

Millie Pant

Kusum Deep

Atulya Nagar

Jagdish Chand Bansal *Editors*

Proceedings of the Third International Conference on Soft Computing for Problem Solving

SocProS 2013, Volume 2



Springer

Advances in Intelligent Systems and Computing

Volume 259

Series editor

Janusz Kacprzyk, Warsaw, Poland

For further volumes:

<http://www.springer.com/series/11156>

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

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

Hani Hagrass, 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

Millie Pant · Kusum Deep
Atulya Nagar · Jagdish Chand Bansal
Editors

Proceedings of the Third International Conference on Soft Computing for Problem Solving

SocProS 2013, Volume 2

 Springer

Editors

Millie Pant
Department of Paper Technology
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand
India

Kusum Deep
Department of Mathematics
Indian Institute of Technology Roorkee
Roorkee, Uttarakhand
India

Atulya Nagar
Department of Mathematics
and Computer Science
Liverpool Hope University
Liverpool
UK

Jagdish Chand Bansal
Department of Applied Mathematics
South Asian University
New Delhi
India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

ISBN 978-81-322-1767-1

ISBN 978-81-322-1768-8 (eBook)

DOI 10.1007/978-81-322-1768-8

Springer New Delhi Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014931408

© Springer India 2014

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. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. 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.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

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

Preface

SocProS is a 3-year old series of International Conferences held annually under the joint collaboration between a group of faculty members from IIT Roorkee, South Asian University Delhi and Liverpool Hope University, UK.

The first SocProS was held at IE(I), RLC, Roorkee, December 20–22, 2011, with General Chairs as Prof. Kusum Deep, Indian Institute of Technology Roorkee and Prof. Atulya Nagar, Liverpool Hope University, UK. It was a huge success and attracted participation from all over the world, including places like UK, US, France, South Africa, etc.

The second SocProS was held at JKLU, Jaipur, December 28–30, 2012 and was as successful as SocProS 11. Encouraged by the success of first two SocProS Conferences, this year this flagship conference—SocProS 13, which is the Third International Conference on Soft Computing for Problem Solving is being held at the Greater Noida Extension Centre of IIT Roorkee during December 26–28, 2013.

This year SocProS 13 was held as a part of the Golden Jubilee Celebrations of Saharanpur Campus of IIT Roorkee.

Like other SocProS conferences, the focus of SocProS 13 lies in Soft Computing and its applications to solve real life problems occurring in different domains ranging from medical and health care to supply chain management to image processing and cryptanalysis, etc.

SocProS 2013 attracted a wide spectrum of thought-provoking research papers on various aspects of Soft Computing with umpteen applications, theories, and techniques. A total of 158 research papers are selected for publication in the form of proceedings, which is in Volumes 1 and 2.

The editors would like to express their sincere gratitude to the Plenary Speakers, Invited Speakers, Reviewers, Programme Committee Members, International Advisory Committee, Local Organizing Committee, without whose support the quality and standards of the Conference as well as these Proceedings would not have seen the light of the day.

On the Institutional side, we would like to express our gratitude to Saharanpur Campus of Indian Institute of Technology Roorkee Campus, Roorkee, India to provide us a platform to host this Conference. Thanks are also due to the various sponsors of SocProS 2013.

We hope that the papers contained in this proceeding will prove helpful toward improving the understanding of Soft Computing at teaching as well as research level and will inspire more and more researchers to work in the field of Soft Computing.

Roorkee, India
New Delhi, India
Liverpool, UK

Millie Pant
Kusum Deep
Jagdish Chand Bansal
Atulya Nagar

About the Book

The proceedings of SocProS 2013 serve as an academic bonanza for scientists and researchers working in the field of Soft Computing. This book contains theoretical as well as practical aspects of Soft Computing, an umbrella term for techniques like fuzzy logic, neural networks and evolutionary algorithms, swarm intelligence algorithms etc.

This book will be beneficial for the young as well as experienced researchers dealing with complex and intricate real world problems for which finding a solution by traditional methods is very difficult.

