}(\tau_i) = \{p_j | \text{dist}(\tau_i, p_j) \leq R_C \text{ and } p_j \in P\}$$

Therefore, we have to place a relay ζ_r on any one of the potential positions from $\text{ComSet}(\tau_i)$ to ensure connectivity of τ_i .

- (7) W denotes the set of relay nodes, which provides k -connectivity to all targets and ω_i denotes the set of relay nodes that those are providing k -connectivity to τ_i . In other words, $\omega_i = \{\zeta_r | \text{dist}(\tau_i, \zeta_r) \leq R_C, \forall r, 1 \leq r \leq m\}$ and

$$W = \bigcup_{i=1}^n \omega_i$$

- (8) $\text{cov_set}(\zeta_j)$ denotes the set of targets (sensors) that those are within communication range of ζ_j . In other words,

$$\text{cov_set}(\zeta_j) = \{\tau_i | \text{dist}(\tau_i, \zeta_j) \leq R_C \text{ and } \tau_i \in \tau\}$$

2.1 LP Formulation for k -Connected Node Deployment Problem

Now, we address the k -connectivity problem where our basic objective is to minimize the number of relay nodes. Let p_i be a Boolean variable such that

$$p_i = \begin{cases} 1, & \text{If a relay node is placed on } p_i, \forall p_i \in P \\ 0, & \text{Otherwise} \end{cases} \quad (2.1)$$

So, the Linear Programming (LP) formulation of the k -connectivity problem can be formulized as follows:

$$\text{Minimize } Z = |W|$$

Subject to

$$\varpi_i \geq k, \forall i, 1 \leq i \leq n \quad (2.2)$$

$$p_i \leq 1, \forall i, 1 \leq i \leq x \quad (2.3)$$

The constraint (2.2) states that the every target is connected with at least k number of relay nodes, i.e., k -connected. The constraint (2.3) ensures that maximum one relay node can be placed on any potential position $p_i, \forall i, 1 \leq i \leq x$.

2.2 k-Connected Algorithm Using GA

Now we present the GA-based [8, 9] approach for providing the k -connectivity to all targets with chromosome representation, initial population selection, crossover, and mutation in the subsequent subsections as follows.

2.2.1 Chromosome Representation

We represent the chromosome as a string of relay nodes. The length of each chromosome is kept same as the number of target points. In a chromosome, i th gene value is a set of relay nodes those are placed on $ComSet(\tau_i)$ to ensure connectivity.

Let us assume a scenario with 11 targets 20 potential positions. Table 1 represents the individual target and the potential positions within its range. Figure 1 shows a chromosome for $k = 2$. Here, the gene value at position 7 is $\{10, 12\}$ which implies that the ζ_{10} and ζ_{12} are placed on p_{10} and p_{12} , respectively, and therefore, now τ_7 is connected with ζ_{10} and ζ_{12} . Similarly, τ_{10} and τ_{11} are connected with $\{\zeta_{14}, \zeta_{17}\}$ and $\{\zeta_{12}, \zeta_{14}\}$, respectively.

2.2.2 Fitness Function

The fitness value of a chromosome represents its qualification on the basis to provide k -connectivity to all targets using minimum number of relay nodes. We

Table 1 Targets and relay nodes

Targets (τ_i)	ComSet (τ_i)
τ_1	{1, 2}
τ_2	{1, 2, 7, 8, 20}
τ_3	{3, 4, 5, 20}
τ_4	{4, 5, 6}
τ_5	{5, 6, 10, 11}
τ_6	{7, 9, 14, 17}
τ_7	{10, 11, 12}
τ_8	{7, 8, 16}
τ_9	{16, 17, 18, 19}
τ_{10}	{14, 15, 17, 18}
τ_{11}	{12, 13, 14, 15}

Target no.	1	2	3	4	5	6	7	8	9	10	11
Gene value (ω_i)	{1, 2}	{1, 2}	{3, 5}	{5, 6}	{10, 11}	{7, 14}	{10, 12}	{7, 8}	{17, 19}	{14, 17}	{12, 14}

Fig. 1 Chromosome generated from Table 1

build a fitness function to evaluate the individual chromosomes of the initial population. Our aim is to find a minimal set of relay nodes with the consideration that this set provides k -connectivity to all the targets. Therefore, the fitness function for proposed work is to minimize $Z = |W|$.

Now, let us calculate the fitness value of the chromosome as shown in Fig. 1.

$$W = \bigcup_{i=1}^n \omega_i = \{ \{1, 2\} \cup \{1, 2\} \cup \{3, 5\} \cup \{5, 6\} \cup \{10, 11\} \cup \{7, 14\} \cup \{10, 12\} \cup \{7, 8\} \cup \{17, 19\} \cup \{14, 17\} \cup \{12, 14\} \} = \{1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 14, 17, 19\}$$

i.e, $Z = |W| = 13$

Therefore, with reference to chromosome (refer Fig. 1) only 13 out of 20 relay nodes are enough to provide 2-connectivity to all targets.

2.2.3 Selection

The selection process determines which of the chromosomes from the current population will mate (crossover) to create new chromosomes. For the selection process, we select some valid chromosomes with higher fitness value. The individuals with better fitness values have better chances of selection. There are several selection procedures, e.g., Tournament selection, Rank selection, Roulette-wheel, etc. In our work, we have used Roulette-wheel selection.

2.2.4 Crossover

The crossover operation takes place between two randomly selected chromosomes. To produce the new offspring from the selected parent chromosomes, we use 1-point crossover whereby a point is chosen at random, and the two parent chromosomes exchange their information after that point. The whole process is shown in Fig. 2.

2.2.5 Mutation

The mutation is applied at a selected gene rather than randomly selected gene. In mutation operation, that relay node is selected from a chromosome which has

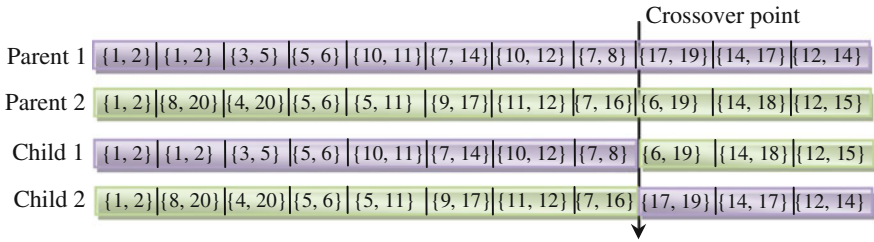


Fig. 2 Crossover operation

minimum appearance and tries to replace it with that relay node which has higher presence in the chromosome.

2.3 Node Deployment Based on Greedy Approach

In this method, the relay nodes are deployed for providing k -connectivity to all targets using greedy methods. For the same, there are two algorithms, e.g., Algorithm 1 and Algorithm 2, which are depicted in Figs. 3 and 4, respectively. In algorithm 1, we firstly generate $cov_set(\zeta_i)$.

This set stores all the targets, which is covered by ζ_i . After execution of Algorithm 1, Algorithm 2 will start its execution (refer Fig. 4). In Algorithm 2, output of Algorithm 1 acts as an input. The step 1 of Algorithm 2 sorts the ζ_i on $|cov_set(\zeta_i)|$, i.e., number of targets covered by ζ_i . After then sub-step 2.1 selects that relay node ζ_i covers maximum number of targets. Sub-step 2.2 forms the resultant set of the relay nodes (W) which provide the k -connectivity to all targets.

Fig. 3 Algorithm for generating $cov_set(\zeta_i)$

Algorithm 1: Formation of $cov_set(\zeta_i)$

Input: Coordinates of targets and relay nodes
Output: Generation of $cov_set(\zeta_i)$, $\forall \zeta_i \in \zeta$
Procedure:
Step 1: /* Formation of $cov_set(\zeta_i)$, $\forall \zeta_i \in \zeta$ */
 for all $\zeta_i \in \zeta$
 for all $\tau_i \in \tau$
 if ($dist(\tau_i, \zeta_i) \leq R_C$)
 1.1 **then** $cov_set(\zeta_i) = \tau_i$
 endif
 end for
 end for
Step 2: *Stop.*

Fig. 4 Greedy algorithm to find out W

Algorithm 2: GreedyToFind(W)

Input: $cov_set(\zeta_i), \forall i, 1 \leq i \leq m$.

Output: Set of relay nodes which provide k -connectivity.

Procedure:

Initialization: $W = \{\}$

Step 1: Sort $\zeta_i, \forall i, 1 \leq i \leq m$ in according of $|cov_set(\zeta_i)|$.

Step 2: **While** ($\tau_r, \forall r, 1 \leq r \leq n$ is not k -connected)

2.1: Select the ζ_i with highest $|cov_set(\zeta_i)|$.

2.2: $W = W \cup \{\zeta_i\}$

End while

Step 3: **Stop.**

3 Experimental Results

We performed the experiments on the proposed algorithm using MATLAB R2012b. We depict the experimental results for the minimum number of relay nodes required for k -connectivity. For simulation purpose, we considered two different network scenarios (WSN #1 and WSN #2), and both of them have the sensing field of 100×100 and 300×300 square meter area, respectively. In both the scenarios, we have formed the grid of having size 10 m and position of relay nodes at intersecting points of these grids except boundary of the scenario. For execution of our proposed GA-based algorithm, we considered an initial population of 60 chromosomes. During selection operation of GA, we have selected 10% best population using Roulette-Wheel for performing 1-point crossover operation. We compare the simulated results of both algorithms.

Figure 5a, b shows the comparison of algorithms in terms of number of relay nodes required to provide $k (\geq 2)$ connectivity in scenarios WSN #1 and WSN #2. With this figure, it is clearly shown that GA-based approach required less number of relay nodes to provide k -connectivity than greedy-based approach. GA-based proposed algorithm performed better than proposed Greedy-based approach because GA always gives the global optimal solution while greedy gives the local optimal solution. Moreover, in

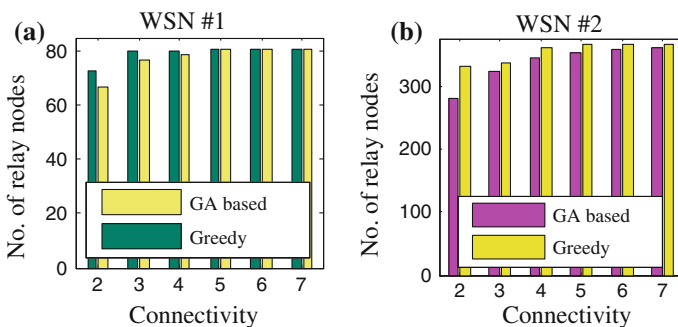


Fig. 5 Number of relay nodes in **a** WSN #1, **b** WSN #2

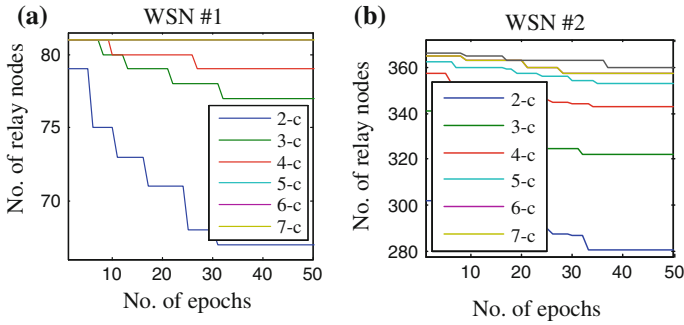


Fig. 6 Convergence of number of relay nodes in **a** WSN #1, **b** WSN #2

greedy, we make the local best choice on the basis of fitness function but in GA, choice is made with considering that in each iteration, it strengthens the value of fitness function. Figure 6a, b represents the convergence of fitness value of proposed algorithm in WSN #1 and WSN #2 scenarios, respectively. From the figure, it is clearly depicted that proposed algorithm starts giving the optimized value between 30 and 40 epochs. Both algorithms provide k -connectivity so these also provide the fault tolerance up to failure of $k - 1$ relay nodes complementary.

4 Conclusions

In this paper, we have proposed genetic algorithm-based node deployment algorithm for providing the k -connectivity to every targets (sensors). Here, we have given explanation for LPP formulation and step-by-step execution of all basic steps of GA with suitable examples. We have also proposed node deployment policy based on greedy approach. In the greedy approach, we arrange all the relay nodes based on number of connected targets, after then we start selecting the relay nodes. The algorithm stops execution till all the targets are not k -connected. From the experimental results, it is seen that GA-based approach is better than the greedy-based node deployment strategy. Our future research will be carried out to design energy aware relay node placement with generation of energy aware node disjoint sets.

References

1. Liu, X., Mohapatra, P.: On the deployment of wireless data back-haul networks. *IEEE Trans. Wireless Commun.* **6**, 1426–1435 (2007)
2. Konstantinidis, A., Yang, K., Zhang, Q., Zeinalipour-Yatzi, D.: A multi-objective evolutionary algorithm for the deployment and power assignment problem in wireless sensor networks. *Comput. Netw.* **54**, 960–976 (2010)

3. Younis, M., Akkaya, K.: Strategies and techniques for node placement in wireless sensor networks: a survey. *Ad Hoc Network* **6**, 621–655 (2007)
4. Meguerdichian, S., Koushanfar, F., Potkonjak, M., Srivastava, M.B.: Coverage problems in wireless ad-hoc sensor networks. *IEEE Infocom*, pp. 1380–1387 (2001)
5. Martins, F.V.C., Nakamura, F.G., Quintao, F.P., Mateus, G.R.: Modeland algorithms for the density, coverage and connectivity control problem in flat WSNs. In: *Proceedings of the International Network Optimization Conference (INOC'07)*, pp. 1145–1152 (2007)
6. Lloyd, E.L., Xue, G.: Relay node placement in wireless sensor networks. *IEEE Trans. Comput.* **56**(1), 134–138 (2007)
7. Nitesh, K., Jana, P.K.: Relay node placement algorithm in wireless sensor network. In: *Proceedngs of IACC-2014*, pp. 220–225. *IEEE* (2014)
8. Kuila, P., Gupta, S.K., Jana, P.K.: A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm Evol. Comput.* **12**, 48–56 (2013)
9. Gupta, S.K., Kuila, P., Jana, P.K.: GAR: an energy efficient GA-based routing for wireless sensor networks. In: *Proceedings of ICDCIT 2013*. LNCS (Springer), vol. 7753, pp. 267–277 (2013)

Segmentation of Cotton Bolls by Efficient Feature Selection Using Conventional Fuzzy C-Means Algorithm with Perception of Color

Sandeep Kumar, Manish Kashyap, Akashdeep Saluja
and Mahua Bhattacharya

Abstract Ad hoc method for segmentation of mature or nearly mature cotton bolls is proposed based on proper feature vector selection and efficient application of Fuzzy c-means (FCM) on images. Perception of color is used as fundamental criteria for segmentation. The results obtained are compared with conventional FCM and supremacy of the proposed work is presented. Since the technique is ad hoc, it will work only for the said purpose in the natural setting of cotton fields. Any improper acquisition of images of cotton bolls, like intense illumination or deep shadows (which is of course absent in natural settings) will produce improper results.

Keywords Fuzzy c-means · Image segmentation · Cotton boll segmentation · Perception of color

1 Introduction

To analyze the images of cotton bolls, the first necessary step is to segment the portion of cotton boll from the surrounding region, in which we are interested. Current work deals with segmentation of white portion of cotton bolls which can be

S. Kumar (✉) · M. Kashyap · A. Saluja · M. Bhattacharya
ABV Indian Institute of Information Technology and Management,
Gwalior 474010, M.P, India
e-mail: sandeep2006iiitm@gmail.com

M. Kashyap
e-mail: manishkashyap.iiit@gmail.com

A. Saluja
e-mail: akashdeep1412@gmail.com

M. Bhattacharya
e-mail: mahuabhatta@gmail.com

used to do further analysis in terms of shape-based features or texture-based features. The segmentation finally leads to classification of a randomly picked image whether it is mature or immature. All of this work is done to improve the efficiency of cotton boll picking systems (mechanical or electrical) so that immature cotton bolls are not picked by the system. Conventional process of cotton picking is manual which is time-consuming, error prone, with chances of wastage of cotton. The proposed technique is based on computer vision methodology which will automate the picking process and reduce error and wastage.

Similar kind of work, like [1, 2] describe the extraction of shape-based features of cotton top which would extract the cotton surface. In the same context, [3] established the statistical analysis and experiments. They gave a computer vision model which distinguished the mature cotton on the basis of its surface color and the background of the cotton field. Yong and Chang-ying [4] gave a new method of study for cotton identification and its surrounding. The method established the differences between the different parts of cotton such as cotton leaf and cotton stem. This model was originated to identify mature cotton from the surrounding. Establishment of the relation between shape-based features with real environment is an important task.

Similarly, [5] also established the work on the basis of statistical analysis. They explained the computer vision model that found the maturity of cotton in the environment using the color fringing knowledge. This technique is also known as purple fringing.

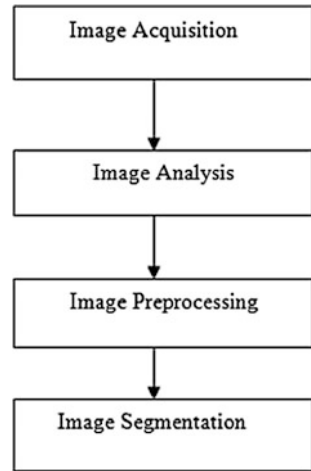
The rest of the paper is organized as follows. Section 2 illustrates the basic steps in image analysis. Section 3 illustrates the conventional fuzzy c-means algorithm in relevant detail. Section 4 is an important section where stress has been put on the fact using Generic FCM in various situations efficiently. Section 5 provides some facts that are useful for segmenting white portion of cotton and hence the ad hoc nature of this method comes from that. Section 6 illustrates the proposed method. Section 7 contains the results and Sect. 8 concludes the paper and gives the future work.

2 Image Analysis Steps

Image acquisition: The image acquisition system is the most important part of this work. The accuracy of this algorithm under different physical and environment variables depend on the acquisition of images. The images are taken by a digital camera of high mega pixel. The images were taken at different stages of growth of cotton flowers in natural variable lighting conditions. The locations were selected randomly on the basis of experiments and all the images are taken at different angles of sunlight. During collection of the images, the camera was placed above the cotton bolls scene and focused on the field at an inclined angle. A digital camera is needed to find the distance between the camera and the object in the scene.

Analysis of images was carried out in the laboratory using MATLAB software. All the images are in 24-bits color image of size 256×192 pixels and stored in

Fig. 1 Flowchart of image analysis



RGB color space in JPG file format. The algorithm is tested on RGB images taken before the preprocessing stage (Fig. 1).

Image analysis: This step involves extraction of important information from the images. This is the pre-required step before the preprocessing step. The step also includes the automatic analysis of the digital images.

Image Preprocessing: The images are often corrupted by noise during the image acquisition process which would hinder further image preprocessing. The presence of noise in cotton image obscures the important information present in the images and also hinders the process of image analysis. The Gaussian filter is used to remove the noise present in image without significant blurring, especially at edges [6–8]. The Gaussian is easily differentiated and integrated [9].

The Gaussian filter is the best known example of low pass filter (LPF) implemented with a nonuniform kernel. The mask coefficients for the Gaussian blur filter are samples from a 2D Gaussian function. The function is mentioned as

$$h(x, y) = \exp\left[\frac{-(x^2 + y^2)}{2\sigma^2}\right], \quad (1)$$

where h is the filter mask, (x, y) is the coordinates of image and σ is the variance.

Image Segmentation: This step involves the division of images into different segments to find out the region of interest (ROI) for further image analysis. After this, the segments are labeled [10, 11]. The main goal of the segmentation is to locate the objects and boundaries in images. The regions which are segmented may be similar or different on the basis of shape, color, intensity, or texture. The image is segmented by the variant fuzzy C-mean (FCM) technique.

3 Conventional Fuzzy C-Means Algorithm

Let us define the sample space of ‘ n ’ data points by the following set X [12]:

$$X = \{x_1 x_2 x_3 x_4 \dots x_n\},$$

where these ‘ n ’ data samples can be anything we wish to classify. For example, physique of human beings in which every data point, i.e., x_i can be a column vector comprising features like height, weight, etc. So for each data point x_i , we define the following feature vector:

$$x_i = \{x_{i1} x_{i2} x_{i3} \dots x_{im}\}$$

i.e., every data point is comprised of ‘ m ’ features and we have ‘ n ’ such data points. Every feature x_{ij} is to be normalized since ‘ m ’ features can all have different units. To classify the set X (the sample space or the universe of elements) into ‘ c ’ different classes of their own kind, lets again define a family of sets ‘ A ’ defined by $\{A_i, i = 1, 2 \dots c\}$ as a fuzzy c -partition of the universe. The partial membership of single point can have more than one class. It is defined as

$$\mu_{ik} = \mu_A(x_k) \in [0, 1], \tag{2}$$

where μ_{ik} is the present membership value of k th data point placed in the i th cluster. The membership here has a restriction that summation of membership values over all classes should be unity.

The family of fuzzy partition matrices, M_{fc} , for the classification involving ‘ c ’ classes and n data points, is defined as

$$M_{fc} = \left\{ U \mid \mu_{ik} \in [0, 1]; \sum_{i=1}^c \mu_{ik} = 1; 0 < \sum_{i=1}^c \mu_{ik} < n; \right\}, \tag{3}$$

where $i = 1, 2, 3, \dots, c$ and $k = 1, 2, 3, \dots, n$.

To determine the best value of matrix ‘ U ’ we need to minimize the following objective function with the symbols having usual meaning:

$$J_m(U, v) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2 \tag{4}$$

$$d_{ik} = d(x_k - v_i) = \left[\sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2} \tag{5}$$

Cluster centers are defined as follows:

$$v_{ij} = \frac{\sum_{k=1}^n \mu_{ik}^{m'} \cdot x_{kj}}{\sum_{k=1}^n \mu_{ik}^{m'}}, \quad (6)$$

where $m' \in [1, \infty)$. These variables restrict the values of fuzziness in the classification and the proper best fuzzy c -partition would be the smallest value of the partitions.

4 Generic FCM as a Tool

FCM is not just a classifier of given data, it is a complete tool to make classification better. To understand this better, let us consider the example of images. It is known that FCM as a generic classifier creates ‘ c ’ different classes without any prior knowledge about the spatial dependence on input data. However, in the images, no pixel is completely independent of its neighborhood. This fact was first considered by [13] and a modification to the objective function was suggested which lead to a better classification.

On similar grounds, increasing the dimensionality of the feature vector was also considered in some previous works. Data in higher dimensions are easily separable compared to low dimensional data. But for increasing the dimensions of the feature vector, justified analysis is to be done. So, the following two methods are important while designing an ad hoc application for segmentation [14–18].

- (a) Modification of objective function [$J(U, v)$].
- (b) Increasing the dimensionality of feature vector ‘ m ’.

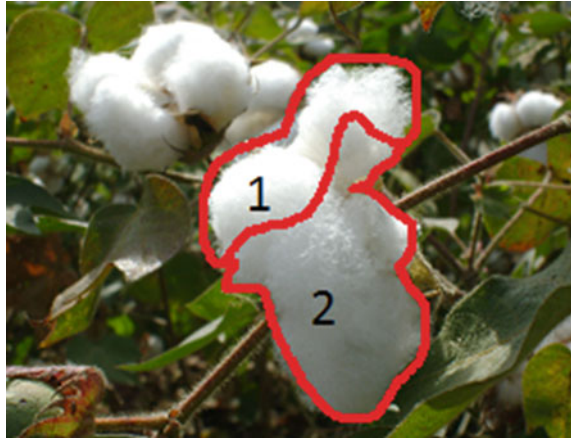
5 Cotton Bolls-Specific Facts

To segment the cotton bolls (i.e., the white portion) of the cotton from its surroundings, the following points must be noted:

- (a) Although the surface of cotton boll is perceived as white by humans but due to the presence of shadows on the surface, the surface no longer remains white. It may appear gray with the amount of darkness proportional to its facing toward the sun. It can be seen in Fig. 2.

Region 1 shown in Fig. 2 corresponds to near-white region, whereas region 2 corresponds to gray shades, although a general visual perception of whiteness is created.

Fig. 2 Two sub-segments of the visually perceived same segment



Color is absent on cotton boll (considering gray shade as no color). Due to the presence of ‘fur’-like texture on the surface, the illumination on the surface is not smooth, i.e., various grayscale intensity values can be found in clusters in nearby spatial locations.

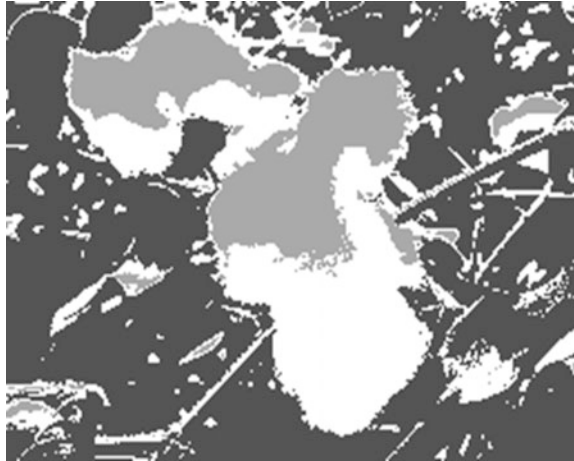
6 Proposed Method

Based on the facts presented in the previous sections, the following method for segmentation of cotton images is proposed as explained in the next paragraph.

Traditional methods of image segmentation using FCM or its variants consider the color of every pixel to be the only available feature for purposes of classification. However, for the current case, as described in the previous sections, surface of cotton is not uniformly colored. This may result in improper classification of the surface itself into two or more regions corresponding to dark and light sides as shown in Fig. 3.

However, it is known that the surface of cotton will be grayscale in nature while the background will be in various colors due to the presence of leaves and other objects. So, it is proposed that the feature vector corresponding to every pixel be modified so that instead of representing the color for every pixel, it should represent the ‘GRAYNESS’ or ‘COLORFULNESS’. Assuming RGB framework, the color of every pixel can be represented as a column vector of size (3×1) corresponding to the red, green, and blue frames. But classification merely based on these three features will produce results similar to Fig. 3. Hence, we modify the (3×1) feature vector as follows.

Fig. 3 Improper segmentation using conventional FCM



The new feature vector (3×1) (for every pixel) should be taken as

- (1) Value of RED component of color for that pixel MINUS its grayscale value.
- (2) Value of GREEN component of color for that pixel MINUS its grayscale value.
- (3) Value of BLUE component of color for that pixel MINUS its grayscale value.

Because this difference (or 1-norm) for every pixel and for every frame (Red, Green, or Blue) will be near zero if in the colored image the value is near grayscale, or it will be nonzero (higher in magnitude) if the pixel is colorful. This makes the process more robust for segmentation. The proposed technique shows more efficient segmentation of cotton boll compared to conventional FCM for further analysis. Hence, it is expected that the two wrongly classified regions in the above results should come out to be the same region. The results are shown below.

7 Experiments and Results

The results are generated on images of size 256×192 using the standard FCM implementation having feature vector size as 3 described above. The number of data samples is the total number of pixels in the image. Maximum number of iterations is set to 100 in case convergence is not achieved (rare case). Since this problem is a well-posed problem [14] the number of iterations, i.e., 100 is quite sufficient.

Figure 4 is the original colored cotton image, Fig. 5 is an application of FCM on it, and Fig. 6 is the result of application of the proposed method. Similarly, Figs. 7, 8 and 9 follow. It can be clearly verified that the white portion of cotton has been separated in the proposed method but not in conventional application of FCM.

Fig. 4 Original colored cotton image



Fig. 5 By conventional FCM

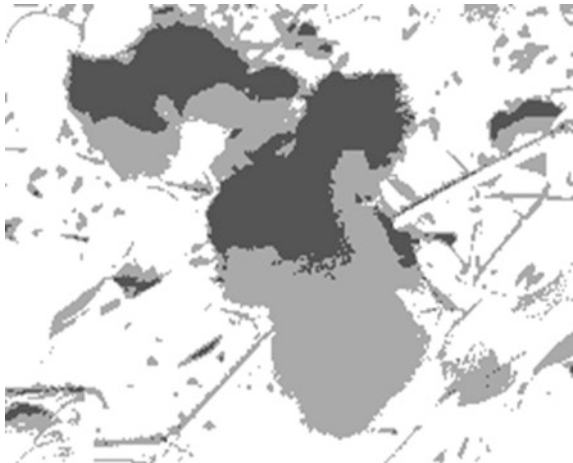


Fig. 6 By proposed method

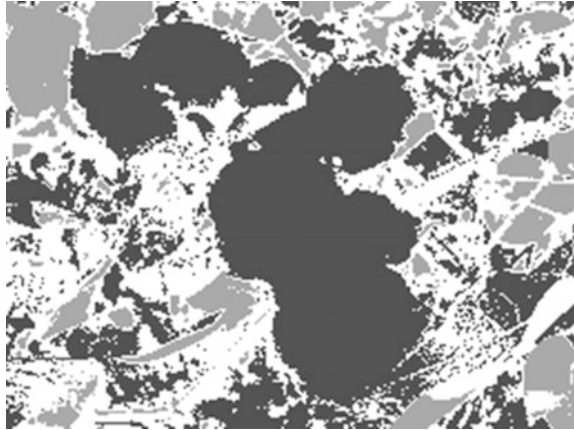


Fig. 7 Original colored cotton image



Fig. 8 By conventional FCM

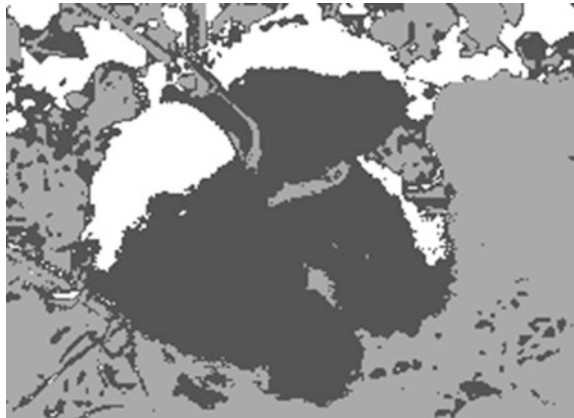
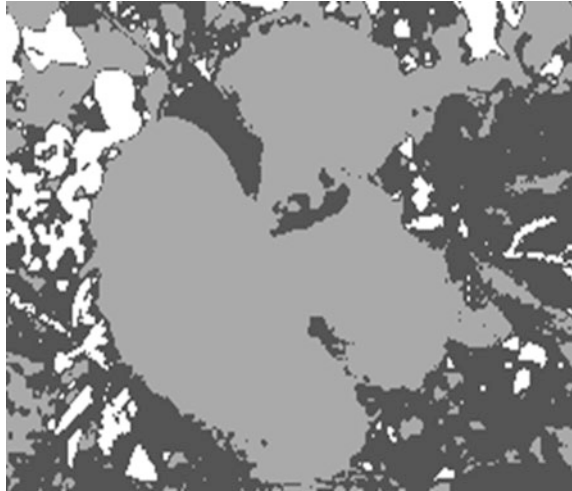


Fig. 9 By proposed method



8 Conclusion and Future Work

As is evident from the results, a better method for segmentation of cotton images has been proposed based on FCM with the color perception without modifying the objective function of FCM. Due to the increased complexity of FCM by increasing the size of feature vector, we leave it as the future work to test the proposed method on various resolutions (multi-resolution analysis). The aim of the proposed technique is to use FCM as a tool efficiently without modifying its objective function, since the change in the objective function may increase its complexity of computation.

Acknowledgments Authors like to acknowledge the support given by Department of Science & Technology, Ministry of Science & Technology, Government of India, Central Institute for Cotton Research, Nagpur under Indian Council of Agricultural Research (ICAR) and Central Mechanical Engineering Research Institute, Ludhiana under the Council of Scientific and Industrial Research (CSIR). This work is under the collaborative networking project Center for Precision and Conservation Farming Machinery (CPCFM) entitled: Vision-Based Expert System for Picking of Cotton.

References

1. Wang, M., Xu, K., Wei, J., Yuan, J.: A research for intelligent cotton picking robot based on machine vision. In: International Conference on Information and Automation, 20–23 June 2008
2. Wang, W., Qu, D., Ma, B., Wang, Y.: Cotton top feature identification based on machine vision and image processing. In: Computer Science and Automation Engineering (CSAE) (2011)

3. Yong, W., Chang-ying, Ming-xia, S.J.: Model and analysis of color for different parts of ripe cotton in picking period. In: Transactions of the CSAE, vol. 23, pp. 183–185 (2007)
4. Yong, W., Chang-ying, J.: Study on discrimination of mature cotton in early scenes. *Acta Agricul. Jiangxi* **18**(6), 141–143 (2007)
5. Yong, W., Chang-ying, J., Ming-xia, S.: Study on the recognition of mature cotton based on the chromatic aberration in natural outdoor scenes. *Acta Agricul. Zhejiang* **19**, 385–388 (2007)
6. Basu, M.: Gaussian-based edge-detection methods—a survey. *IEEE Trans. Syst., Man, Cybern. C* **32**, 252–260 (2002)
7. Marr, D., Hildreth, E.: Theory of edge detection. *Proc. R. Soc. A-Math. Phys. Sci. B* **207**, 187–217 (1980)
8. Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* **1986**(8), 679–698 (1986)
9. Witkin, A.P.: Scale space filtering. In: Proceeding 8th International Joint Conference Artificial Intelligence, pp. 1019–1022. Karlsruhe, Germany (1983)
10. Nicole R.: Study on the matlab segmentation image segmentation. In: J. Name Stand. Abbrev
11. Kwon, D. et al.: A image segmentation method based on improved watershed algorithm and region merging. *IEEE Trans. Circuits Syst. Video Technol.* **17**, 517–529 (2007)
12. Ross, J.: *Fuzzy Logic with Engineering Applications*, 3rd edn
13. Krinidis, S., Chatzis, V.: A robust fuzzy local information C-means clustering algorithm. *IEEE Trans. Image Process.* **19**(5), 1328–1337 (2010)
14. Brandt, M.E. : An error convergence simulation study of hard vs. fuzzy c-means clustering. In: *IEEE World Congress on Computational Intelligence*, vol. 3, pp. 1835–1839 (1994)
15. Gonzalez, R.C., Woods, R.E., Eddins, S.L.: *Digital Image Processing Using Matlab* (2008)
16. Bezdek, J.: *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum, New York (1981)
17. Paudel, D.R., Eric, F., Abidi, H.N.: Evaluation of cotton fiber maturity measurements. *Ind. Crops Prod.* **45**, 435–441 (2013)
18. Balafar, M.A.: Fuzzy C-mean based brain MRI segmentation algorithms. *Artif. Intell. Rev.* **41**(3), 441–449 (2014)

System Dynamics Modeling for Analyzing Recovery Rate of Diabetic Patients by Mapping Sugar Content in Ice Cream and Sugar Intake for the Day

Suhas Machhindra Gaikwad, Rahul Raghvendra Joshi
and Preeti Mulay

Abstract Ice cream is a complex colloidal system, usually formed by structural compounds viz., air bubbles, fat, and ice crystals which are dispersed in a matrix consisting of a solution of sugars, proteins, stabilizers, emulsifiers, dyes, and scents. This paper presents system dynamics model based on the Vensim tool and this model contains equations derived for diabetic and non-diabetic patients and common health conscious people as well. Very effectual results have been obtained from the derived model suggesting best suitable ice cream for mapping diabetes patients, which is different from health conscious people and normal crowd.

Keywords System dynamics · Vensim simulation · Ice cream · Diabetic patient

1 Introduction

Ice cream is a summer time favorite and also a big time treats from all age. Ice cream contains different attributes like sugar, fats, carbohydrates, and proteins. Sometimes ice cream contains wine flavors. In today's smart IT-enabled world, even hoteliers need to adapt to automatic menu suggestions system. Now-a-days technology is used for making decisions about right food choice of suggesting, e.g., prediction of food demand in food courts by using decision tree approaches, nutritional needs recommendation based on fuzzy logic, etc. [1, 2]. To be specific, in this paper, the

S.M. Gaikwad (✉) · R.R. Joshi · P. Mulay
Symbiosis International University, Lavale, Pune 412115, India
e-mail: Suhas.gaikwad@sitpune.edu.in

R.R. Joshi
e-mail: rahulj@sitpune.edu.in

P. Mulay
e-mail: preeti.mulay@sitpune.edu.in

S.M. Gaikwad · R.R. Joshi · P. Mulay
Symbiosis Institute of Technology, Pune, India

focus is on consumption of an ice cream [3–6] by diabetic patients. Till date there exists no model existing to correctly predict which ice cream suits best to which human being, by considering all health and restriction factors. The research work proposed in this paper is implemented by developing a model using the Vensim tool [7] and equations derived for diabetic, non-diabetic, and others. The input to this model includes sugar contents from different ice cream flavors, medical data for patients (based on incremental learning approach) [2], and data of diabetics.

Vensim simulation software is for improving the performance of real systems. Vensim is used for developing, analyzing, and packaging dynamic feedback models. It emphasizes on:

- High quality, with dimensional consistency and Reality Checks™
- Connections to data and sophisticated calibration methods
- Instant output with continuous simulation in SyntheSim
- Flexible model publication
- Model analysis, including optimization and Monte Carlo simulation

Vensim contains many industry-leading technical advances in simulation technology. Vensim also supports a great incremental learning environment. The research work proposed and under implementation, in this paper is an effectual extension of Modified-Closeness Factor-based Incremental Clustering (M-CFBA) [8] approach. Ice cream data, patient's data along with diabetic's details is clustered separately. These clusters are stored in cluster database. These two different cluster sets are mapped (ice cream cluster db and patients' db). This mapping generates incremental learning model, which is incorporated in Vensim tool. The closeness factor-based incremental clustering algorithm [8, 9] is error-based statistical clustering algorithm specially designed for knowledge augmentation. As and when different patient's diabetic's results are published and as and when new flavors of ice creams are added [10, 11], M-CFBA will prove its effectiveness by incremental mapping cluster [12].

2 Development of System Dynamics Model [13] for Ice Cream and Diabetic Patients

A new model called NSRT is derived. The full form for NSRT model is Non-Diabetic patients Sugar (Diabetic) patients Recovery rate Total people model.

2.1 *Derived Equations for NSRT Model*

Equations derived for patients and ice creams are given in Sects. 2.1.1–2.1.4. The data required for input is taken from Pearson's website and other sources [14].

2.1.1 Equation for Non-Diabetic Patient (N) [14]

$$dN/dt = \beta * \alpha * (N/T) * [(A - t)/C] \quad (1)$$

where,

β Sugar per day or Sugar rate (120 mg/dl)

α Infectivity, i.e., probability of getting infected with sugar in ice cream (0.5)

N Non-diabetic patients (990 unit people)

T Total people (unit people)

A Actual sugar intake for the day (220 mg/dl)

t Target sugar (140 mg/dl)

C Correction factor (50)

2.1.2 Equation for Diabetic Patient (S) [14]

$$dS/dt = \beta * \alpha * (N/T) * [(A - t)/C] - S * i \quad (2)$$

where S = Diabetic Patient (10 unit people), i = Sugar content from ice cream (normally 13 mg/dl) and values, meanings for rest of terms in equation are same as that of the Eq. (1).

2.1.3 Equation for Recoveries of People (R) [14]

$$dR/dt = S * \theta \quad (3)$$

where θ = Recovery rate (1/365 unit day). Here, also values and meanings for rest of the terms in equation are same as that of the Eq. (2).