The different areas covered in the proceedings are: Image Processing, Cryptanalysis, Supply Chain Management, Newly Proposed Nature Inspired Algorithms, Optimization, Problems related to Medical and Health Care, Networking etc.

Contents

A Preliminary Study on Impact of Dying of Solution on Performance of Multi-objective Genetic Algorithm	1
Rahila Patel, M. M. Raghuvanshi and Latesh Malik	
A Closed Loop Supply Chain Inventory Model for the Deteriorating Items with JIT Implementation	17
S. R. Singh and Neha Saxena	
Transient Stability Enhancement of a Multi-Machine System Using BFOA	31
M. Jagadeesh Kumar, S. S. Dash, C. Subramani, M. Arun Bhaskar and R. G. Akila	
Common Fixed Points by Using E.A. Property in Fuzzy Metric Spaces	45
Vishal Gupta and Naveen Mani	
Use of Evolutionary Algorithms to Play the Game of Checkers: Historical Developments, Challenges and Future Prospects	55
Amarjeet Singh and Kusum Deep	
Development of Computer Aided Process Planning System for Rotational Components Having Form Features	63
D. Sreeramulu, D. Lokanadham and C. S. P. Rao	
Comparative Analysis of Energy Efficient Protocols for Prolonged Life of Wireless Sensor Network	75
Gagandeep Singh, H. P. Singh and Anurag Sharma	
Condition Monitoring in Induction Motor by Parameter Estimation Technique	87
P. Kripakaran, A. Naraina and S. N. Deepa	

Word Recognition Using Barthannwin Wave Filter and Neural Network	99
Abhilasha Singh Rathor and Pawan Kumar Mishra	
Graph Coloring Problem Solution Using Modified Flocking Algorithm	113
Subarna Sinha and Suman Deb	
Augmented Human Interaction with Remote Devices Using Low Cost DTMF Technology	125
R. Singathiya, Neeraj Jangid, Prateek Gupta and Suman Deb	
Multi-Objective Ant Colony Optimization for Task Scheduling in Grid Computing	133
Nitu and Ritu Garg	
An Evaluation of Reliability of a Two-Unit Degradable Computing System Using Parametric Non-linear Programming Approach with Fuzzy Parameters	143
Kalika Patrai and Indu Uprety	
An Inventory Model with Time Dependent Demand Under Inflation and Trade Credits	155
Yogendra Kumar Rajoria, Seema Saini and S. R. Singh	
Digital Image Processing Method for the Analysis of Cholelithiasis.	167
Neha Mehta, S. V. A. V. Prasad and Leena Arya	
Theoretical Study on Vibration of Skew Plate Under Thermal Condition	173
Anupam Khanna and Pratibha Arora	
Efficient Approach for Reconstruction of Convex Binary Images Branch and Bound Method	183
Shiv Kumar Verma, Tanuja Shrivastava and Divyesh Patel	
Construction of Enhanced Sentiment Sensitive Thesaurus for Cross Domain Sentiment Classification Using Wiktionary	195
P. Sanju and T. T. Mirnalinee	
Comparative Analysis of Energy Aware Protocols in Wireless Sensor Network Using Fuzzy Logic	207
Rajeev Arya and S. C. Sharma	

Domain Specific Search Engine Based on Semantic Web 217
 Shruti Kohli and Sonam Arora

1-Error Linear Complexity Test for Binary Sequences 225
 Hetal G. Borisagar, Prasanna R. Mishra and Navneet Gaba

Neural Method for Site-Specific Yield Prediction 237
 Pramod Kumar Meena, Mahesh Kumar Hardaha, Deepak Khare
 and Arun Mondal

**Computational Intelligence Methods for Military Target
 Recognition in IR Images** 247
 Jai Prakash Singh

**Flash Webpage Segmentation Based on Image Perception
 Using DWT and Morphological Operations** 253
 A. Krishna Murthy, K. S. Raghunandan and S. Suresha