2.1.4 Equation for Total People (T) [14]

$$T(t) = N(t) + S(t) + R(t) \quad (4)$$

where,

$T(t)$ Total people

$N(t)$ Non-diabetic people

$S(t)$ Diabetic patient

$R(t)$ Recoveries of people

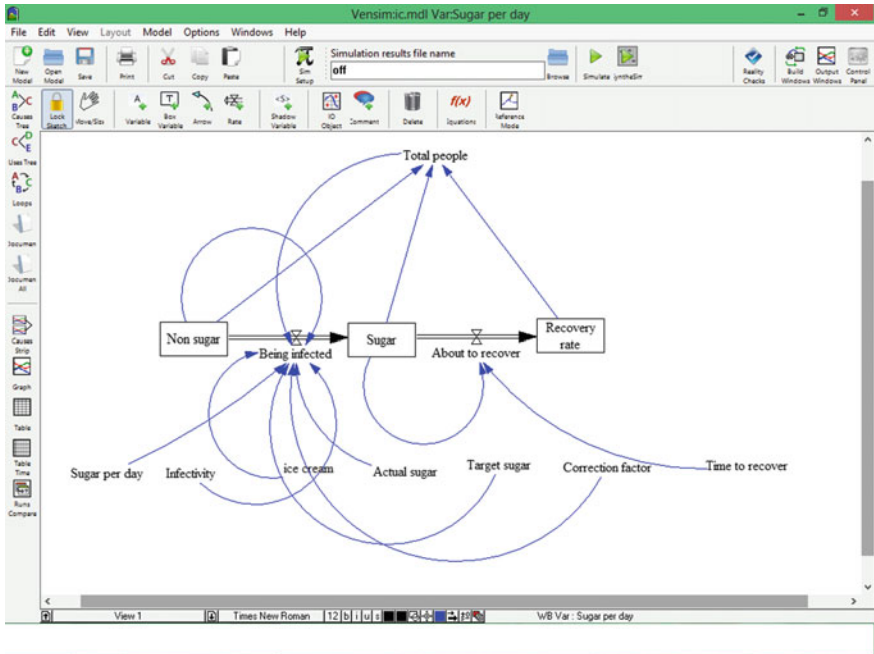


Fig. 1 System dynamics model for NSRT

3 System Dynamics Model for NSRT

The system dynamics model [15–17] shown in Fig. 1 considers Eqs. (1)–(4) to predict the results shown by Vensim simulation tool. Here, three variables viz., N = non-diabetic people, S = diabetic patient, and R = recoveries of people are considered.

4 Vensim Simulation Results

The results of NSRT simulation model are shown by graphs in Figs. 2 and 3.

Figure 2 shows that for the value of i as 13 mg/dl and sugar intake for the day as 220 mg/dl for the patient whose age is 71 years, blue curve shows that non-diabetic people will take 3 days, green curve shows that diabetic people will take 35 days to match with blood sugar level before consuming ice cream with sugar content as 13 mg/dl and red curve shows that recovery in this case will become steady after 45 days.

The graph in Fig. 3 shows that if change value of i from 13 to 15 mg/dl and sugar intake for the day from 220 to 200 mg/dl for patient having age 71 and 60,

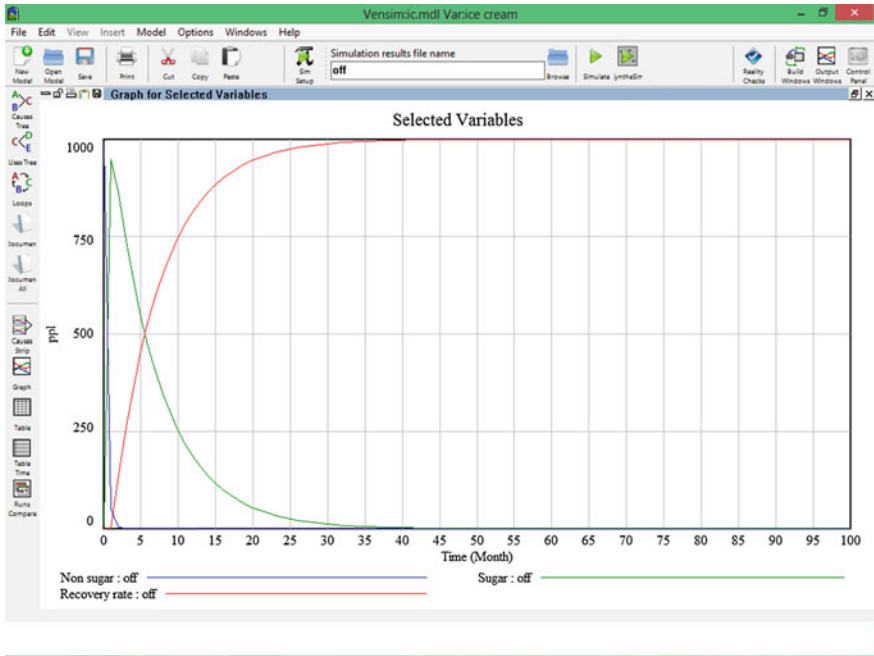


Fig. 2 Results of NSRT model by considering value of i as 13 mg/dl

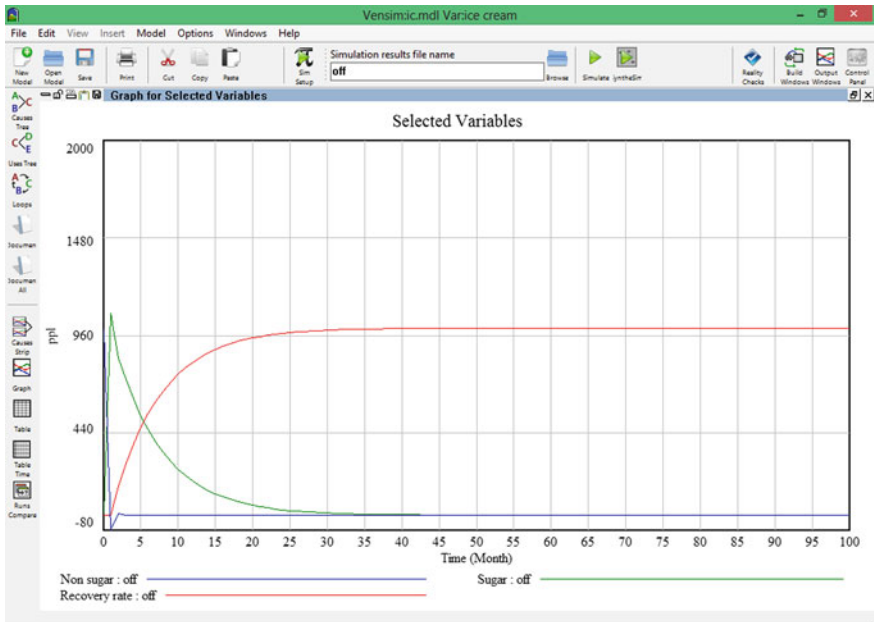


Fig. 3 Results of NSRT model by considering value of i as 15 mg/dl

respectively [12] the blue curve shows that non-diabetic patients will take 2 days, green curve shows that diabetic patients will take 30 days to match with blood sugar level before consuming ice cream in with sugar content as 13 mg/dl, and red curve shows that recovery in this case will become steady after 40 days.

5 Conclusion

This paper mainly focuses on use of system dynamics model and its results by considering equations derived for sugar content in ice cream, diabetic patients, and non-diabetic patients. It is clear from graphs in Figs. 2 and 3, if we vary the value of i , i.e., sugar content in ice cream and sugar intake per day for patient under consideration then recovery will fluctuate after certain days of month. This technique of developing system dynamics simulation model is useful to ice cream industries, for diagnosis of diabetic patients, every individual, hotelier, etc., and to name a few. As mentioned in the introduction of this paper, till date no such simulation model for predicting a suitable ice cream to a diabetic patient is available so through this research attempt is made to do so. The correct prediction about which ice cream suits best to which human being is done by considering two different types of ice cream and its simulation result gives days required to recover based on the sugar level of diabetic and non-diabetic people, all health and people with a restriction factor.

6 Future Extensions

The results given by system dynamics model can be compared further with ice crème and patients' clusters set in order to achieve better validations and verifications before predicting the suitability of an ice cream to patients. Also ice cream, wine, and food and wine and desert paired can be worked out in proposed system dynamics model.

References

1. Bozkir, A.S., Sezer, E.A.: Predicting food demand in food courts by decision tree approaches. *Procedia Comput. Sci.* **3**, 759–763 (2011)
2. Priyono, R.A., Surendro, K.: Nutritional needs recommendation based on fuzzy logic. *Procedia Technol.* **11**, 1244–1251 (2013)
3. Salter, R.M.: Nova: a modern platform for system dynamics, spatial, and agent-based modeling. *Procedia Comput. Sci.* **18**, 1784–1793 (2013)
4. Woo, T.H.: Dynamical modeling investigation for economy of nuclear power plants (NPPs) in global nuclear energy market. *Int. J. Electr. Power Energy Syst.* **43**, 369–374 (2012)

5. Xi, X., Poh, K.L.: Using system dynamics for sustainable water resources management in Singapore. *Procedia Comput. Sci.* **16**, 157–166 (2013)
6. vensim.pdf
7. Varela, P., Pintor, A., Fiszman, S.: How hydrocolloids affect the temporal oral perception of ice cream. *Food Hydrocoll.* **36**, 220–228 (2014)
8. Wei, S., Yang, H., Song, J., Abbaspour, K.C., Xu, Z.: System dynamics simulation model for assessing socio-economic impacts of different levels of environmental flow allocation in the Weihe River Basin, China. *Eur. J. Oper. Res.* **221**, 248–262 (2012)
9. Al-Nazer, A., Helmy, T., Al-Mulhem, M.: User's profile ontology-based semantic framework for personalized food and nutrition recommendation. *Procedia Comput. Sci.* **32**, 101–108 (2014)
10. Mulay, P., Dr. Kulkarni, P.A.: Knowledge augmentation via incremental clustering, new technology for effective knowledge management. *ACM digital library. Int. J. Business Inform. Syst.* **12**(1), 68–87 (2013)
11. Machhindra Gaikwad, S., Dr. Mulay, P., Raghvendra Joshi, R.: Analytical hierarchy process to recommend an ice cream to a diabetic patient based on sugar content in it. *Procedia Elsevier* (2015)
12. Joshi, P.M., Dr. Kulkarni, P.A.: A novel approach for clustering based on pattern analysis. *Int. J. Comput. Appl.* **25**(4), 0975–8887 (2011)
13. Fernández, C., Espinós, F.J., López, M.C., García-Diego, F.J., Cervera, C.: Representation of a mathematical model to predict methane output in dairy goats. *Comput. Electron. Agric.* **91**, 1–9 (2013)
14. Hussin Alsonosi Omar, A., Abu Hasan, Y.: Numerical simulations of an SIR epidemic model with random initial states. *ScienceAsia* **39S**, 42 (2013)
15. Yuan, X., Shen, L., Ashayeri, J.: Dynamic simulation assessment of collaboration strategies to manage demand gap in high-tech product diffusion. *Robot. Comput.-Integr. Manufact.* **26**, 647–657 (2010)
16. Choi, K., Bae, D.-H.: Dynamic project performance estimation by combining static estimation models with system dynamics. *Inf. Softw. Technol.* **51**, 162–172 (2009)
17. Sahin, O., Siems, R.S., Stewart, R.A., Porter, M.G.: Paradigm shift to enhanced water supply planning through augmented grids, scarcity pricing and adaptive factory water: a system dynamics approach. *Environmental Modelling & Software* (2014)

Enhanced Stable Period for Two Level and Multilevel Heterogeneous Model for Distant Base Station in Wireless Sensor Network

Pawan Singh Mehra, M.N. Doja and Bashir Alam

Abstract In last decade, Wireless Sensor Network (WSN) has gained popularity because of its little or no infrastructure-based network communication. The key concerned in WSN is energy optimization. Several researches are being carried out focusing on the parameters that affect the energy of the network. Introduction of heterogeneity enhances the network lifetime. In this paper, we propose a heterogeneity-based energy efficient clustering scheme for distant base station. In this protocol, cluster head selection is based on localized parameters of the node. The proposed work is simulated for two level and multilevel heterogeneous energy model. Simulation results validate the extended stability of proposed work. The proposed protocol remarkably outperforms SEP and DEEC.

Keywords Wireless sensor network · Heterogeneous · Clustering · Cluster head · Base station · Energy efficiency

1 Introduction

With the advancement in technology, the size of the hardware device is decreasing day by day leading to development of microsensors. Wireless Sensor Network (WSN) is a composition of large number of micro sensors deployed in an environment to collect useful data. This data is directed to Base station or sink which is located nearby the target area. Base station can act as gateway or master node which

P.S. Mehra (✉) · M.N. Doja · B. Alam

Department of Computer Engineering, Faculty of Engineering & Technology,
Jamia Millia Islamia, Jamia Nagar, New Delhi, India
e-mail: pawansinghmehra@gmail.com

M.N. Doja
e-mail: ndoja@yahoo.com

B. Alam
e-mail: babashiralam@gmail.com

forwards the data to the remote user for further processing. WSN can be thought of as a team work where all sensors contribute to fulfill the requirement of the application but it is not necessary that all the sensor nodes in the network are actively participating at the same time. WSN are deployed in several applications which require monitoring and tracking [1]. Tracking in WSN includes animal tracking in habitat, human tracking in business, enemy tracking in battlefield or car or bus tracking in Traffic. Monitoring in WSN can be done for Environment (e.g., temperature, pressure, humidity, etc.), Inventory monitoring in Warehouse, Patient monitoring in Hospital, Structural Monitoring in Huge buildings, Machinery monitoring in factory. WSN can be deployed deterministically or nondeterministically [2]. In deterministic deployment sensor nodes are placed in the target area with some specific consideration in mind, e.g., structural monitoring or inventory monitoring where as in nondeterministic deployment, sensor nodes are dropped on the fly, i.e., the location of the sensors is unknown, e.g., forest fire detection or volcanic eruption detection. However, location of base station in nondeterministic deployment is at distant place because it may not be possible to place the base station in between the disaster area.

In this paper, we have proposed two sets of distributed protocol: two level model and multilevel model. In both the models, the location of base station is kept at distant place. The selection of the cluster head plays a crucial role for longer lifetime of a network. In the proposed work, the selection is based on localized parameters of each node and the node with best candidature gets selected as cluster head.

Rest of the paper is continued as follows. Relevant literature survey is discussed in Sect. 2. Heterogeneous model is discussed in Sect. 3. Section 4 describes the proposed work. Performance and evaluation of proposed work is shown in Sect. 5. Concluding remarks are discussed in Sect. 6.

2 Literature Review

Heinzelman et al. proposed LEACH [3]. This scheme is designed for homogeneous wireless sensor network. Here homogeneous means all nodes are having same energy level ab initio. In this decentralized scheme, a round comprises of two phases: setup phase and steady phase. In setup phase, cluster heads are selected on the basis of random number generated by each node. If the random number generated is less than the defined threshold $T(n)$, then the node gets selected as cluster head. The threshold is calculated by

$$T(n) = \begin{cases} \frac{P_{opt}}{1 - P_{opt} \left(r \bmod \frac{1}{P_{opt}} \right)} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where r is the round number and p_{opt} is the optimal number of cluster heads in each round. Every node in network gets chance to become cluster head exactly once in $1/p_{opt}$ rounds. This scheme does not consider the remnant energy of the node while selecting the cluster head. Randomization of cluster head selection may at times lead to no cluster head in a round.

In [4], Jindal et al. discusses LEACH and its descent protocols with some of its improvements. AbdelSalam et al. proposed BEES [5], an algorithm which is inspired by bees colony. It has four phases: tiling, backbone, selection, and clustering. This algorithm is capable of overcoming the challenges like clustering, localization, and data aggregation but clustering process is complex.

Smaragdakis et al. introduced SEP [6] which is heterogeneity aware protocol. In this protocol, the author discusses the impact of heterogeneity and instability. SEP is based on LEACH which does not require knowledge of energy level of network. This protocol considers two types of nodes: normal nodes and advanced nodes. The candidature of each node to become cluster head is based on weighted election probability which considers the initial energy level of each node. The weighted probability for normal and advanced node is calculated as follows:

$$P_{nrm} = \frac{P_{opt}}{1 + \alpha m} \tag{2}$$

$$P_{adv} = \frac{p_{opt}(1 + \alpha)}{1 + \alpha m} \tag{3}$$

where P_{opt} is the optimum percentage of cluster head in each round, α is additional energy factor and m is the fraction of advance nodes. Since this protocol provides better chance to advance nodes to become cluster head than normal nodes. But this protocol does not consider the residual energy of the node. This protocol does not consider the distance between the cluster head and the base station as the communication is single hop.

Qing et al. proposed DEEC [7] which is based on SEP in weighted probability. This protocol considers two levels of energy and extends the protocol for multilevel energy of sensor nodes deployed in field. The election of cluster head is based on the ratio of residual energy to the average energy of the network. The weighted probability is calculated by following

$$P_i = \begin{cases} \frac{P_{opt}E_i(r)}{(1+\alpha m)\bar{E}(r)} & \text{for normal node} \\ \frac{p_{opt}(1+\alpha)E_i(r)}{(1+\alpha m)\bar{E}(r)} & \text{for advance node} \end{cases} \tag{4}$$

where r is the current round, $E_i(r)$ is the residual energy of the node and $\bar{E}(r)$ is the average energy of the network for current round. This protocol also extends the

model for multilevel energy where the level of energy of the nodes lie between the close set $[E_0, E_0(1 + \alpha)]$. The cluster heads are rotated to balance the load of the network. This protocol is also random based so there may be a round in which no cluster head is chosen. This protocol does not consider the local parameters of the node and is unfit for application where the location of base station is at distant place from the target area.

Wang et al. proposed DECP [8] for two levels heterogeneity. In this the selection of cluster head is based on the communication cost and remnant energy of the node. The author uses Average Power Distinction (APD) to calculate the power level of each sensor nodes. All the nodes in the network broadcast their residual energy level. All the nodes calculate the neighbor distance, current energy, communication cost, and broadcast it to the neighbors. Node with minimum cost is elected as cluster head and noncluster head nodes join the nearest cluster head. This protocol exhibits better stability period.

Varma et al. proposed base station initiated dynamic routing protocol based on clustering [9]. In this protocol, heterogeneity in terms of energy, computational capability and location awareness is considered for some fractions of nodes. These nodes act as cluster head. The nodes are deployed deterministically in the field. Only one hop transmission provision is made for cluster head. The clusters are divided into levels. Cluster head nearer to the base station has lower level where as cluster head farther from the base station has higher level. Data is forwarded from higher level to lower level. This protocol succeeds in extending the lifetime of the network but centralized approach limits the application of the proposed work.

Kumar et al. proposed EEHC [10] for three levels of node: normal nodes, advance nodes, and super nodes. It is based on LEACH for threshold for randomization. EEHC considers the initial energy of the node for cluster head candidature. Selection of cluster head is based on weighted probability for each level of nodes. This approach guarantees rotation of cluster head role among the nodes but does not consider the residual energy of the node. There are chances of no cluster head for a round because of randomization.

Chand et al. proposed heterogeneous HEED [11] for three levels of sensor nodes. In this protocol, homogeneous model of HEED [12] is extended. Fuzzy logic-based system is used to select the cluster heads among the nodes in the network. Parameters like residual energy, node density and distance are put as input to the inference engine to select the best available node for cluster head selection. This protocol significantly improves the life time of the network.

3 Heterogeneous Network Model

In this section, network model for the proposed work is discussed. We assume that N numbers of nodes are randomly deployed in $M \times M$ network field. p_{opt} is chosen as 10 % of alive nodes in the network. The nodes are location unaware. Capabilities

of all the nodes in the network are similar except the energy level. The nodes in the network always have data to be forwarded to base station which is located far away. Application of such network can be seen for remote monitoring like military surveillance, forest fire detection, disaster area, etc. The network is assumed to have fixed topology with clustering hierarchy.

Heterogeneity in sensor nodes of wireless sensor network are of three types: computational, energy and link [11]. Computational heterogeneity is described in terms of more processing power for complex data and extended memory, energy heterogeneity means the node has high power source and link heterogeneity is meant for long and reliable transmission with high bandwidth. Link and computational heterogeneity depend upon power source, thus, it is better to exploit energy heterogeneity in sensor nodes. The amount of energy consumption in routing data to sink in heterogeneous model is less as compared to homogeneous model and the links are more reliable [10].