Model Order Reduction of Time Interval System: A Survey 265
 Mahendra Kumar, Aman and Siyaram Yadav

**Design and Implementation of Fuzzy Logic Rule Based
 System for Multimodal Travelling Network** 279
 Madhavi Sharma, Jitendra Kumar Gupta and Archana Lala

**An Integrated Production Model in Fuzzy Environment
 Under Inflation** 289
 S. R. Singh and Swati Sharma

**R&D and Performance Analysis of Indian Pharmaceutical
 Firms: An Application of DEA** 303
 Varun Mahajan, D. K. Nauriyal and S. P. Singh

**An Imperfect Production Model with Preservation Technology
 and Inflation Under the Fuzzy Environment.** 315
 S. R. Singh and Shalini Jain

**Assessing the Soft Computing Approach for Problem Solving
 in the Multi Criteria Decision Making for Performance
 Assessment in Indian Institutions Through
 Case Study Approach** 325
 Pooja Tripathi, Jayanthi Ranjan and Tarun Pandeya

Implementing Breath to Improve Response of Gas Sensors for Leak Detection in Plume Tracker Robots	337
Ata Jahangir Moshayedi and Damayanti Gharpure	
Use of Simulation and Intelligence Based Optimization Approach in Bioprocess.	349
Pavan Kumar and Sanjoy Ghosh	
Landuse Change Prediction and Its Impact on Surface Run-off Using Fuzzy C-Mean, Markov Chain and Curve Number Methods	365
Arun Mondal, Deepak Khare, Sananda Kundu, P. K. Mishra and P. K. Meena	
A LabVIEW Based Data Acquisition System for Electrical Impedance Tomography (EIT).	377
Tushar Kanti Bera and J. Nagaraju	
Crop Identification by Fuzzy C-Mean in Ravi Season Using Multi-Spectral Temporal Images	391
Sananda Kundu, Deepak Khare, Arun Mondal and P. K. Mishra	
Small World Particle Swarm Optimizer for Data Clustering.	403
Megha Vora and T. T. Mirnalinee	
Fast Marching Method to Study Traveltime Responses of Three Dimensional Numerical Models of Subsurface	411
C. Kumbhakar, A. Joshi and Pushpa Kumari	
A Heuristic for Permutation Flowshop Scheduling to Minimize Makespan	423
Deepak Gupta, Kewal Krishan Nailwal and Sameer Sharma	
A Survey on Web Information Retrieval Inside Fuzzy Framework	433
Shruti Kohli and Ankit Gupta	
Squaring Back off Based Media Access Control for Vehicular Ad-hoc Networks	447
Kamal Kant Sharma, Mukul Aggarwal and Neha Yadav	
A Study on Expressiveness of a Class of Array Token Petri Nets	457
T. Kamaraj, D. Lalitha and D. G. Thomas	

Non-dominated Sorting Particle Swarm Optimization Based Fault Section Estimation in Radial Distribution Systems. 471
 Anoop Arya, Yogendra Kumar and Manisha Dubey

Approximate Solution of Integral Equation Using Bernstein Polynomial Multiwavelets. 489
 S. Suman, Koushendra K. Singh and R. K. Pandey

A Novel Search Technique for Global Optimization 497
 Kedar Nath Das and Tapan Kumar Singh

Cryptanalysis of Geffe Generator Using Genetic Algorithm 509
 Maiya Din, Ashok K. Bhateja and Ram Ratan

Genetic Algorithm Approach for Non-self OS Process Identification. 517
 Amit Kumar and Shishir Kumar

Gbest-Artificial Bee Colony Algorithm to Solve Load Flow Problem. 529
 N. K. Garg, Shimpi Singh Jadon, Harish Sharma and D. K. Palwalia

Neural Network Based Dynamic Performance of Induction Motor Drives 539
 P. M. Menghal and A. Jaya Laxmi

A Bio-inspired Trusted Clustering for Mobile Pervasive Environment. 553
 Madhu Sharma Gaur and Bhaskar Pant

Comparison of Adaptive Social Evolution and Genetic Algorithm for Multi-objective DTLZ Toolkit 565
 Nitish Ghune, Vibhu Trivedi and Manojkumar Ramteke

A Reference Watermarking Scheme for Color Images Using Discrete Wavelet Transform and Singular Value Decomposition 577
 Ravinder Katta, Himanshu Agarwal and Balasubramanian Raman

Local Search Methods for the Winner Determination Problem in Multi-Unit Combinatorial Auctions 589
 Abdellah Rezoug and Dalila Boughaci

Strongly Biased Crossover Operator for Subgraph Selection in Edge-Set Based Graph Representation	601
Sakshi Arora and M. L. Garg	
“Color to Gray and Back” Using DST-DCT, Haar-DCT, Walsh-DCT, Hartley-DCT, Slant-DCT, Kekre-DCT Hybrid Wavelet Transforms	613
H. B. Kekre, Sudeep D. Thepade and Ratnesh N. Chaturvedi	
Music Genre Classification Using Music Information Retrieval and Self Organizing Maps	625
Abdul Nafey Ahmad, Chandra Sekhar and Abhinav Yashkar	
Implementation of a Private Cloud: A Case Study	635
Prachi Deshpande, S. C. Sharma and S. K. Peddoju	
Modified Activity of Scout Bee in ABC for Global Optimization	649
Kedar Nath Das and Biplab Chaudhur	
A Multistage Model for Defect Prediction of Software Development Life Cycle Using Fuzzy Logic	661
Harikesh Bahadur Yadav and Dilip Kumar Yadav	
Generalized Second-Order Duality for a Class of Nondifferentiable Continuous Programming Problems	673
Iqbal Husain and Santosh Kumar Srivastava	
A Low Cost Electrical Impedance Tomography (EIT) Instrumentation for Impedance Imaging of Practical Phantoms: A Laboratory Study	689
Tushar Kanti Bera and J. Nagaraju	
A Broyden’s Method Based High Speed Jacobean Matrix Calculator (JMC) for Electrical Impedance Tomography (EIT)	703
Tushar Kanti Bera, Samir Kumar Biswas, K. Rajan and J. Nagaraju	
Predicting Total Number of Failures in a Software Using NHPP Software Reliability Growth Models	715
Poonam Panwar and A. K. Lal	

Design Optimization of Shell and Tube Heat Exchanger Using Differential Evolution Algorithm 729
 Pawan Singh and Millie Pant

Compression of Printed English Characters Using Back Propagation Neural Network 741
 Sunita, Vaibhav Gupta and Bhupendra Suman

Conceptual Design of EPICS Based Implementation for ICRH DAC System 757
 Ramesh Joshi, Manoj Singh, S. V. Kulkarni and Kiran Trivedi

An Efficient Network Management and Power Saving Wake-On-LAN. 767
 Pranjal Daga, D. P. Acharjya, J. Senthil and Pranam Daga

Authentication in Cloud Computing Environment using Two Factor Authentication 779
 Brijesh Kumar Chaurasia, Awanish Shahi and Shekhar Verma

Expedited Artificial Bee Colony Algorithm. 787
 Shimpi Singh Jadon, Jagdish Chand Bansal, Ritu Tiwari and Harish Sharma

Indian Ehealth Services: A Study. 801
 Shilpa Srivastava, Milli Pant and Namrata Agarwal

Cluster Based Term Weighting Model for Web Document Clustering. 815
 B. R. Prakash, M. Hanumanthappa and M. Mamatha

Performance of Poled Organic Polymeric Films by ANOVA in Terms of SHG Conversion Efficiency 823
 Renu Tyagi, Yuvraj Singh Negi and Millie Pant

Performance Analysis of Bio-Inspired Techniques 831
 Samiksha Goel, Arpita Sharma and V. K. Panchal

On a Class of Nondifferentiable Multiobjective Continuous Programming Problems. 845
 Iqbal Husain and Vikas Kumar Jain