In this proposed work, we have two models in terms of energy; two level model and multilevel model. In two level model, some fraction (i.e., m) of nodes are bundled with extraneous energy and are termed as advance nodes. Rests of the nodes are normal nodes. E_o is the initial energy for normal nodes and advance nodes have α times more energy than normal nodes. Therefore, we have mN advance nodes in the network with energy level $E_o(1 + \alpha)$. Total energy of the network for two levels can be calculated by:

$$E_{total} = N(1 - m)E_o + NmE_o(1 + \alpha) = NE_o(1 + \alpha m) \tag{5}$$

For multilevel model of the proposed work, the initial energy of the nodes is closed under the set $[E_o, E_o(1 + \alpha)]$. Total energy of the multilevel model can be calculated by:

$$E_{total} = \sum_i^N E_o(1 + \alpha_i) \tag{6}$$

where α_i lies between E_o to α_{max} .

In the energy expenditure analysis, we assume the same energy model used in [7]. To transmit l bit message to a distance d , energy dissipated by radio is given by:

$$E_{Tx}(l, d) = \begin{cases} lE_{elec} + l\epsilon_{fs}d^2, & d < d_o \\ lE_{elec} + l\epsilon_{mp}d^4, & d \geq d_o \end{cases} \tag{7}$$

where E_{elec} is the energy required per bit to run the transceiver circuit and $\epsilon_{fs}d^2$ and $\epsilon_{mp}d^4$ depends upon transmitter amplifier model.

The energy expended by a cluster head in a round is calculated as

$$E_{CH} = n(lE_{elec} + l\epsilon_{fs}d^2 + E_{DA}) \quad (8)$$

where n is the number of its cluster members and E_{DA} is energy dissipated in data aggregation.

4 Proposed Work

The proposed work is divided into two phases: Clustering phase and Data dissemination phase. Base station broadcast a hello message to the network so that the sensor nodes estimate the location of the base station through RSSI or LQI which is commonly implemented by IEEE 802.11 standard. To setup the clusters, the nodes broadcast their energy level into the network. Every node in the network calculates its local parameter of the surrounding ,i.e., average distance to neighbor nodes (T_i), its current residual energy level ($E_i(r)$), energy exhausted in previous round (δ_i), total energy dissipated till current round ($\omega_i(r)$), number of times it has been chosen (C) and not chosen as cluster head (C_o) and its distance to base station (D_{BS}) and determine its weightage for cluster head candidature by the formula given below:

$$W_i = \frac{E_i(r) \times C_o \times T_i}{C \times \omega_i(r) \times \delta_i \times D_{BS}} \quad (9)$$

Each node broadcast its Weightage (W_i) and the node with highest weightage gets elected as cluster head. Cluster head sends a join message along with the TDMA schedule to all the nodes within cluster communication range (R_c). The communication range for cluster radius in the proposed work is calculated by:

$$R_c = \sqrt{\frac{M^2}{P_{opt} \times \Pi}} \quad (10)$$

The nodes join the cluster by sending acknowledgement to the cluster head. Likewise next node from leftover nodes in the network with highest weightage broadcast its candidature and forms clusters. After completion of clustering phase, dissemination phase comes into play. All the sensor nodes collect data from the surrounding and transmit it to their respective cluster heads. Cluster heads after aggregating the data directly transmit it to the base station which is located at a distant place as per the TDMA schedule so that no collision takes place and thus completes one round of the proposed protocol.

5 Performance Evaluation

We have simulated SEP and DEEC along with the proposed work for different heterogeneity parameters. Simulation of the proposed work is carried out for two level and multi level. SEP is extended for multilevel to obtain the results. Several hundred iterations are done to obtain the normalized value of results. Performance evaluation is done on the basis of stability period as discussed in [6] and number of successful packet transmission from node to base station. A packet is meant to be successfully transferred if it completes its journey from creation at a node to the base station. Longer stability period justifies the better energy dissipation rate of the protocol. We have considered first-order radio model as used in [7] and the parameters are shown in Table 1.

In Fig. 1, 10 % nodes are chosen as advance nodes for two level model. Proposed protocol achieves 47 and 102 % longer stability period than SEP and DEEC for two level model and for multilevel model, proposed work attains 117 and 35 % longer stability period than SEP and DEEC.

In Fig. 2, fraction of advance nodes is 30 % of the total number of nodes deployed in the field. Proposed work overcome the stability period of SEP and DEEC with 41 and 268 %, respectively for two level model and for multilevel model, the span of stability period for proposed protocol is 137 and 402 % more as compared to SEP and DEEC respectively.

Table 1 Parameters of radio model used in simulation

Parameter description	Symbol	Value
Energy dissipated by amplifier to transmit at a short distance	ϵ_{fs}	10 pJ/bit/m ²
Energy dissipated by amplifier to transmit at a long distance	ϵ_{mp}	0.0013 pJ/bit/m ⁴
Energy consumed in electronic circuit to transmit or receive the signal	E_{elec}	50 nJ/bit
Data aggregation energy	E_{DA}	5 nJ/bit/report

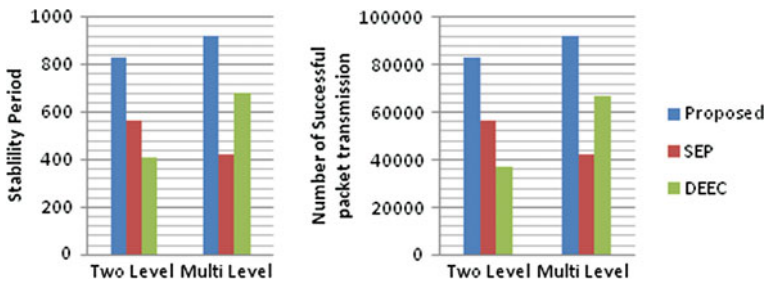


Fig. 1 Stability period and successful packet transmission for SEP, proposed work and DEEC when $E_o = 0.5$ J, $\alpha = 0.5$, $m = 0.1$, $N = 100$, field size (100, 100), and location of base station (175, 50)

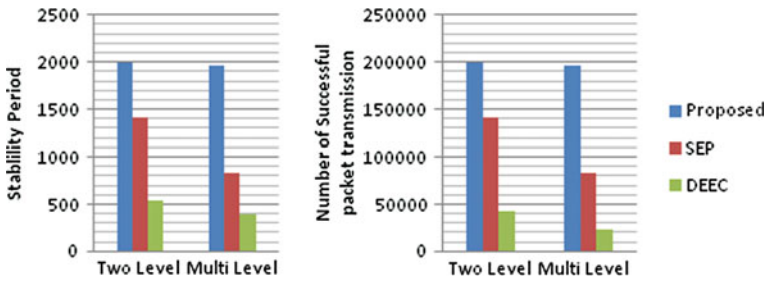


Fig. 2 Stability period and successful packet transmission for SEP, proposed work and DEEC when $E_0 = 1$ J, $\alpha = 1$, $m = 0.3$, $N = 100$, field size (100, 100), and location of base station (175, 50)

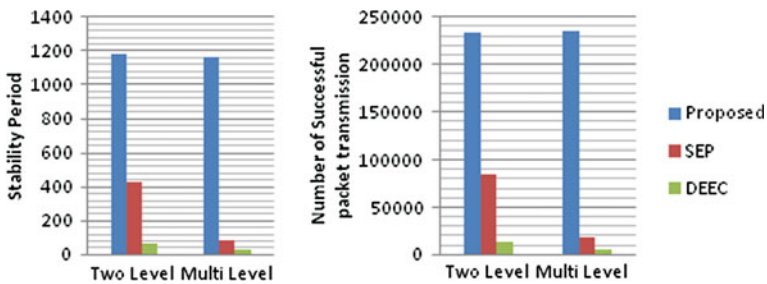


Fig. 3 Stability period and successful packet transmission for SEP, proposed work and DEEC when $E_0 = 2$ J, $\alpha = 2$, $m = 0.5$, $N = 200$, field size (200, 200), and location of base station (350, 50)

Figure 3 exhibits noteworthy performance of the proposed protocol. The size of the field is increased and the number of nodes deployed in the field is set to (200 × 200), 200 respectively. Fraction of advance nodes is set as $m = -0.5$ and initial energy of the nodes is kept to 2 J. For two level model, the proposed work achieves 176 and 1607 % better stability period as compared to SEP and DEEC, respectively, where as for multilevel model, proposed work remarkably gains 1208 and 3783 % more stable period over SEP and DEEC, respectively.

6 Conclusions

We describe the proposed work for heterogeneous wireless sensor network where base station is far away from base station. In this protocol, best candidate for cluster head is chosen keeping in mind the parameters that affect the energy of node to prolong the network lifetime. Proposed protocol performs remarkably well in two level as well as multilevel model.

References

1. Mehdi Afsar, M., Mohammad, H., Tayarani, N.: Clustering in sensor networks: a literature survey. *J. Network Comput. Appl.* **46**, 198–226 (2014)
2. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. *Computer Networks* **52** (12), 2292–2330 (2008)
3. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: *Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00)* (2000)
4. Jindal, P., Gupta, V.: Study of energy efficient routing protocols of wireless sensor network and their further researches: a survey. *J. Comput. Sci. Commun. Eng.* (2013)
5. AbdelSalam, H.S., Olariu, S.: Bees: bio inspired backbone selection in wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* (2012)
6. Smaragdakis, G., Matta, I., Bestavros, A.: SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. In: *Second International Workshop on Sensor and Actor Network Protocols and Applications* (2004)
7. Qing, L., Qingxin, Z., Mingwen, W.: Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Computer Commun.* **29**(12), 2230–2237 (2006)
8. Wang, X., Zhang, G.: DECP: A distributed election clustering protocol for heterogeneous wireless sensor networks, ser. *Lecture Notes in Computer Science*, vol. 4489. Berlin, Germany: Springer Berlin Heidelberg (2007)
9. Varma, S., Nigam, N., Tiwary, U.S.: Base station initiated dynamic routing protocol for heterogeneous wireless sensor network using clustering. In: *Fourth International Conference on Wireless Communication and Sensor Networks 2008 WCSN 2008*, pp. 1, 6, 27–29 (2008)
10. Kumar, D., Aseri, T.C., Patel, R.B.: EEHC: energy efficient heterogeneous clustered scheme for wireless sensor networks. *Comput. Commun.* **32**(4), 662–667 2009
11. Chand, S., Singh, S., Kumar, B.: Heterogeneous HEED Protocol for wireless sensor network. *Wirel. Personal Commun.* **77**(3), 2117–2139 (2014)
12. Younis, O., Fahm, Y.S.: HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. *IEEE Trans. Mobile Comput.* **3**(4), 366–379 (2004)

Dynamic Timetable Generation Using Constraint Satisfaction Algorithm

Urmila Kalshetti, Deepika Nahar, Ketan Deshpande,
Sanket Gawas and Sujay Sudeep

Abstract Manual method of generating timetable has always been a time-consuming, laborious, and tedious task. It is neither efficient nor effective in terms of utilization of resources. The complicated relationships between time periods, classes (lectures), classrooms, and instructors (staff) make it difficult to attain a feasible solution. In this paper, timetabling problem is modeled as a constraint satisfaction problem. The algorithm dynamically builds the timetable adjusting resources in order of complexity. The main focus is to satisfy all the hard constraints and maximum soft constraints without any conflicts among resources. In order to reach a subsolution state, we use various heuristics that guide the search. Along with this, chronological backtracking and look-ahead techniques are also discussed. This software is ergonomic in nature as it also provides a way to alter the given inputs.

Keywords Timetabling problem · Constraint satisfaction · Hard constraints · Soft constraints · Chronological backtracking · Look-ahead

U. Kalshetti (✉) · D. Nahar · K. Deshpande · S. Gawas · S. Sudeep
Department of Computer and Information Technology, Pune Vidyarthi Griha's COET,
Pune 411009, India
e-mail: urmila.kalshetti@gmail.com

D. Nahar
e-mail: deepika.nahar@gmail.com

K. Deshpande
e-mail: ketanrd.009@gmail.com

S. Gawas
e-mail: sanketgawas1993@gmail.com

S. Sudeep
e-mail: sujaysudeep93@gmail.com

1 Introduction

Timetabling concerns all activities with regard to making a timetable that must be subjective to different constraints. Timetable generation in educational institutes is a problem that has a number of constraints, which, if not properly managed, cause many resource clashes, and results in poor management. Manual approach to generate timetable is tedious, time-consuming process, and sometimes error prone as well. An educational timetable must meet a number of requirements and should satisfy the desires of all entities involved simultaneously. Timetabling problem is frequently measured as a resource provision problem in operations research, where resources like classrooms, lectures, staff, labs, and subjects are to be consigned into timeslots of a timetable to accomplish an objective function focusing on constraints among resources. It is a thought-provoking task to find a near perfect or ideal solution when the number of resources reduces and constraints increases. The process of creating a timetable varies from one institute to other as the number of resources, constraints, and requirements change invariably [1]. All the problems related to building a timetable are known to be NP-complete [2]. This means that, as the size of the problem grows, so does the time required to arrive at the conclusive solution.

Considering all this, where the complexity of the problem and restricted time available does not permit precise solution, it is necessary to aim at optimized heuristic methods to obtain good feasible solution.

2 Related Work

The timetabling problem basically deals with a broad set of constraints. These constraints are categorized as hard and soft constraints, where the hard constraints are those, which cannot be violated and the latter are not compulsorily required to be satisfied. Because of these constraints, the problem becomes intricate and hence, a heuristic approach needs to be suitably preferred. A number of different heuristic approaches and theories have been proposed since 1960s' to find the most favorable and approximate solution to the timetabling problem. This indicates that there are many snags in timetabling that need to be solved in view of the availability of more powerful computing facilities and advancement of information technology (Deris et al. 1997). Gotlieb (1963) developed the first non-heuristic approach to solve the timetabling problem [3]. This being a traditional approach, suffered a lot of setbacks with respect to discrete search space and multiple objective functions. There onwards a number of approaches were developed. However, each approach had its own drawbacks. As a result of this insufficiency provided by classical methods, a number of researchers tried their hands on heuristic-based nonclassical techniques [2]. The various nonclassical approaches to solve the time tabling problem are:

1. Linear/Integer Programming
2. Simulated Annealing

3. Genetic Algorithm
4. Constraint Satisfaction Programming

The Linear/Integer programming technique is associated with mathematical programming. This technique deals with variables, boundaries and prespecified conditions [4, 5]. The variables are restricted within the boundaries, which necessarily have to satisfy the given conditions. This technique performs inefficiently from the computational point of view increasing the size and complexity of the problem. Simulated Annealing (SA) involves local search optimization technique for finding solution to optimization problem. This technique is highly resource demanding thus increasing the computational time. To improve the performance, shared memory multiprocessors systems and distributed memory multiprocessors based on parallelization are used [6, 7]. Genetic Algorithms (GA) are based on the theory of natural selection. The principle behind their working operation is that a set of random solutions is generated. Out of these, only the fittest solutions survive after the randomly generated solutions are competed with each other. Each solution in the set is more or less similar to chromosome which ultimately forms a population [8]. The algorithm then uses three basic genetic operators: reproduction, crossover, and mutation. With the help of these operators and fitness function the GA searches for the optimal solution based on guided random search. Disadvantages include no stopping criteria, resource intensive, and time consuming. Constraint Satisfaction Programming/Algorithm (CSP) is a search procedure that operates in space of constraint sets [8]. It consists of two states, the initial state and the goal state. The initial state contains constraints that are originally given, whereas the goal state are any state that has been constrained enough where 'enough' must be defined for each problem. It is a two-step process: 1. Constraints are determined and broadcasted as far as possible throughout the system and 2. Then, if still the solution is not found the search begins. A guess about approximate value is made and added as a new constraint. Propagation then occurs with this new constraint [8]. CSP technique, if used, reduces the search space thereby reducing the time required to search for solution. This minimizes backtracking. Furthermore, memory required is also smaller as the search space is reduced.

Apparently, no solitary system will prove as a best and precise solution to solve the timetabling problem. However, as indicated there seems to be an inclination toward the usage of Constraint Satisfaction Programming for the same. A lot of work has been done in timetabling problem domain using CSP. The two tools, ILOG Solver and ILOG Scheduler, were used by Lixi Zhang and SimKim Lau to integrate with CSP for solving the timetabling problem, this is referenced in [9]. But, their program was unable to write data on data file. The use of graphical interface for interaction and subject database connection, for maintaining a backend data was absent. Also, no preferences (sequence in which the constraints should be executed) about the constraints were mentioned. Since the algorithms used in the ILOG Scheduler were bound consistent, it was not easy to propagate changes over

bounds of domain variables, a reference of this is given in [10]. In a thesis report presented, Thomas Muller has restricted the use of CSP technique to solve only those problems, which define hard constraints [11].

As we have relatively done a research about the systems implemented so far, we can say that CSP has been most prominently used in generating near optimal solutions to solve timetabling problem. Our motivation behind choosing this technique is that, it possibly generates an automated and a clash-free solution for timetable thereby expanding the scope to making it generic. Hence, we use it as the fundamental principle for our proposed methodology.

3 Timetable Generation as Constraint Satisfaction Problem

The core task in constraint programming is to allocate activities in exact time and slots, with respect to the various constraints so as to satisfy as nearly as possible a set of desirable objectives [12]. Technically speaking, a constraint is nothing but a demand that a certain system needs to fulfill to generate best results [11]. Constraint satisfaction problem (CSP) deals with assignment of values from its domains to each variable such that no constraint is violated. CSP has three components: variables, values, and constraints. In general, CSP consists of: a finite set of variable $X = \{x_1, \dots, x_n\}$ with respective domains $D = \{D_1, \dots, D_n\}$ which list the possible values for each variable $D_i = \{v_i, \dots, v_k\}$ and a set of constraints $C = \{C_1, \dots, C_t\}$. The constraints limit the possible values that a variable can have. Constraints can be split into two parts:

1. Hard constraints
2. Soft constraints

Hard constraints are those that cannot be avoided. For example, no two lectures in the same classroom at the same time, no two lectures for a staff at the same time, etc. Soft constraints are those, which are not compulsorily satisfied but can be minimized. For example, no gaps between two lectures, not more than three lectures per day for a particular staff.

Having defined timetabling as a constraint satisfaction problem we can now put forward an example for the same. Consider, the set X as finite set of variables containing classrooms, staff, and subjects. $X = \{\text{classroom } (X_1), \text{ subject } (X_2), \text{ staff } (X_3)\}$. Likewise, domain D consists of set of respective domain values for each of the variables defined. $D = \{\text{classroom } (D_1), \text{ subject } (D_2), \text{ staff } (D_3)\}$. Now, if, X_1 relates to *classroom* (D_1), X_2 relates to *subjects* (D_2), X_3 relates to *staff* (D_3) then, D_1 relates to the set of values for that classroom, D_2 relates to the set of values for that subject, D_3 relates to the set of values for that staff. $\text{Classroom } (D_1) = \{001,$

002, 003}. $Subject (D_2) = \{DS, SA, BI\}$. $Staff (D_3) = \{John, Alfred, Thomas\}$. This mapping elucidates the basics of timetabling as constraint satisfaction problem.