Three Echelon Supply Chain Design with Supplier Evaluation 867
 Kanika Gandhi, P. C. Jha, Kannan Govindan and Diego Galar

A Carbon Sensitive Multi Echelon Reverse Logistics Network Design for Product Value Recovery	883
Jyoti Dhingra Darbari, Vernika Agarwal and P. C. Jha	
Multi Period Advertising Media Selection in a Segmented Market	905
Sugandha Aggarwal, Arshia Kaul, Anshu Gupta and P. C. Jha	
Fuzzy Multi-criteria Approach for Component Based Software System Under Build-or-Buy Scheme	929
P. C. Jha, Ramandeep Kaur, Sonam Narula and Sushila Madan	
A Novel Lossless ECG Compression Technique for Transmission in GSM Networks	947
Diana Moses and C. Deisy	
Disaster Relief Operations and Continuous Aid Program in Human Supply Networks: Are they congruent?—An analysis	959
V. G. Venkatesh, Rameshwar Dubey and Sadia Samar Ali	
Graduate School Application Advisor Based on Neural Classification System	975
Devarsh Bhonde, T. Sri Kalyan and Hari Sai Krishna Kanth	
About the Editors	987
Author Index	989

A Preliminary Study on Impact of Dying of Solution on Performance of Multi-objective Genetic Algorithm

Rahila Patel, M. M. Raghuwanshi and Latesh Malik

Abstract Genetic Algorithm (GA) mimics natural evolutionary process. Since dying of an organism is important part of natural evolutionary process, GA should have some mechanism for dying of solutions just like GA have crossover operator for birth of solutions. In nature, occurrence of event of dying of an organism has some reasons like aging, disease, malnutrition and so on. In this work we propose three strategies of dying or removal of solution from next generation population. Multi-objective Genetic Algorithm (MOGA) takes decision of removal of solution, based on one of these three strategies. Experiments were performed to show impact of dying of solutions and dying strategies on the performance of MOGA.

Keywords Multi-objective genetic algorithm (MOGA) · Diversity · Convergence · Dying of solutions · Dying strategies

1 Introduction

In biology and ecology, death is the end of an organism. Contemporary evolutionary theory sees death as an important part of the process of natural selection. It is considered that organisms less adapted to their environment are more likely to die having produced fewer offspring, thereby reducing their contribution to the - gene pool. The gene pool of a species or a population is the variety of genetic information in its living members. A large gene pool (extensive genetic diversity)

R. Patel (✉) · L. Malik
G.H. Rasoni College of Engineering, Nagpur, India
e-mail: rahila.patel@gmail.com

L. Malik
e-mail: latesh.malik@ghrce.edu.in

M. M. Raghuwanshi
Rajiv Gandhi College of Engineering and Research, Nagpur, India
e-mail: m_raghuwanshi@rediffmail.com

is associated with robust populations that can survive bouts of intense selection. Meanwhile, low genetic diversity reduces the range of adaption possible. Replacing native with alien genes narrows genetic diversity within the original population, thereby increasing the chance of extinction [1, 2].

Algorithms based on strategies of evolution are called as evolutionary algorithms. Genetic Algorithm (GA) [3] is one such algorithm based on natural genetics and survival of fittest. Genetic algorithms are used to solve single-objective and multi-objective optimization problems [4]. Genetic algorithm used for solving single-objective problem are simpleminded and return a single optimal solution whereas for solving multi-objective problem, genetic algorithm has to select appropriate solutions. The decision of selection of solution becomes complicated in presence of multiple conflicting objectives. Every multi-objective evolutionary algorithm has two goals; convergence and diversity, so they need two different mechanisms for fulfillment of these goals. Algorithms like NSGA-II [5] use non-dominated sorting and crowding distance based selection strategies for convergence and diversity. Many such multi-objective evolutionary algorithms with explicit mechanism for convergence and diversity control are found in the literature [5–10].