4 Mathematical Model for Timetabling Problem

In order to compute the complexity of the timetabling problem, a mathematical model for the same is proposed. It includes the following:

I set of staff

J set of courses

K set of subjects

L set of time periods available

M set of days available

L_i Load of staff i who belongs to I

L_k Load of subject k which belongs to K

K_i set of subjects that could be taught by staff i belonging to I

C_l Number of classrooms available during time period l belonging to L

4.1 Constraints

1. $X_{ikl} = 1$ if staff i teaches subject k at time period l
 = 0 otherwise

2. Each staff can only teach at most one subject in a particular time period.

$$\sum_k X_{ikl} \leq 1$$

3. Number of class rooms available during time period,

$$\sum_i X_{ikl} \leq C_l$$

4. For every time period subject should be allocated ,i.e., no gaps allowed,

$$X_{ikl} = 1$$

5. Each subject has to be allocated to a particular time period

$$\sum_i \sum_k X_{ikl} = 1$$

6. Load of staff per week

$$\sum_m X_{im} \leq L_i$$

7. For a particular course, only one or zero subjects conducted in every time period

$$\sum_k \sum_l X_{jkl} \leq 1$$

8. Subject load

$$\sum_m X_{km} = 1$$

5 Proposed Methodology

In this section, we present a detailed modus operandi, showcasing its unique features of our proposed system. The algorithm is developed considering a sample case of Savitribai Phule Pune University. However, it is currently restricted to only the 3 years i.e. Second Year (SE), Third Year (TE), Final Year (BE) out of the entire 4 year course of Engineering. Our proposed system manages the constraints and allocates resources on the basis of a heuristic approach. This includes the scheduling of resources one by one. The algorithm is based on Constraint Satisfaction using the look-ahead scheme.

5.1 Proposed System Features

- The system produces intermediate timetables, such as timetable for classes, labs, staff, and a master timetable.
- The system disseminates workload among timeslots for a week in an equivalent manner.
- It prioritizes allocation of staff according to their respective timeslots.

5.2 *Look-Ahead Scheme*

The main drawback of backtracking is—*thrashing* [13]. This means that it finds out the inconsistent and partial solutions of a problem repetitively in a loop. As thrashing is NP-complete, it is almost improbable for a complete effective solution to it. However, using the look-ahead scheme can categorically solve this issue. The pruning potential of backtracking algorithm can be dynamically improvised by look-ahead scheme. This scheme is based on moving forward to gather appropriate data. The look-ahead strategy is invoked when the algorithm tries to instantiate a particular value to the subsequent variable. Based on this inference a decision made would perpetually restrict the potential search. So, when the forward constraint propagation is done up to a certain limit then the algorithm can carry out the following steps: (a) Make a decision regarding which variable to initialize next if there is no prearranged order. As a result the most heavily constrained variables are selected which have the least feasible values. (b) Make a decision about the value to be assigned to the next variable. Look-Ahead strategies incur an extra cost considering the computational time required to instantiate the request, but they prove to be more beneficial overall.

5.3 *Proposed Look-Ahead Strategy*

Considering the various approaches and techniques a novel heuristic algorithm, which dynamically creates a timetable satisfying all the hard constraints and maximum soft constraints, has been established.

5.3.1 **Input**

- Details of staff, subjects, practical, labs, start and end time of all classes, number of batches for practical, break time, list of classrooms.

5.3.2 **Variables, Terms, and Abbreviations Used**

- *SE, TE, BE*—Academic years of engineering course.
- *common labs*—Labs shared between two or more practical batches.
- w = Sum of workload of each practical
- n = Number of batches
- s = Sum of common for both years, common lab for each year and alternative labs
- d = Number of days

- $Nclr$ = Number of classrooms
- $Ncls$ = Number of classes
- $lecture\ slots$ = Timeslots where theory lectures can be conducted considering one for each day
- $load$ = Number of practical batches conducted by common staff
- $theory\ load$ = Workload of theory subjects conducted by common staff
- $yr\ preference$ (*Year Preference*) = Let staff ' t_1 ' teach both SE and TE . So if the timing for SE is 8:30–3:30, then $yr\ preference$ is given to that particular staff of SE , which has staff timing from 10:30–5:30
- $st\ preference$ (*Staff Preference*) = If the staff timing and year timing are same then preference is low and if staff time and year time are different then preference is high.
- $common\ staff$ = The staff which tutors more than one academic year (SE , TE , BE) of a particular institute
- $common\ time$ = Relates to that mutual time where all the staff is present.
- $compulsory\ subject$ = Compulsory subject is that subject which has the maximum load for a particular day in accordance to the number of days remaining in that particular week

5.3.3 Look-Ahead Algorithm

1. For every possible combination of practical time (SE , TE , BE)
2. Check for $common\ labs$, If false go to 4
3. Check lab allocation for current combination of practical time,
If false go to 9
4. Check classroom availability, If false go to 9
5. Check workload satisfiability, If false go to 9
6. If workload of every $common\ staff$ is satisfied
7. Call for timetable generation algorithm
8. Else go to 9
9. Try next combination, go to 2

5.3.4 Lab Intersection

1. Get practical load for both years having $common\ lab$
2. Get $common\ lab$ for each year
3. Get alternative lab
4. If $(w \times n)/2 \leq (s \times d)$ return true.
5. Else return false

5.3.5 Classroom Availability

1. For every timeslot
2. If $NCl_r < NCl_s$ having theory lectures, return false
3. Return true

5.3.6 Workload Satisfiability

1. For every *common staff*
2. For every year, initially find out *lecture slots*
3. For every remaining year the current *common staff* is teaching for
4. If same practical time, check if staff takes practical for both years
5. If true, check *load* satisfiability
6. If false, return false
7. If false, do nothing
8. Else false, If Practical time for this year is in *common time*
9. Check if *common time* for current & remaining year is same
10. If true, check if staff conducts practical for remaining year
11. If true, consider load & subtract from *lecture-slot*
12. Check *workload satisfiability* (theory load) for current year of *common staff*
13. If *theory load > lecture slot*, return false

5.3.7 Timetable Generation Algorithm

1. Decide *yr-preference* for every year (*SE, TE, BE*)
2. Decide the *st-preference* for every single staff present for respective year.
3. Allocate practical batches primarily
4. Allocate *compulsory subjects*.
5. Allocate remaining subjects slot-wise for every year (*SE, TE, BE*).
6. Check for soft constraints.
7. Finally, check *compulsory subjects* for every year

6 System Block Diagram

See Figs. 1 and 2.

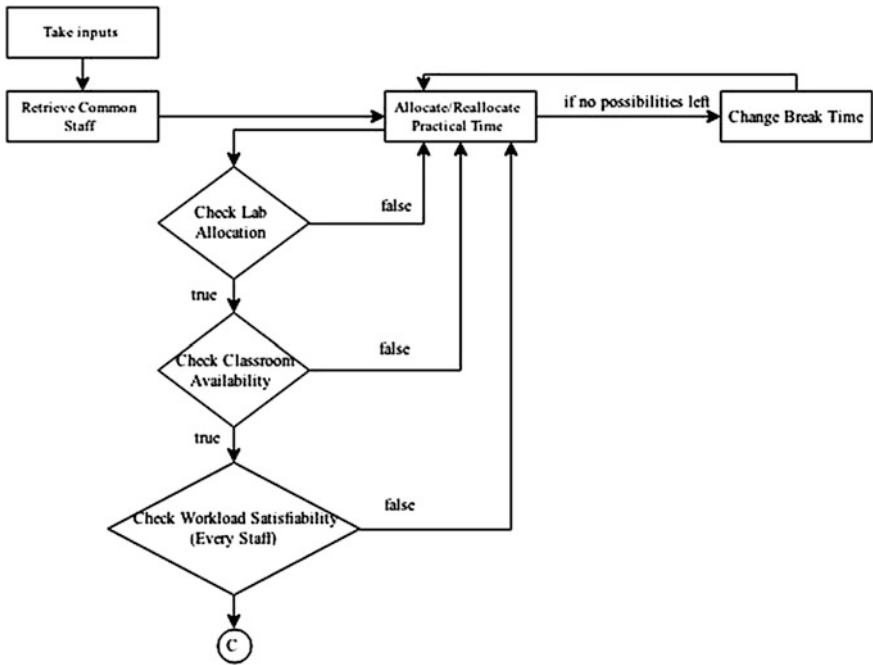


Fig. 1 Phase 1-The look-ahead procedure

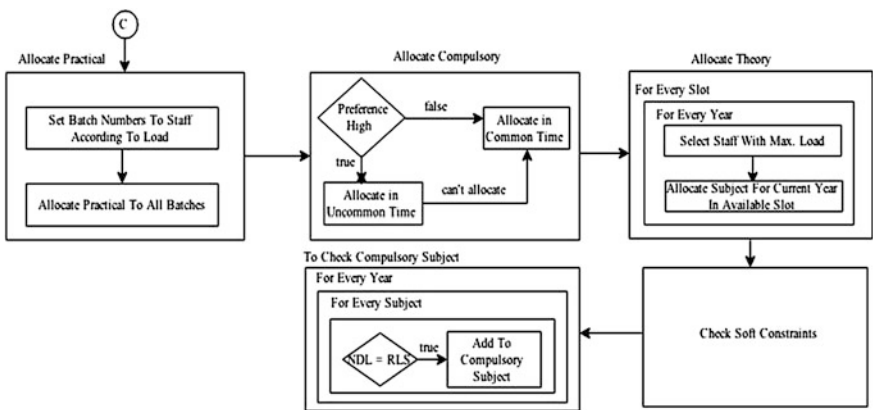


Fig. 2 Phase 2-The timetable generation algorithm (NDL No of days left and RLS Remaining load of subject)

7 Conclusion

We have proposed an algorithm, which allows us to find optimal solution to the problem. Consequently, we are on the verge of implementing a system, to solve the timetabling problem of the entire university. Our proposed work includes, developing a timetable schedule automatically on the basis of given input with minimum human intervention. The algorithm is completely developed by us by integrating the look-ahead technique with Constraint Satisfaction Algorithm. The hard constraint is knowingly tackled in an effective manner so that there is no clash among any resources. We have also tried to satisfy maximum of the soft constraints to the best possible level. Thus, as a result of creating an automation of the timetable problem, many man-hours will be reduced eventually, which otherwise are comparatively very high. Distinct timetables for the individual classes, faculties, and labs would also be generated automatically by this system. Further, the system will generate the master timetable for any department of the college.

References

1. Lien-Fu, L., Nien-Lin, H., Liang-Tsung, H., Tien-Chun, C.: An artificial intelligence approach to course timetabling. In: Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '06)
2. Oyebanjo, A.: Development of a University Timetable Automation System, OTA, May 2012
3. Gotlieb, C.C.: The construction of class-teacher timetables. In: Proceedings of IFIP Congress, pp. 73–77. North-Holland Pub. Co., Amsterdam (1962)
4. Lawrie, N.L.: An integer programming model of a school timetabling problem. *Comput. J.* **12**, 307–316 (1969)
5. Tripathy, A.: School timetabling—a case in large binary integer linear programming. *Manag. Sci.* **30**(12), 1473–1489 (1984)
6. Abramson, D.: Constructing school timetables using simulated annealing: Sequential and parallel algorithms management-science (1991)
7. Dikman, R., Luling, R., Simon, J.: Problem independent distributed simulated annealing and its applications. Technical Report—Paderborn Center for Parallel Computing (1993)
8. Elaine, R.: Artificial Intelligence. ISBN-13: 978-0-07-008770-5
9. Zhang, L., Lau, S.: Constructing university timetable using constraint satisfaction programming approach. In: International Conference on Computational Intelligence for Modelling, Control and Automation (2005)
10. Hana, R., Keith, M.: University Course Timetabling with Soft Constraints, USA
11. Tomáš, M.: Constraint Based Timetabling. Prague (2005)
12. Sandhu, K.S.: Automating class schedule generation in the context of a University timetabling information system. Griffith University (2001)
13. Rina, D.: Constraint Processing. ISBN 1-55860-890-7

Application of Genetic Algorithm for Evolution of Quantum Fourier Transform Circuits

Swanti Satsangi and C. Patvardhan

Abstract Quantum Fourier Transform finds a variety of applications in quantum computing. It is the most important building block in a number of quantum algorithms like Shor's algorithm, phase estimation algorithm, etc. This paper illustrates the ability of Genetic algorithm for evolving these quantum fourier transform circuits on a classical computer. Circuits for two, three, four, and five qubits have been discussed in the paper, however. the algorithm has been generalized for evolving circuits for any number of qubits.

Keywords Genetic algorithm · Quantum fourier transform · Quantum circuits

1 Introduction

Quantum computation (QC) is a flourishing research area that makes use of properties of quantum mechanics like superposition and entanglement to perform operations on data [1]. These operations are performed in the form of unitary transformations on n -qubit registers which are quantum mechanical analog of n -bit registers [2]. A set of these transformations forms a quantum circuit. Due to the unitary nature, these circuits are always logically reversible unlike the classical case. Also, the qubits that enter a quantum circuit are not individual states, rather, they are superposition of states, i.e., in joint or entangled states.

S. Satsangi (✉)

Department of Physics and Computer Science, Dayalbagh Educational Institute,
Agra, India
e-mail: swantis@gmail.com

C. Patvardhan

Department of Electrical Engineering, Dayalbagh Educational Institute,
Agra, India
e-mail: cpatvardhan@gmail.com

© Springer India 2016

S.C. Satapathy et al. (eds.), *Proceedings of the Second International Conference on Computer and Communication Technologies*, Advances in Intelligent Systems and Computing 379, DOI 10.1007/978-81-322-2517-1_74

773

The recent spurt in interest in quantum computing can be attributed to two important reasons. The first is that it has been shown that quantum computation-based algorithms can sometimes provide speedups that are not possible at all in the classical world. This is true, for example, in the celebrated Shor's Algorithm which provides an exponential speedup over the classical counterparts in factoring integers. There are other problems for which quantum computation has been shown to provide good performance vis-à-vis the classical equivalents albeit not in as spectacular terms as the Shor's Algorithm. These problems include Grover's Search Algorithm, Deutsch-Jozsa Algorithm, and the quantum Fourier transform. The other reason is that continued miniaturization has been fuelling performance improvements in computers in keeping with the Moore's Law for the last five decades. But as the scales approach nanoscales quantum effects would have to be necessarily taken into account as the physics changes from classical to quantum.

Despite the spurt in interest due to reasons mentioned above quantum algorithms are notoriously hard to design primarily because human intuition is rooted in classical. Therefore, alternative means of automatically evolving quantum algorithms are being explored.

In the last decade, researchers have come up with the idea of automatic design of Quantum Algorithms using evolutionary algorithms. Out of several branches of evolutionary algorithms, genetic programming and genetic algorithms have been most frequently used for this purpose [3].

This effort investigates the use of Genetic algorithm (GA) to evolve circuits for Quantum Fourier Transform (QFT) on a classical computer. The rest of the paper is organized into following sections. Section 2 discusses the various implementations of evolutionary algorithms in evolving quantum circuits in the past. Section 3 gives a brief overview of quantum Fourier transform. Section 4 describes the approach adopted in this paper for evolving QFT circuits. Finally Sect. 5 discusses the experimental results.

2 Earlier Attempts at Evolving Quantum Circuits

Different evolutionary schemes and representations of quantum circuits have been proposed by researchers in the past for various circuits. Evolutionary algorithms have also been successfully used to figure out various complex optimization problems with little prior knowledge.

Williams and Gray [4] implemented evolutionary algorithm for designing of quantum circuits for the first time. Spector et al. [5] presented three different Genetic Programming (GP) schemes for quantum circuit evolution. Similar efforts involving GP were made by Rubinstein [6]. Leier and Banzhaf [7] introduced a GP system using linear tree structures. It could achieve more "degrees of freedom" in the construction and evolution of quantum circuits.

A generic genetic algorithm was introduced by Lukac and Perkowski [8, 9] which evolved arbitrary quantum circuit specified by a goal unitary matrix.

They encoded quantum circuit as a structured string of quantum gates which could reflect the parallel and sequential framework of quantum circuit. In their work, Ruican et al. [10], reported an automated quantum circuit synthesis approach, using a genetic algorithm and evolved various benchmark circuits. They considered the circuits as successive rippling of the so-called gate sections and used a gate-database in order to specify the gates that will be used in the synthesis process. Debarati et al. [11] have also shown that evolutionary computation can be used for automated quantum circuit development in real time. The experiments mainly have been done on arbitrary quantum circuits containing NOT, CNOT, Toffoli, Hadamard, and multiple controlled NOT gates.

Although discovering new quantum algorithms is an important goal, attempts have also been made in the literature to evolve circuits employing quantum gates that are functionally equivalent to classical circuits in realizing the same boolean function. The major advantage of this approach is that the classical circuits could be irreversible but the quantum circuits are reversible [1, 4].

Efforts have been made to implement Reversible Logic (RL) circuits using CMOS technology [12]. Lukac et al. [8] provide a strategy for the design of reversible circuits using Quantum Gates. Mukherjee et al. [11] and Mukhopadhyay et al. [13] have evolved quantum circuits for some basic boolean functions, e.g., half and full adder, encode, decoder, multiplexer, etc. Satsangi et al. [14] et al. proposed the use of variable length encodings of the chromosome strings for evolving several quantum circuits that implement standard Boolean functions.

Genetic algorithms have also found application in evolution of a number of special problems like quantum teleportation and quantum fourier Transform. Yabuki and Iba applied GA for the evolution of quantum teleportation which was reported to have fewer gates than the best known teleportation circuit [15]. Massey et al. [16] evolved a human competitive algorithm with genetic programming for the quantum fourier transform circuit for up to five qubits. A hybrid quantum inspired evolutionary algorithm was developed by Yang et al. [17]. It was successfully implemented for evolving three qubit Quantum Fourier transform circuits apart from the other quantum circuits using only adjacent two qubits gates and a strategy of combining the cost and correctness of quantum circuits into the fitness function. Another application of Hybrid Quantum Inspired Evolutionary Algorithm (HQIEA) has been shown by Satsangi et al. [18]. The paper illustrates evolution of Quantum equivalents of classical circuits like Binary-Gray converter, Boolean function solver, half adder and full adder.

3 Quantum Fourier Transform

Quantum Fourier Transform is described as a linear transformation on the vector of amplitudes of a quantum state and is therefore, the analog of the classical discrete fourier transform. A practical difference between the classical and the quantum fast fourier transform is that the latter provides the result of the calculation as a

superposition of quantum states, which cannot be directly read according to the postulates of quantum mechanics. Quantum Fourier Transform also improves the complexity to $O(\log 2 N)$ as compared to $O(N \log N)$ of its classical counterpart.

3.1 Mathematical Model

The quantum fourier transform converts a given quantum state denoted by $\sum_{i=0}^{N-1} x_i|i\rangle$ to another state denoted by $\sum_{i=0}^{N-1} y_i|i\rangle$ given as [2]:

$$y_k = \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} x_j \omega^{jk}$$

where $(x_0 \dots x_{N-1})$ and $(y_0 \dots y_{N-1})$ are the input and output probability amplitudes, respectively, and $\omega = e^{\frac{2\pi i}{N}}$ is the N th root of unity.

The unitary matrix F_N for QFT is represented as:

$$F_N = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \omega^3 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \omega^6 & \dots & \omega^{2(N-1)} \\ 1 & \omega^3 & \omega^6 & \omega^9 & \dots & \omega^{3(N-1)} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \omega^{3(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{bmatrix}$$

3.2 Circuit Model

The quantum fourier transform can be implemented in the form for a circuit for any number of qubits using a set of Hadamard and rotation gates. For N number of qubits, the QFT circuit consists of a set of $\frac{N(N-1)}{2}$ Hadamard and Phase gates. In addition to these, there are $\frac{N}{2}$ Swap gates. Figure 1 shows a 3-qubit Quantum Fourier Transform circuit evolved using GA while Fig. 2 shows a generalised QFT circuit for n-qubits.

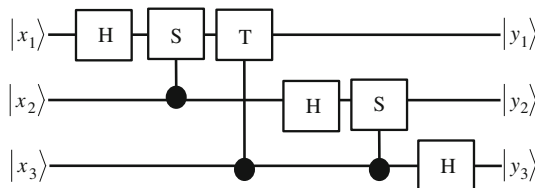


Fig. 1 A 3 qubit QFT circuit (swap gate not shown)

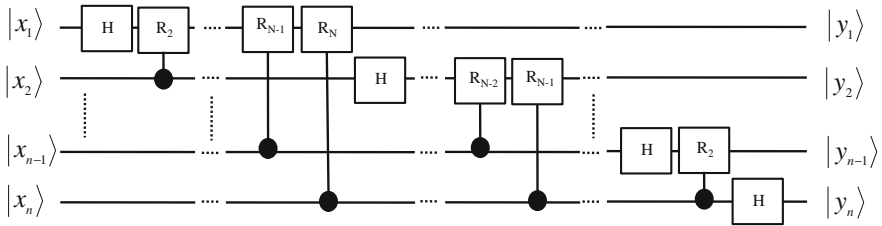


Fig. 2 A generalized representation of QFT circuit for n qubits

4 Evolution of Quantum Fourier Transforms Circuits

The circuit model adopted in this paper has been discussed below in detail.

4.1 Gate Library

The gate library consists of the Hadamard gate and the controlled R_k ($k = 2, 3, 4, \dots$) gates which are used in construction of standard QFT circuits found in the literature [1]. In a separate set of experiments, a few additional gates were included in the library. These were CNOT and controlled V/V^+ gates. It was observed that this extended gate library did not change the total number of gates in a circuit, rather substituted the Phase (R_2) gate with controlled- V or a combination of CNOT and controlled- V^+ gates. Figure 3a shows the 2 qubit QFT circuit obtained using the standard gate library and Fig. 3b, c show its equivalent circuit using the extended library. Thus, the controlled-Phase gate was substituted by a combination of Hadamard and control- V gates as in Fig. 4b and the control- V gate in turn was substituted by a combination of conjugate of control- V (CV^+) and a CNOT gate as in Fig. 4c.

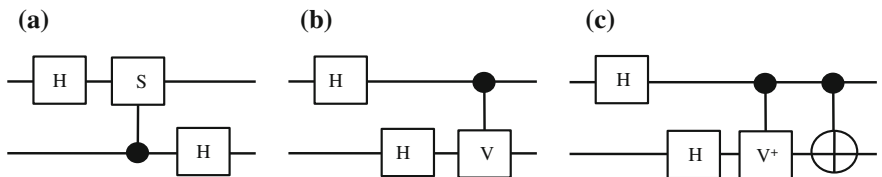


Fig. 3 Various 2 qubit QFT circuits evolved using GA

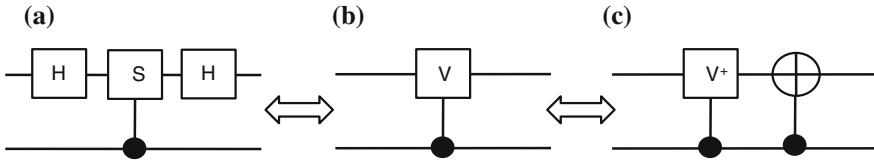


Fig. 4 Equivalent circuits

4.2 Gate Representation

Each gate is represented by an integer in a way that the same gate acting on different qubits is coded with different integers. For instance, CNOT(1,2) is denoted by a different integer as compared to CNOT(2,1). The integer allotted to each gate depends on the types of gates being used for the given problem and the number of qubits for which the circuit has to be generated. This coding scheme is different from the second-order representation mentioned by Massey et al. [16] for evolving QFT circuits using GP.

4.3 Population

Population with m chromosomes is randomly generated initially and is updated in each generation. Each of the m chromosomes contains p number of gates but this size of chromosome, p , is fixed for the entire population. Since the gates are represented by integers, a chromosome is simply a sequence of integers which forms a circuit. For instance, a circuit with five gates would be represented as [2 5 8 4 6] where each integer is denoting a gate from the library. In the work by Massey et al. [16], these individuals were not quantum circuits, but higher level constructs which had to be decoded to generate quantum circuits.

4.4 Fitness Function

Matrix-based approach for fitness assessment has been adopted in this paper, which calculates component wise difference metric on the obtained and the desired output matrices in order to find the exact equivalent of the desired circuit [4, 11]. The fitness is calculated as:

$$\text{Fitness} = \text{Match}/\text{Total}$$

where, the total number of elements in the obtained matrix is summed up as total and the number of elements that exactly match with their corresponding ones in the expected matrix, is match. The algorithm stops when the perfect fitness is obtained, i.e., $\text{Fitness} = 1$.

Another method for calculating the fitness could be by summing up all the deviations between the outputs of the evolved circuit and the desirable outputs which was adopted in [16].

Evaluation of individuals can also be done based on correctness, cost, and feasibility procedure [3]. The correctness procedure deals with the process in which the desired matrix is obtained. The feasibility procedure evaluates whether or not a given circuit is practically feasible on quantum computer and the cost procedure optimizes the obtained cost within a minimum and maximum range.

4.5 Selection Process

Tournament selection was tested against Roulette wheel selection. A number of experiments were performed using the two techniques separately. It was observed that the inclusion of tournament in the selection process resulted in an increase in the evolution time for a given circuit as compared to roulette wheel selection. This was due to the fact that in random selection of contesting candidates in the tournament, the probability of the fitter chromosomes getting selected for tournament and eventually as parent for next generation, is same as its weaker counterpart. On the other hand, roulette wheel being a fitness proportionate technique, ensured that the fitter candidates gets a better chance of survival and hence proved to be better suitable for the problem at hand.

4.6 Probability of Crossover and Mutation

The probability of crossover (P_c) and mutation (P_m) have been meticulously selected after a series of experiments with different values. In order to decide the value of probability of crossover, the probability of mutation was fixed to $P_m = 0.1$, P_c was varied from 0.1 to 0.7 and the time taken to evolve a given circuit was noted. It was observed that a probability of 0.7 gives the minimum evolution time as shown in the Table 1.

Similarly, P_c was fixed to 0.7 and P_m was varied from 0.1 to 0.4 and the evolution time was noted. No significant difference was found in the evolution time with mutation and hence P_m was chosen as 0.1. Any probability lower than this resulted in a considerable increase in evolution time.

Table 1 Variation in evolution time with change in the probabilities of crossover and mutation

Pc (with Pm = 0.1)	Time taken to evolve a 2 qubit QFT circuit (in seconds)	Pm (with Pc = 0.7)	Time taken to evolve a 2 qubit QFT circuit (in s)
0.1	78.263088	0.1	1.5604
0.3	31.725830	0.2	1.6186
0.5	10.0994	0.3	1.5427
0.7	1.5604	0.4	1.3237

5 Experimental Results

Using the fitness function mentioned in the previous section, a population of size 20 and probabilities of crossover and mutation as 0.7 and 0.1 respectively, a series of experiments were performed for 2, 3, 4, and 5 qubit. Unitary matrix for QFT was given as input for each case along with the desired number of gates. A circuit with the desired number of gates was successfully evolved in each run of the algorithm. The algorithm was run 10 times for each case. The evolution time and the number of generations reported in the Table 2 are the average time and generations recorded over 10 runs (Table 3).

Table 2 Summary of experimental results averaged over 10 runs

No. of qubits	No. of gates	No. of generations	Time taken (in s)
2	3	553	3.6
3	6	2041	12.9
4	10	68,300	781.2
5	15	141,362	9103.4

Table 3 Circuits produced by algorithm using standard and extended libraries

No. of qubits	Circuit using standard library	Circuit using extended library
2	H1, CR2 H2	H1, H2, CV(1,2)/H1, H2, CV + (1,2), CNOT(1,2)
3	H1, CR2(2,1), CR3(3,1), H2, CR2 (3,2), H3	H1, CR2(2,1), H2, CR3(1,3), H3, CV (2,3)
4	H1, CR2(2,1), CR3(3,1), CR4(4,1), H2, CR2(3,2), CR3(4,2), H3, CR2 (4,3), H4,	CV(2,1), H1, CV(1,2), H2, CR3(1,3), H3, CR3(2,4), CR4(4,1), H4, CV(2,3)
5	H1, CR2(2,1), CR3(3,1), CR4(4,1), CR5(5,1), H2, CR2(3,2), CR3(4,2), CR4(5,2), H3, CR2(4,3), CR3(5,3), H4, CR2(5,4), H5,	H2, H1, CV(1,2), CR4(4,1), CR2(3,2), CR3(2,4), CR3(1,3), H3, CR5(1,5), H4, CV(3,4), CR3(3,5), CR2(4,5), CR4 (5,2), H5

Massey et al. [16] evolved a human competitive algorithm with Genetic Programming for the Quantum Fourier transform circuits. Their research was conducted using software suite Quantum Programs and Circuits through Evolution (Q-PACE). Initially the evolved algorithm could construct QFT circuits for only 1, 2 and 3 qubits that were not as good as the best known circuits for QFT but later it was improved to a size-independent algorithm capable of generating a correct circuit for any supplied n . The drawback of using GP was that it evolved an algorithm which had to be further decoded and executed in order to implement an exact QFT circuit.

Another effort in this direction was made by Ding et al. [3]. Using a novel hybrid quantum inspired evolutionary algorithms, various quantum circuits like entangle 2 and 3, control-S and QFT-3 were evolved using only one qubit and adjacent two qubit gates. They introduced a reward-punish factor in the fitness function and the algorithm terminated only when a satisfiable cost was achieved. This constraint not only resulted in increase in the evolution time but also a reduced probability of finding a circuit satisfying the stopping criteria as compared to the GA algorithm discussed in the current paper.

6 Conclusion

In the realm of Quantum circuit design, evolutionary algorithms have proved to be very useful. The GA approach discussed in this paper is able to evolve optimum circuits for QFT with same number of gates as present in the best know circuits for QFT. Interestingly, a number of equivalent circuits have also been obtained for the same transformation matrix using both standard gate library as well as the extended library with CNOT and Controlled V/V^+ gates.

This paper discussed successful implementation of genetic algorithm for evolving Quantum Fourier transform circuits but the approach is not limited to QFT circuits. It is applicable to any reversible problem where the transformation matrix is available and is capable of obtaining more than one circuit for a given problem, provided there exists more than one circuit for it. Since genetic algorithms appear to be promising, one can venture into designing more complex circuits using these in future.

References

1. Nielsen, M., Chuang, I.: Quantum Computation and Quantum Information. Cambridge University Press, Cambridge (2000)
2. Wikipedia: http://en.wikipedia.org/wiki/Quantum_computing
3. Ding, S., Jin, Z., Yang, Q.: Evolving quantum circuits at the gate level with a hybrid quantum-inspired evolutionary algorithm. In: Soft Computing, Springer-Verlag, pp. 1059–1072 (2008)

4. Williams, C.P., Alexander, G.G.: Automated design of quantum circuits. QCQC'98, LNCS (1999)
5. Spector, L., Barnum, H., Bernstein, H.J., Swamy, N.: Quantum computing applications of genetic programming. *Advances in Genetic Programming* (1999)
6. Rubinstein, B.I.P.: Evolving quantum circuits using genetic programming. In: *Proceedings of the 2001 Congress on Evolutionary Computation* (2001)
7. Leier, A., Banzhaf, W.: Evolving Hogg's quantum algorithm using linear-tree GP. In: *Proceedings of the 2003 International Conference on Genetic and Evolutionary Computation: Part 1*, pp. 390–400 (2003)
8. Lukac, M., Perkowski, M.: Evolving quantum circuits using genetic algorithm. In: *Proceedings of the 2002 NASA/DOD Conference on Evolvable Hardware* (2002)
9. Lukac, M., Perkowski, M., Goi, H., Pivtoraiko, M., Yu, C.H., Chung, K., Jee, H., Kim, B., Kim, Y.: Evolutionary approach to quantum and reversible circuits synthesis. *Artif. Intell. Rev.* **30**, 361–417 (2003)
10. Ruican, C., Udrescu, M., Prodan, L., Vladutiu, M.: Adaptive and natural computing algorithms. *Lect. Notes Comput. Sci.* **4431**, 174–183 (2007)
11. Mukherjee, D., Chakrabarti, A., Bhattacharjee, D.: Synthesis of quantum circuits using genetic algorithm. *Int. J. Recent Trends Eng.* **2**(1) (2009)
12. Veiri, C., Josephine, A., Frank, M.: A fully reversible asymptotically zero energy microprocessor. MIT AI Laboratory (1998)
13. Mukhopadhyay, D., Si, A.: Quantum multiplexer designing and optimization applying genetic algorithm. *Int. J. Comput. Sci.* **7**(5) (2010)
14. Satsangi, S., Gulati, A., Kalra, P.K., Patvardhan, C.: Application of genetic algorithms for evolution of quantum equivalents of boolean circuits. *Int. J. Electr. Comput. Electron. Commun. Eng.* **6**(3) (2012)
15. Yabuki, T., Iba, H.: Genetic algorithms for quantum circuit design—evolving a simpler teleportation circuit. In *Late Breaking Papers at GECCO* (2000)
16. Massey, P., Clark, J.A., Stepney, S.: Human-competitive evolution of quantum computing artifacts by genetic programming. *Evol. Comput.* **14**(1), 21–40 (2006)
17. Yang, Q., Zhong, S., Ding, S.: A simple quantum inspired evolutionary algorithm and its application to numerical optimization problems. *J. Wuhan University (Natural Science Edition)* **52**(1), 21–24 (2006)
18. Satsangi, S., Patvardhan, C.: Design of reversible quantum equivalents of classical circuits using hybrid quantum inspired evolutionary algorithm. *International Advanced Computing Conference*. pp. 12–13 (2015)

An Improved Polygon Clipping Algorithm Based on Affine Transformation

Mugdha Sharma and Jasmeen Kaur

Abstract Today, Computer Graphics is used almost in all the domains including gaming, entertainment, education, CAD/CAM, etc. One of the most important operations in computer graphics is clipping, such as line clipping and polygon clipping. Its importance emerges from the fact that polygon clipping can be applied in VLSI CAD, GIS, garment industry, etc. Many algorithms exist at the present moment but the intersection calculations incur huge costs. Our paper proposes an algorithm based on affine transformation which eliminates degeneracies while clipping self-intersecting and multi-polygons. Experimental results show that the new algorithm outperforms Greiner-Hormann and Vatti Algorithms for real-time datasets which are used in the packing industry.

Keywords Computer graphics · Polygon clipping · Affine transformation · Degeneracy

1 Introduction

The domain of computer graphics has replaced the cumbersome and expensive equipments used in engineering sciences with specialized graphics hardware and software to create and manipulate images. Today, computer graphics is used in industry, business, government, education and entertainment.

The problem of clipping aims to extract those regions of the shape (line, circle, ellipse, etc.,) within a given specified window. If specific regions of a polygon inside a given window are selected, the problem is termed as ‘polygon clipping’. In

M. Sharma (✉) · J. Kaur
Rukmini Devi Institute of Advanced Studies, Delhi, India
e-mail: mugdha.sharma@rdias.ac.in

J. Kaur
e-mail: jasmeen.kaur@rdias.ac.in

such a case, the polygon which is clipped is termed as the ‘clip polygon’ and the window against which clipping is done is called the ‘constraint polygon’.

There are various algorithms for polygon clipping but almost all of them are capable of clipping only convex polygons and that too against rectangular constraint polygons. And if someone has tried to innovate within the existing algorithms, then it has led to various complex calculations which makes the algorithm slow and inefficient. So in this paper, we have proposed a new algorithm that uses shearing on the constraint polygon and the intersecting polygon edge and the inverse affine transformation yields the required results with less time complexity. The algorithm is suitable to clip self-intersecting and multi-polygons with the elimination of degeneracies.

So, to achieve the goal of developing a better polygon clipping algorithm, the following specific objectives are developed for this paper:

- To explore the polygon clipping algorithms and propose a novel approach that supports clipping of self-intersecting and multi-polygons with the help of Affine Transformations.
- To evaluate the potential of computer graphics and polygon clipping in packing industry.
- To interpret and analyze the performance of the model developed.
- At last, to forward recommendations based on the outcomes of the study.

In packing industry, a check is done whether an object (subject polygon) can be placed within a larger bin (Constraint Polygon 1). When the subject polygon does not fit completely within the constraint polygon, calculation of intersections cannot be avoided. These intersections are computed based on which the clipped-out polygon can be determined for placing in some other suitable bin (Constraint Polygon 2). The paper focuses on clipping the subject polygon against constraint polygon 1.

The paper is structured as follows. Section 2 provides the details of the work done in the field of clipping algorithms for polygons and various other techniques of computer graphics. Section 3 describes the different types of polygons and Affine Transformations. Section 4 discusses the proposed approach. In Sect. 5, the new proposed approach has been analyzed by comparing it with other polygon clipping algorithms like Cyrus-Beck, Skala, and Liu. After which the conclusion and scope for the future is provided in Sect. 6.

2 Related Work

Most of the existing algorithms clip only convex polygons and work against rectangular constraint polygons. The re-entrant polygon clipping algorithm proposed by Sutherland and Hodgeman [1] performs clipping successfully only for convex polygons and generates extraneous lines (degenerate lines) for concave

polygons. Liang-Barsky and Cyrus-Beck Algorithms [2] provide a parametric solution to the clipping problem. These algorithms are twice as fast as the re-entrant algorithm, but fail to handle degeneracies. Maillot's Algorithm [3] is more accurate and faster than the previous algorithms since it uses integer arithmetic. But this algorithm is limited to convex polygons.

The use of clock orientation in Weiler-Atherton [4] empowers it to clip concave and multi-polygons but makes it unsuitable for self-intersecting polygons. The Vatti Algorithm [5] is based on the sweep-line approach and deals polygon with holes and self-intersecting polygons. Greiner-Hormann Algorithm [6], which is predominantly used for interactive clipping operations in computer games is yet another solution that outperforms Vatti polygon clipping algorithm but does not deal properly with degeneracies. Owing to its simpler data structure, Liu's Algorithm [7] is more memory efficient than Greiner-Hormann. Kim-Kim Algorithm [8] eliminates the perturbation vertices used in Greiner-Hormann Algorithm. Even though Foster et al. Algorithm [9] is a solution that combines Liu and Kim-Kim algorithms, it incurs huge costs for self-intersecting and multi-polygons. Anurag Chakraborty's algorithm [10] extends Weiler-Atherton's algorithm and other concave polygon clipping algorithms to handle self-intersecting polygons. But this algorithm cannot extend Greiner-Hormann to clip self-intersecting polygons.

None of the aforementioned algorithms could overcome solving of equations or other complex calculations to determine intersections between clip and constraint polygons. Our algorithm uses Affine Transformations (Shearing) on the constraint polygon and the intersecting polygon edge and the inverse Affine Transformation yields the required results with less time complexity. Our algorithm avoids complex calculations and provides superior results for huge number of intersections.

3 Foundations

A polygon is a geometric object bounded by a finite set of straight line segments, i.e., *edges*. The points where two edges meet are the *vertices* of the polygon. A polygon can be *concave* or *convex* or *self-intersecting* or polygon with holes (*multi-polygons*). In a convex polygon, the line segment joining any two points interior to the polygon lies totally inside the polygon and all interior angles of a convex polygon is at most 180° . A concave (re-entrant) polygon which is a non-convex polygon can be divided into a set of convex polygons. Polygons may also include holes and self-intersections.

In general, large concave polygons are difficult to process and when processed may yield incorrect results. So, conversion of a concave polygon into a set of nonconvex polygons is unavoidable in certain situations. Polygon Triangulation can be used to split concave polygons (Fig. 1). Similarly, self-intersecting polygons

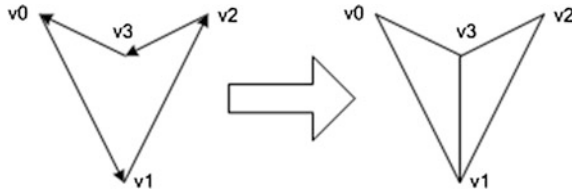


Fig. 1 Concave polygon and its triangulation into a set of convex polygons

may be split into a set of non-self-intersecting contours for easy processing. To identify the hole contours within a multi-polygon, inside-outside tests can be conducted. Two such methods are explained in Sect. 3.1.

3.1 Inside-Outside Tests

3.1.1 Parity Rule (Odd-Even Rule)

When the number of polygon edges crossed so far by a line joining a pixel P in a polygon to any distant point is even or odd, the pixel P is exterior or interior, respectively.

3.1.2 Winding Number Rule

When the number of times the edges wind around a pixel P in a polygon in anti-clockwise direction is nonzero or zero, P is exterior or interior respectively.

Figure 2 shows the different cases of Clipping.