Genetic algorithm uses selection, crossover and mutation operators to evolve a set of solutions of current generation. Selection operator is then used to select fit solutions from current generation as next generation population [3]. Weaker Solutions which are not selected die out and do not reappear. Thus dying is implicit part of selection mechanism. Solutions having better fitness produce fitter offspring and selection strategies are likely to select both parent and their offspring. Solutions (parent solution) present in first generation inherit their properties to offspring solutions by using cross over operator. In subsequent generations good properties of parent solutions are carrying forward by offspring solutions generated by crossover operator. Problem with this selection strategy is that, after few generation whole population is dominated by presence of few solutions from initial population and their offspring i.e. trail of very few solutions from initial population reach to final generation and most of the solutions die out somewhere in between generations. Thus selection strategy based on survival of fittest reduces diversity and convergence of solutions. So, a new mechanism is needed to avoid excessive dying of solutions.

In this work dying of solution has been made explicit part of evolutionary process and three strategies for dying of solution or parent removal (ParRem) have been proposed and implemented. A solution is deterministically removed from next generation population by using one of the three proposed dying strategies. Impact of dying rate of solutions on the performance of GA has been studied. Idea of gradual dying is modeled and a new framework of MOGA has been used to demonstrate the same.

Rest of the paper is organized as [Sect. 2](#) presents strategies for dying of solution [Sect. 3](#) proposes new framework for MOGA [Sect. 4](#) covers experimental setup and results and [Sect. 5](#) compares strategies of dying with other algorithm. Finally paper concludes with few conclusions and future work.

2 Strategies for Dying of Solution

A thought; dying of parent solution; opposite of selection of parent solution, is materialized in this work. Proposed three strategies of dying are given below.

ParRem1: Remove solution having minimum distance from one or more solutions of next generation population. In this strategy before removing a solution from next generation population distance between all the solutions is checked. Since similarity means uniformity and dissimilarity means diversity, according to this strategy similar solution should die out and dissimilar solution should remain in population in order to have diversity in population. The solution having minimum distance from its neighbor solution will be removed from next generation population. Distance between solutions is calculated in objective space and distance measure used is Euclidean distance. Pseudo code for distance calculation is given below.

Initialize min_d = 9999,

for j = 1 : N-1.

for i = j + 1 : N

*d = distance between first parent(i^{th}) and second parent (j^{th})
parent*

if (d < min_d)

min_d = d;

idx = i;

end if

end for

end for

where N is population size, min_d is minimum distance, idx is index of parent having minimum distance. In NSGA-II for crowding distance calculation all the solutions in the front are sorted first and then crowding distance is calculated among solution and its two neighboring solutions. Here in ParRem1 sorting is not required. Distance from every solution to all the solutions in the population is calculated and then the solution having minimum distance with any of the solution will be removed.

ParRem2: Remove solution having maximum Summation of Normalized Objective Value (SNOV) [10]. In this strategy to make effective range of all the objective functions equal objective values are normalized. After normalization effective range of all the objective function will be zero to one. Assign SNOV to each solution in next generation population. The SNOV will be treated as single scalar fitness of the solution. Now remove solution having highest SNOV value (for minimization of objective). Pseudo code for SNOV calculation is given below.

for $m = 1 : M$ (number of objectives)

Find the maximum and minimum objective values of the m th objective and calculate the range of the m th objective.

Normalize the m th objective values of all members using the equation:

$$f_m(x) = \frac{f_m(x) - f_{\min}}{f_{\max} - f_{\min}}$$

where f_m is the normalized m th objective value.

end for

for $i = 1 : N$

Sum all normalized objective values of the member to obtain a single value.

end for

ParRem3: Remove solution having poor fertility count. Frequency of reproduction is an important parameter in determining species survival: an organism that dies young but leaves numerous offspring displays, according to Darwinian criteria, much greater fitness than a long-lived organism leaving only one [2]. In this strategy, parameter fertility_count keep record of frequency of reproduction of a parent solution. Initially, zero fertility_count is assigned to all the solution in the population. If a parent solution produces offspring better than the parent solution then fertility_count of parent solution is incremented by one and is assigned to better offspring solution. From next generation population, a solution having minimum fertility_count will be removed. Pseudo code is given below.

for $i = 1$ to N

fertility_count(i) = 0;

end for

for $gen = 1$ to $maxgen$

for $i = 1$ to N

If offspring solution wins the multi-level tournament then fertility_count of first parent solution is incremented by 1 and assigned as fertility_count to winner offspring solution.

end for

Remove solution having lowest fertility_count

...