3.2 Affine Transformations

Affine Transformations are special kind of transformations that preserve collinearity, parallel line segments and ratio of distances but do not preserve angles. Translation, rotation, scaling, reflection, and shear are affine transformations. The inverse of affine transformation is also an affine transformation. When affine transformation is applied to an oblique line (neither horizontal nor vertical, i.e., $m \neq \infty$ or $m \neq 0$), the transformed line becomes horizontal or vertical.

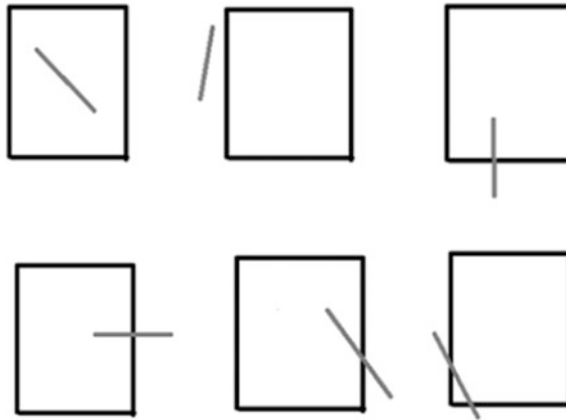


Fig. 2 Different cases of clipping, **a** completely inside, **b** completely outside, **c** horizontal edge intersecting the window, **d** vertical edge intersecting the window, **e** oblique line intersecting the window, **f** oblique line intersecting the window at two of its boundaries

Affine Transformation along x -axis and y -axis are given as:

$$x' = x - (y/m); y' = y \tag{1}$$

$$x' = x; y' = y - mx \tag{2}$$

4 The Proposed Approach

Degenerate edges are extraneous edges formed while clipping concave, self-intersecting and multi-polygons. The result of clipping a polygon is a set of vertices either the original input vertices or intersection vertices. Joining these vertices may yield such extraneous edges. Most of the algorithms like Greiner-Hormann, Vatti, etc., use topological information to handle such edges. This may result in high space complexity. To avoid this, such polygons are converted into a set of convex polygons.

A stack S is used for splitting a polygon with self-intersections and multi-polygons. All the vertices of the polygon are processed in a uniform orientation and pushed into the stack. When a match occurs, all the vertices are popped till the matching vertex is found as given in Fig. 3. The pseudo-code is given in Fig. 4. The resulting subsets of polygons, thus obtained are processed separately and are given as input to the new proposed clipping algorithm using affine transformations.

The proposed clipping algorithm takes only convex subject polygons as input and works for rectangular constraint polygons or window W . Concave polygons are

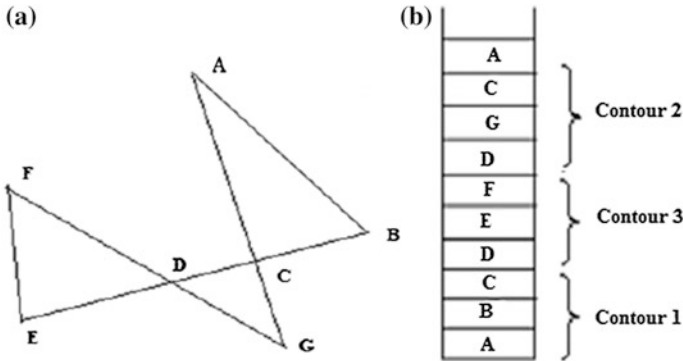


Fig. 3 a Self-intersecting Polygon, b splitting it into different contours (which may be convex or concave type)

Fig. 4 Steps to divide a self-intersecting polygon into a set of non-self-intersecting contours

```

/* stack structure*/
struct stack {
    int s[size];
    int top;
}st;

/*To divide self intersecting polygon*/
st.top=-1;
for(i=0;i<n;i++)
begin : push(a[i]);
    if(i!=0)
        begin : j=st.top;
            for(k=st.top-1;k>=0;k--)
                begin : if(st.s[j]==st.s[k])
                    begin : cntr++;
                        for(count=0;count<(j-k);count++)
                            begin : item=pop();
                                contour[cntr][count]=item;
                            endfor
                        endfor
                    endif
                endfor
            endif
        endfor
    endif
endfor
endfor
    
```

triangulated and the resultant triangles are given as input to the clipping algorithm described below.

Let X_L, X_R, Y_T, Y_B be four edges of the window W . The algorithm processes each edge of the subject polygon. Let (x_1, y_1) and (x_2, y_2) be the two endpoints of the edge.

The steps of the algorithm are as follows:

1. Completely inside: The x and y coordinates of the subject polygon must lie within (X_L, X_R) and (Y_B, Y_T) . i.e., $X_L \leq x_1, x_2 \leq X_R$ AND $Y_B \leq y_1, y_2 \leq Y_T$
2. Completely outside: The x and y coordinates of the subject polygon must lie outside (X_L, X_R) and (Y_B, Y_T) . i.e., $(x_1, x_2) \leq X_L$ OR $(x_1, x_2) > X_R$ OR $(y_1, y_2) \leq Y_B$ OR $(y_1, y_2) > Y_T$

3.
 - A. Horizontal edge intersecting the constraint polygon ($y_1 == y_2$): Reorder or swap the two vertices if $x_1 > x_2$.
 - a. If $x_1 \leq X_L$ (*left of the left edge of the window*), then the intersection point is (X_L, y_1) .
 - b. If $x_2 \geq X_R$ (*right of the right edge of the window*), then the intersection point is (X_R, y_1) .
 - B. Vertical edge intersecting the constraint polygon ($x_1 == x_2$): Reorder or swap the two vertices if $y_1 > y_2$.
 - a. If $y_1 \leq Y_B$ (*bottom of the bottom edge of the window*), then the intersection point is (x_1, Y_B) .
 - b. If $y_2 \geq Y_T$ (*top of the top edge of the window*), then the intersection point is (x_1, Y_T) .
4. An oblique edge intersecting the constraint polygon: Calculate the shear factor $sh_x = (x_2 - x_1)/(y_2 - y_1)$.
 - A. If $y_1 > y_2$, reorder the vertices.
 - a. Apply x -axis shear to the pair of vertices and to the rectangular window which gets transformed to vertical edge parallel to x -axis and parallelogram (X'_L, X'_R, Y'_T, Y'_B) , respectively.
 - b. If x_1' lies between X'_{LB} and X'_{RB} , check whether $(y_1' \leq Y'_B)$ AND $(y_2' \geq Y'_B)$ and if true, assign $y_1' = Y'_B$. Perform inverse x -axis shear transformation to calculate the intersection vertex.
 - c. If x_1' lies between X'_{LT} and X'_{TR} , check whether $(y_1' \leq Y'_T)$ AND $(y_2' \geq Y'_T)$ and if true, assign $y_2' = Y'_T$. Perform inverse x -axis shear transformation to calculate the intersection vertex.
 - B. If $x_1 > x_2$, swap the vertices and determine shear factor $sh_y = 1/sh_x$.
 - a. Apply y -axis shear to the pair of vertices and to the rectangular window which gets transformed to horizontal edge parallel to y -axis and parallelogram (X'_L, X'_R, Y'_T, Y'_B) , respectively.
 - b. If y_1' lies between Y'_{LB} and Y'_{LT} , check whether $(x_1' \leq X'_L)$ AND $(x_2' \geq X'_L)$ and if true, assign $x_1' = X'_L$. Perform inverse y -axis shear transformation to calculate the intersection vertex.
 - c. If y_1' lies between Y'_{BR} and Y'_{TR} , check whether $(x_1' \leq X'_R)$ AND $(x_2' \geq X'_R)$ and if true, assign $x_2' = X'_R$. Perform inverse y -axis shear transformation to calculate the intersection vertex.

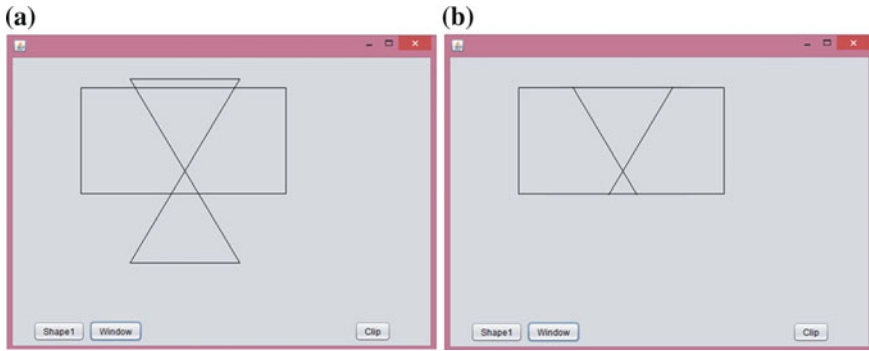


Fig. 5 a A self-intersecting polygon with a clipping window. b Clipped polygon

Figure 5 shows the output of the proposed clipping algorithm which has been implemented in JAVA. We can also see that the degeneracies have been eliminated in this proposed algorithm.

5 Analysis

For testing the efficiency of our algorithm, comparison with other polygon clipping algorithms like Cyrus-Beck, Skala, and Liu is done. From Table 1, it is inferred that the proposed method requires least number of computations, thereby reducing the time complexity.

Our algorithm is implemented using JAVA and is run using Netbeans IDE. An artificially created dataset containing 480 test instances of convex and concave polygons with 3–8 vertices is used. In general, the constraint polygon is set very large in many applications and so, most of the clip polygons lie entirely within it,

Table 1 Comparison of number of arithmetic operations to clip a polygonal edge against edge of the window by existing and new algorithms

Algorithm	Line does not intersect the polygon edge			Line intersects the polygon edge		
	Division, multiplication	Addition, subtraction	Comparison	Division, multiplication	Addition, subtraction	Comparison
Cyrus-Beck	5	6	6	5	6	6
Skala	2	1	4	5	6	4
Liu	1	1	4	5	6	4
Our proposed algorithm	0	0	2 (Step 1 and 2 in Sect. 4)	2	5	1 (Step 3 A and B in Sect. 4) 5 (Step 4 A and B in Sect. 4)

Table 2 Execution time (in s) of the new algorithm for n number of edges

n	Execution time (in s)
3	0.0035
4	0.0037
5	0.0059
6	0.0087
7	0.0127
8	0.0153

thereby reducing the number of intersection calculations. In any algorithm, the determination of intersections is crucial, unavoidable and time-consuming.

Table 2 shows the running time for clipping the polygon datasets against an optimal constraint polygon. Nearly, 50 test cases for each ‘ n ’ is found and the average is computed. The number of intersections is quite negligible larger constraint polygons are chosen.

6 Conclusion and Future Scope

Our algorithm can clip convex, concave, self-intersecting, and multi-polygons successfully. It also resolves the problem of degeneracies that arise while clipping nonconvex polygons by applying polygon triangulation instead of using any of the concave polygon splitting methods. From the experimental results, it is shown that it is simple and fast. In future, we aim to extend the polygon clipping algorithm to clip the subject polygon against any arbitrary-shaped windows.

References

1. Sutherland, E., Hodgman, G.W.: Reentrant polygon clipping. *Commun. ACM* **17**(1), 32–42 (1974)
2. Liang, Y.-D., Barsky, B.A.: An analysis and algorithm for polygon clipping. *Commun. ACM* **26**(11), 868–877 (1983)
3. Maillot, P.G., New, A.: Fast method for 2D polygon clipping: analysis and software implementation. *ACM Trans. Graphics* **11**(3), 276–290 (1992)
4. Weiler, K., Atherton, P.: Hidden surface removal using polygon area sorting, computer graphics. In: *Proceedings of SIGGRAPH*, vol. 11(2), pp. 214–222, July 1977
5. Vatti, B.R.: A generic solution to polygon clipping. *Commun. ACM* **35**(7), 56–63 (1992)
6. Greiner, G., Hormann, K.: Efficient clipping of arbitrary polygons. *ACM Trans. Graph.* **17**, 71–83 (1998)
7. Liu, Y.K., Wang, X.Q., Bao, S.Z., Gombosi, M., Zalik, B.: An algorithm for polygon clipping, and for determining polygon intersections and unions. *Comp. & Geosci.* **33**, 589–598 (2007)
8. Kim, H., Kim, M.-J.: An extension of polygon clipping to resolve degenerate cases. *Comp.-Aided Des. & Appl.* **3**(1–4), 447–456 (2006)

9. Foster, E.L., Overfelt, J.R.: Clipping of arbitrary polygons with degeneracies. *ACM Trans. Graphics* 1–5 (2014)
10. Chakraborty, A.: An extension of Weiler-Atherton algorithm to cope with the self-intersecting polygon. Cornell University Library, *Computer Graphics*. arXiv:1403.0917v1 Submitted on 4 March 2014

Alternative Design Space Analysis for Electronic Commerce System

P. Rajarajeswari, D. Vasumathi and A. Ramamohanreddy

Abstract Electronic commerce system is a popular way of web application services for business people, customers, and Employees. We observed so many challenges of electronic commerce system in an information system. Electronic commerce system is used for doing business transactions, funds transfer that involves transfer of information through the internet. Electronic commerce system is used for buying the products or selling the products through the internet. This paper presents the model of electronic commerce system with different design alternatives. Implementation of different design alternatives is used for modeling of electronic commerce system. In this paper, alternative design space analysis is used for deriving different design alternatives of the system. This paper presents an architecture view of the system based on UML modeling language. Alternative design space analysis can also applicable for designing industrial applications in an electronic commerce system.

Keywords Electronic commerce system · Software architecture · Object oriented design modeling · Alternative design space analysis

1 Introduction

Software architecture solves the design problems for large type of systems. It provides the correct arrangement of component and connectors. It gives high-level design structure of software system. It is developed based on requirement

P. Rajarajeswari (✉) · D. Vasumathi
Department of Computer Science and Engineering,
Jawaharlal Nehru Technological University, Hyderabad, India
e-mail: perezpicse@gmail.com

A. Ramamohanreddy
Department of Computer Science and Engineering,
Sri Venkateswara University, Tirupathi, India
e-mail: ramamohansvu@yahoo.com

analysis of functional and non-functional requirements. It is achieved with the result of assembling a certain number of architectural elements in a well manner [1]. It also achieves the performance requirements of the system. ADL language and UML languages are used for describing software architecture design process [2].

Electronic commerce system uses computer and communication technologies for doing sales process of organizations. Various companies are able to purchase the items by using electronic commerce system. This type of system consists of different modes for purchasing the items by using web server model. Description of electronic commerce system is given in Sect. 2. In Sect. 3 model of electronic commerce system is presented. Alternative design space analysis is applied for electronic commerce system that is given in the Sect. 4. Heuristic rules are given in the Sect. 5. Implementation of domain model for electronic commerce system is given in Sect. 6. Conclusions are presented in the Sect. 7.

2 Electronic Commerce System

E-Commerce (EC) system is one of the applications for Web services. It provides various services for customers, partners, and employees [3]. E-commerce company services are able to pay the amount based on market demands. This type of system manages the production of various companies, deliver the product within the given time based on market demands, trade, and management [4]. E-commerce system is used by various companies for reducing the budget cost and tries to avoid wastage of more resources. It can be used for doing business activities. Several factors are related with the progress of electronic commerce system development. Its requirements are altered into its constraints because of the drawback of enterprise size, technical force, and economic strength. Normally electronic commerce system is used for purchasing goods. Computers, networks, and software components are to be used as the devices for performing activities of electronic commerce system.

Essential components of electronic commerce system:

1. Catalog management
2. Content management
3. Search management
4. Access control and security
5. Profiling and Personalizing

2.1 Process Flow of Integrated E-Commerce Model

It consists of three phases.

1. Buyer phase
 2. Exchange phase
 3. Seller phase
1. *Buyer phase* Checkout the item in the database source based on client request. If that item is available in the database source and fulfill order for that item. Ordered item is delivered to seller and purchase that selected item by the buyer. Use and maintain that item based on market demands.
 2. *Exchange phase* Information is transferred in between the buyer and seller phases for purchasing products through the internet. Interactions are made with the help of personal agents for deploying the products to the customers.
 3. *Seller phase* Seller identifies the customer items, arrange that items which are available in the database and fulfill the order of selected items by buyers. Dispose those items to the buyers with the help of sellers. Process flow of integrated e-commerce model is shown below. Buyer phase, exchange phase, seller phase are the main phases for business to customer model in online shopping process (Fig. 1).

2.2 Architecture for Electronic Commerce System

Architecture for electronic commerce system consists of online shopping and electronic procurement system. Online shopping process is performed either in

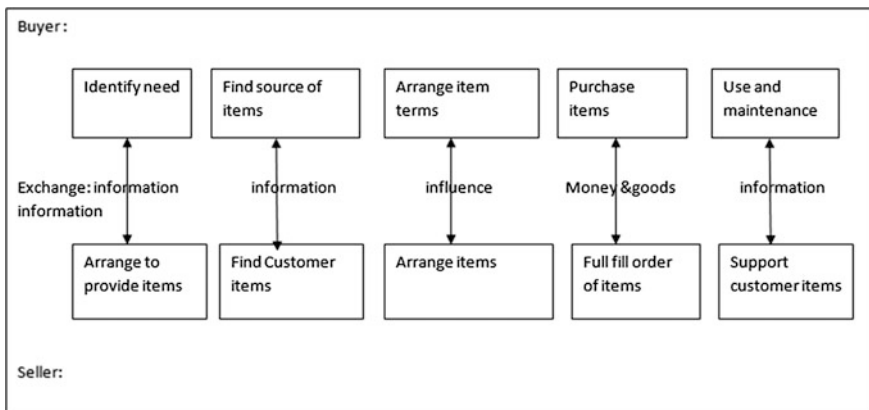


Fig. 1 Process flow of integrated E-Commerce model

Business to Business (B2B) or Business to Consumer (B2C). Government sectors are able to do their business by using electronic procurement system. Today electronic procurement system is of two classes' supplier centric system and buyer centric system. In supplier centric system suppliers are able to manage and update the information based on catalog of products. In buyer centric system buyers are able to maintain and usage of the catalog of products [5].

2.3 Example: Web-Based E-Commerce Architecture System

Web-based ecommerce architecture consists of client, tier-1 to tier-N, and data management system. This type of architecture provides communication in between the various components and gives their functions. Web-based e-commerce architecture provides communication in between client and Servers. e-commerce items are purchased throughout the worldwide by using this type of system (Fig. 2).

3 Models of Electronic Commerce System

Electronic commerce system is one part of e-business which can deal with transactions electronically in between various organizations and third party. Clients are allowed to purchase their products through e-shopping, e-billing process. Electronic commerce system consists of four levels in a conceptual way. These are online shopping, Electronic procurement system, Payee and Payment system [6]. In case of Online shopping item buyers are able to pay the items based on business to business or business to consumer model.

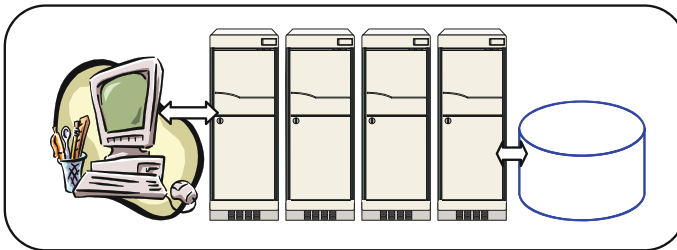


Fig. 2 Web-based e-commerce architecture system

3.1 Conceptual Level Model of Electronic Commerce System

Online shopping, electronic procurement system, payment system payee is the major levels of electronic commerce System. Online shopping system is performed by using business to business model or business to consumer model, catalog and product are the major components in electronic procurement system. In case of Payment system, products or items are purchased by the buyers with the help of cash mode, smartcard, credit card options (Fig. 3).

3.2 Architecture View of Electronic Commerce System

Architecture view of Electronic commerce system is described with the help of UML modeling language [7]. Static diagrams, dynamic diagrams, behavior diagrams are used to describe the structure and behavioral aspects of the system. Among these diagrams class diagram shows the structural relationships in between the components and connectors. Use diagram shows the behavioral aspect of the system.

3.2.1 Class Diagram

Class diagram shows the relation in between the classes and their relations. 1 ... n, 1 ... 1 relations are provided in between the classes and their relationships. It provides structural relationship in between the components and their connectivity. Each class can be considered as one of the component in this system (Fig. 4).

3.2.2 Use Case Diagram

Use case diagram shows use cases and actor. This diagram can be shown below. Each use case shows the operation of the customer (Fig. 5).

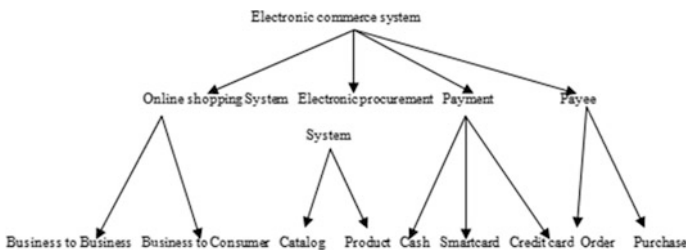


Fig. 3 Conceptual level model of electronic commerce system

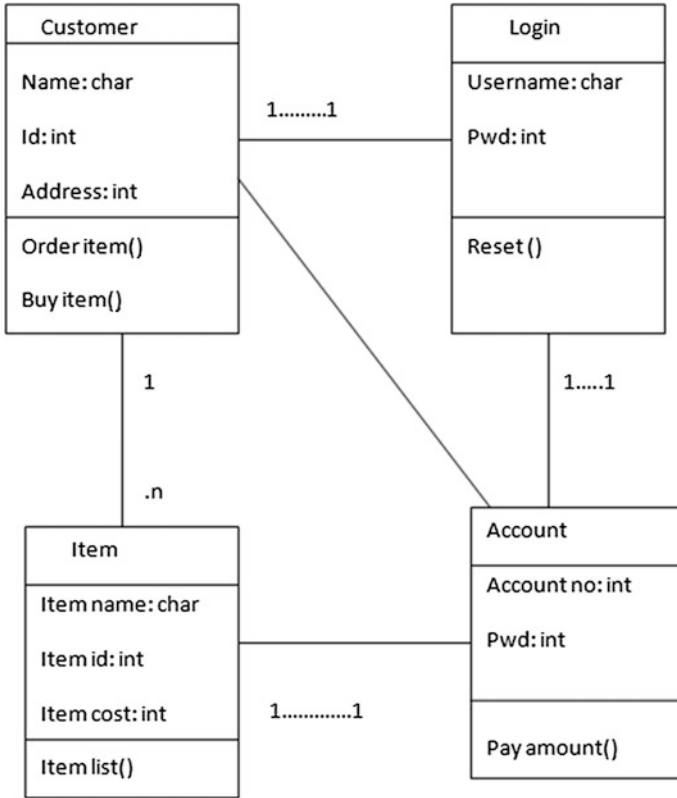
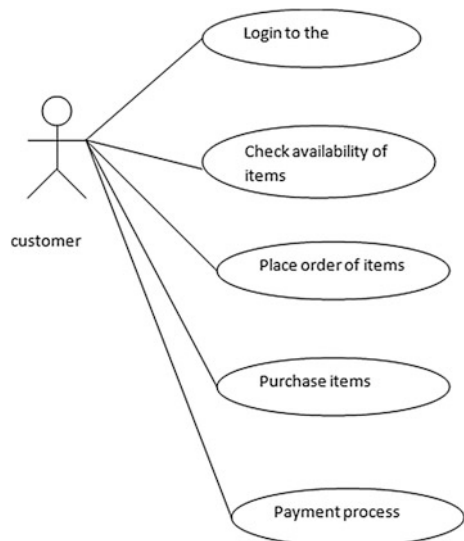


Fig. 4 Class diagram of electronic commerce system

Fig. 5 Use case diagram of electronic commerce system



4 Alternative Design Space Analysis for Electronic Commerce System

Alternative design space analysis is used to define the design space and derive the various design alternatives of system in the design space. Same system can be shown in different ways based on different characteristics is called design alternatives Alternative design space analysis is used for constructing the design space model [8]. In this analysis generate the design space and construct the design space with multidimensional values. Concepts are taken as dimensions for the design space. Main concepts for Electronic commerce system are online shopping, electronic procurement system, payment system, payee. Business to business, business to consumer, catalog, product, cash, smartcard, credit card, order, purchase are the sub-concepts of electronic Commerce system. Concepts of electronic commerce system names are taken as column names and sub-concepts are taken as entry values for each column (Table 1).

Design space model This model is defined with the help of concepts in the multidimensional space [9]. Concepts are considered as coordinates of the design space and sub-concepts are the dimensions for each set of coordinates in the design space. Constraints or design space rules are used for defining the design alternatives of the system in the design space.

Reduction of design space Selection or Elimination of design alternatives is possible by using logical operators to do the reduction of design space.

Representation of Design space We define a model by using design algebra concept in the design space. Design alternatives are derived in the design space for a given design problem. Design space is spanned by a set of independent dimensions. Online shopping, electronic procurement system, payment system, payee are the dimensions of design space. Here ‘ \wedge ’ represents the composition relation that is a mandatory relation, ‘ \vee ’ represents the OR relation, ‘;’ represents the alternative relation, ‘?’ represents the optional relation. Coordinates of the design space are business to customer or business to business that are to be taken as the dimensions for the design space of online shopping item concept. We define the model of online item as follows:

$$\text{Online item} = \{\text{online item system} \wedge \text{payment system} \wedge \text{payee}\}$$

In this case, online shopping item process may be performed by using business to business or business to consumer model.

Payment system may be cash mode, credit card mode and smartcard mode.

Payee maybe order mode, purchase mode.

Consider for example: Define the payment system as follows.

Table 1 Alternatives for the concept of Electronic commerce system

Online shopping system	Electronic procurement system	Payment	Payee
Business to business, Business to customer	Catalog, product	Cash, smartcard, credit card	Order, purchase

Payment system = {Cash mode \vee Credit card mode \vee Smart card mode}

Total set of design alternatives for buying an item through the electronic commerce system maybe 96 ways. Query-based approach is used for selecting and eliminating design space with the help of conditions.

Select concept from design Model where <Condition>

Reduction of design space is possible for the design of electronic commerce system with the selection of items by using query condition which is given below.

Electronic commerce item: Select from online shopping item

Where <online. Item and (business to consumer or business to business)>

Implementation of item is possible by mapping the domain alternatives in the design space. We can specify the model in the following form.

Property set of object can be defined as follows

$P_{\text{object}} = (\text{CL}; \text{OP}; \text{AT})$ {CL: class, OP: operation, AT: attribute} the symbol'; specifies the alternative feature.

$M_{\text{item}} = (\text{TY}, \text{CT}, \text{NO})$ {TY: Type, CT: Cost, NO: Number}

Design space is defined with the mapping relation. Mapping the elements of item to the elements of the object. Design space: $M_{\text{item}} \rightarrow P_{\text{object}}$

Design alternatives for Electronic commerce item are found by using Object oriented design model in the design space which is defined as follows.

Electronic commerce item: Object-electronic item: = Electronic commerce item. weave (Object). Implementation of all the alternative object-oriented electronic commerce item includes in the set of object-electronic item. This set includes 96 design alternatives.

5 Heuristic Rules Method

This type of method uses heuristic rules with the help of conditional statement which is in the form of IF <condition> THEN <consequent> statements. Consequent portion is either selection or elimination operations for reducing the design space [10]. Design alternatives are selected from object-electronic Commerce item may be specified in a query. These rules are decided to include only alternatives for the Electronic commerce item.

Object-electronic commerce item: Select from Electronic commerce item

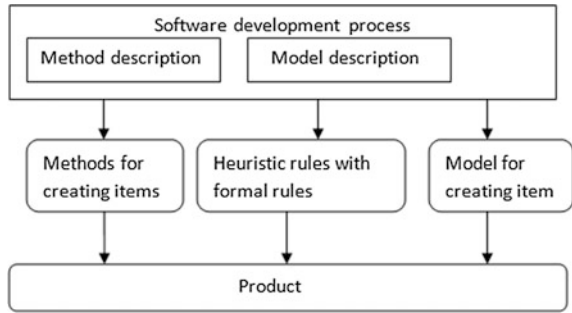
Where <itemtype.CL and item selection.OP>

We got 96 alternatives by using the above rule. This type of design rules may also reduced by Stakeholder constraints [5].

5.1 Modeling for Heuristic Rule Method

Development of software process is performed in a formal way that can be considered as a new technique in case of the knowledge-based engineering. Heuristic

Fig. 6 Modeling of Heuristic rule method



rules are used for doing formal Software development method. In this type of modeling two types of descriptions can be considered. One is model description and other one is method description. Method description provides description for methods which are used in software process. Model description is performed by using heuristic rules is shown (Fig. 6).

6 Implementation of Domain Model for Electronic Commerce System

Four design alternatives are possible for the sub-concept of online shopping item process. The sub-concept of Payment mode can be done in three ways. Payee is in two ways are possible features of Design alternatives. Number of design alternatives for the implementation of this system is 96.

In design algebra concept we calculate the num of design alternatives for doing the computation of product alternatives from a given domain model. Each type of

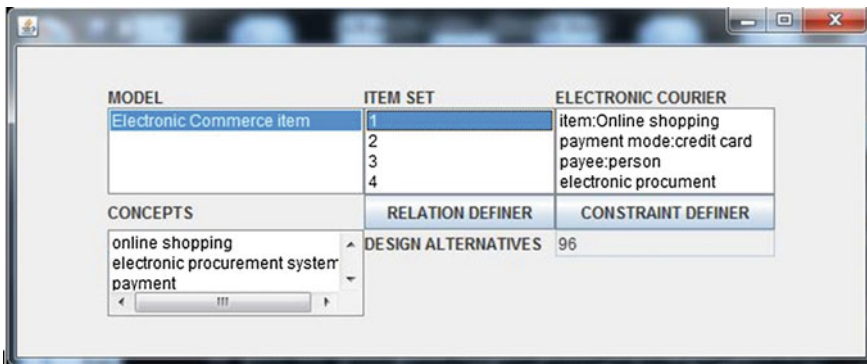


Fig. 7 Model of electronic commerce item

MODEL	Online Shopping Item		ITEM ALTERNATIVES
PROJECT	BUSINESS TO BUSINESS	1	item:online shopping
	BUSINESS TO CUSTOMER	2	payment mode:credit card
		3	payee:person
QUERY	online shopping item.person and (business to business)or(business to customer)		

Fig. 8 Model of online shopping item

design alternatives can be selected based on design constraints. Electronic commerce system is implemented by using java programming language.

Model definer tool for Electronic commerce item and online shopping item is shown below (Figs. 7 and 8).

7 Conclusions

- In this paper, we provide description of Electronic commerce System model in a Conceptual way. We presented alternative design space analysis for Electronic commerce system.
- Various design alternatives are derived and implemented by using design algebra techniques in the design space. In future work design algebra concept can also be used for doing design process of industrial product and transaction system.

References

1. Clements, P., Northrop, L.: Software product lines: practices and patterns, August 30, 2001, 3rd edn. Addison-Wesley, 2001. ISBN-13:078-5342703320
2. Shaw, M., Garlan, D.: Software architectures: perspectives on an emerging discipline, Englewood Cliffs. Prentice-Hall, NJ (1996)
3. Abou, A., Saleh, E.: A proposed framework based on cloud computing for enhancing e-commerce applications. *Int. J. Comput. Appl.* **59**(5), 0975–8887 (2012)
4. Whinstone, A.: The economics of electronic commerce. MacMillan Publishing Company, July 1, 1997. ISBN-10:1578700140
5. Choi, S.-Y., Stahl, D.O., Whinston, A.B.: Economics of doing business in the electronic market place. Macmillan Technical Publishing, Indianapolis (1997)
6. Shaw, M.J.: Electronic commerce: state of art. In: Shaw, M., Blanning, R., Strader, T., Whinston, A. (eds.) *Handbook on electronic commerce*, Chap. 1, pp. 3–24. Springer (2001)

7. Booch, G., Rumbaugh, J., Jacobson, I.: The Unified Modeling Language User Guide, 1st edn, p. 512. Addison-Wesley, 20 Oct 1999. ISBN:0-201-57168-4
8. Tekinerdogan, B.: Synthesis-based software architecture design. PhD Thesis, University of Twente, Department of Computer Science, Netherlands, March, 2000
9. Ossher, H., Tarr, P.: Multi-dimensional separation of concerns using hyperspaces. IBM Research Report 21452, April 1999
10. Riel, A.J.: Object-oriented design heuristics, 1st edn, Addison-Wesley, (1996). ISBN-020163385X

Author Index

A

Aavula, Ravi, 449
Abhinav, E. Meher, 77, 267
Abirami, S., 205, 225
Afreen, Nida, 47
Ahmad, Musheer, 471
Ahmad, Zishan, 471
Alam, Aftab, 109
Alam, Bashir, 463, 751
Albhaishi, Ahmad, 109
Alvi, A.S., 103
Ansari, Abdul Mateen, 109
Anuraj, S., 187
Asha Rani, M., 143

B

Bagchi, Angshuman, 401
Bandyopadhyay, S.K., 55
Banerjee, Arundhati, 401
Banka, Haider, 605
Barga, Mohammad Mujahid, 109
Behera, H.S., 343
Bhadra, Somasree, 35
Bhateja, Vikrant, 1
Bhatia, Saloni, 319, 329
Bhattacharya, Mahua, 731
Bichkar, R.S., 557
Bigul, Sunitha Devi, 449
Biswas, Siddhartha Sankar, 463
Boddu, Rama Devi, 143

C

Chary, Vemana, 67, 267
Chatterjee, Rajeev, 527
Chintalapudi, S. Rao, 519
Chopra, Deepti, 257

D

Dadhich, Ajay, 713

Das, Ajanta, 629
Das, Nibaran, 703
Das, Rupayan, 679
Das, Syaamantak, 527
Desai, Chitra G., 411
Deshpande, Ketan, 761
Dey, Paramita, 125
Dharavath, Ramesh, 497
Doja, M.N., 463, 751
Dwivedi, Amit Krishna, 365

G

Gaikwad, Suhas Machhindra, 743
Gawas, Sanket, 761
Geethanjali, N., 275
Ghafir, Ibrahim, 661
Gireeshkumar, T., 11, 187
Gopala Krishna, P., 479
Guduri, Manisha, 365
Gupta, Mukesh, 713
Gupta, Sindhu Hak, 353
Gupta, Suneet K., 721
Gupta, Swarnima, 651

H

Halder, Chayan, 703
Hameed, Pinjari, 443
Hamodi, Jamil M., 375
Himanshi, 1
Hooda, D.S., 387

I

Islam, Aminul, 1, 365

J

Jain, Divya, 387
Jana, Prasanta K., 509, 605, 721
Jannu, Srikanth, 509
Jose, Akshay Eldho, 11

Joshi, Rahul Raghvendra, 743
Jyothi, B., 289

K

Kakulapati, Vijayalakshmi, 449
Kalshetti, Urmila, 761
Kanungo, D.P., 343
Kashyap, Manish, 731
Kaur, Arvinder, 257
Kaur, Jasmeen, 783
Kaur, Kamaldeep, 257
Kaushal, Hemani, 319, 329
Kavuri, Sai Naveen, 77
Khandelwal, Sarika, 133
Khatoon, Asma, 47
Kishan Rao, K., 143
Koikara, Rosemary, 443
Krishn, Abhinav, 1
Krishna Madan, Y., 1
Krishna Mohan, P.G., 289
Krishna Prasad, M.H.M., 299, 519
Kuila, Pratyay, 721
Kumar, A.V. Ravi, 235
Kumar, B. Sudheer, 275
Kumar, Raghvendra, 21, 29
Kumar, Sandeep, 731
Kumar, Thota Sandeep, 77
Kumar, V. Vijaya, 425
Kundu, Anirban, 35

L

Lalitha Bhaskari, D., 479

M

MadhaveeLatha, Y., 289
Mahesh, Ch., 617, 691
Majumder, Koushik, 679
Malhotra, Akshay, 537
Mandal, Jyotsna Kumar, 527
Manza, Ramesh, 85
Meher Abhinav, E., 67
Mehra, Pawan Singh, 751
Mehra, Rishab, 365
Mewara, H.S., 713
Mishra, Arunodaya Raj, 387
Mitra, Siba, 629
Mohammed, Habeeb Vulla, 567
Mohan, Chandra, 67, 267
Mohan, M. Chandra, 77
Mohan, Vaka Murali, 163
Mulay, Preeti, 195, 743
Murali, V., 353
Murthy, J.V.R., 425
Murugappan, S., 205, 225

N

Nahar, Deepika, 761
Naik, Bighnaraj, 343
Nayak, Biswajit, 153
Nayak, Janmenjoy, 343
Nedungadi, Prema, 639

O

Obaidullah, S.M., 703

P

Pal, Soumitra, 1
Pankaj, Dhakate, 175
Parwekar, Pritee, 593
Pattnaik, Prasant Kumar, 21, 29
Patvardhan, C., 773
Pradeepa, T., 205
Pradhan, Manoj Kumar, 153
Prasad, V. Kamakshi, 617, 691
Praveen, K., 215
Premalatha, P., 187
Prenosil, Vaclav, 661
Puri, Karuna, 195

R

Radhika, Y., 163
Ragunathan, T., 275
Rajarajeswari, P., 793
Rama Mohan Reddy, A., 309
Ramamohanreddy, A., 793
Rangaswamy, B., 275
Ratanpara, Tushar V., 487
Raut, Prachi, 245
Ravindra, K., 617, 691
Ray, Sujay, 401
Reddy, A. Suresh, 77
Reddy, R. Pavan Kumar, 691
Reddy, Suresh, 67, 267
Reddy, V.S.K., 289
Remya, S., 215
Returi, Kanaka Durga, 163
Rizvi, Danish Raza, 471
Rodda, Sireesha, 593
Rongali, Srujana, 671
Roy, Kaushik, 703
Roy, Sarbani, 125, 629
Roy, Sudipta, 55

S

Sadhu, Shayak, 55
Sadhukhan, Siddharth, 55
Sadiq, Mohd., 47
Sahu, Akanksha, 1
Saluja, Akashdeep, 731

Salve, Pradip, [85](#)
 Sardesai, Milind, [85](#)
 Sarkar, Mrinal Kanti, [679](#)
 Sarkar, Subir Kumar, [679](#)
 Sarwade, Nisha, [245](#)
 Satsangi, Swanti, [773](#)
 Saudagar, Abdul Khader Jilani, [567](#)
 Sayal, Rishi, [449](#)
 Shah, Riddhi J., [487](#)
 Shaikh, Salman, [537](#)
 Sharma, Chethan, [443](#)
 Sharma, M.M., [713](#)
 Sharma, Mayank, [713](#)
 Sharma, Mugdha, [783](#)
 Sharma, Pikasha, [577](#)
 Sharma, Rajat, [651](#)
 Sharma, Yogesh, [21](#), [29](#)
 Singh, Abhishek Kumar, [497](#)
 Singh, Monika, [319](#), [329](#)
 Singh, Moutushi, [679](#)
 Sivarathinabala, M., [205](#)
 Smruthy, T.K., [639](#)
 Srinivasa Rao, P.C., [605](#)
 Sudeep, Sujay, [761](#)
 Sumalatha, D.V., [235](#)
 Sunitha, Gurram, [309](#)

Swapna, Ch. Swetha, [425](#)

T

Tarbani, Nitesh M., [103](#)
 Thirupathi, Maragani, [67](#), [77](#), [267](#)
 Thool, Ravindra C., [375](#)

U

Urooj, Shabana, [577](#)

V

Vasantrao, Kardile Vilas, [411](#)
 Vasumathi, D., [793](#)
 Vaya, Dipesh, [133](#)
 Velpula, Vijaya Bhaskar, [299](#)
 Vidhate, Pankaj Uttam, [557](#)
 Vikraman, Rashmi, [225](#)

W

Wankhade, Yogita, [557](#)

Y

Yalavarthi, Radhika, [671](#)
 Yannawar, Pravin, [85](#)