End for

3 Proposed Algorithm

In this section a new Multi-Objective Genetic Algorithm (MOGA) framework is presented which uses simple mechanism for formation of next generation population. Uniform-random method is used for population initialization. Initial population is tuned by using opposition based learning [11]. In the beginning each initial solution is given a unique number as parent number. For crossover operation first parent is sequentially selected and second parent is selected randomly from current population. Crossover operator generates two offspring that form family with first parent. Multi-level tournament operator is used to select one solution from the family. Parent number of first parent is assigned to offspring solution. Cycles of selection-crossover-multilevel tournament selection repeats with each solution in current population. Algorithm decides the generation number in which a parent should be removed by using the formula given below:

$$\text{Gen_Rem} = \text{max gen}/(\text{N} * \text{DR})$$

where max gen is Maximum number of generation, DR is dying rate in % (number of solutions to be removed from population) and N is size of population. For example If max gen = 2,000, DR = 30 % and N = 200 then $\text{Gen_Rem} = 2,000/(200 * 0.3) = 34$. In 34th generation, one parent will be removed on the basis of one of the three strategies of dying, from next generation population and an offspring solution from offspring pool will be inserted in next generation population. This repeats after every 34th generation. At the end of max gen generation final population contains traces of 140 solution and 60 (=200 * 0.3) solutions are removed from population i.e. 70 % parent solutions survive and 30 % parent solutions die out. Algorithmic steps of proposed framework are given below.

Initialization:

Set the parameter values for max gen, DR, N, Probability distribution index for MPX crossover operator [12].

Processing:

1. Generate N solutions randomly then tune them using OBL and keep tuned solutions in *POP*
2. Assign number 1,2,...,N as parent number sequentially to each solution in *POP*
3. Calculate $\text{Gen_Rem} = (\text{max gen}/(\text{N} * \text{DR}))$
4. $i\text{Gen_Rem} = \text{Gen_Rem}$
5. for gen = 1: max gen

*Set off_pool and Next_gen to be the empty set
for i = 1:N*

- For reproduction select first solution sequentially and second solution randomly from *POP*

- Now apply MPX operator on the selected solutions and generate two offspring. Assign parent number of first parent to both the offspring.
- Using Multi-level tournament selection select best among first parent and offspring solutions
- Copy best solution in *Next_gen* and put the offspring which is not selected in the *off_pool*

```

end for
if gen == Gen_Rem
    • Use dying strategy to remove one solution from Next_gen and insert one solution, from off_pool, whose parent number is not same as the removed solution.
    • Gen_Rem = Gen_Rem + iGen_Rem
    • POP = Next_gen
else
    POP = Next_gen
end ifelse
end for
End

```

Initialization and tuning the population: H. R. Tizhoosh introduced the opposition concept through Opposition-Based Learning. In this concept, the opposite of a guess or estimate (i.e., uniform-random) is calculated and compared to the original estimate (random). In 2006, Rahnamayan et al. applied the opposition concept to initialize population in evolutionary Algorithms. This supports more diversity and exploration when starting the search process [11]. In the proposed algorithm population tuning is done by using opposition based learning. A solution is sequentially selected from initial population and its opposite solution is calculated by using opposition based learning as given above. Now tournament is played between solution and opposite solution and winner is selected as tuned solution and loser is discarded. This procedure is repeated for every solution.

Tournament selection: In this algorithm multi-level tournament selection is used. In first level of tournament selection, tournament is played between two offspring solutions. In second level, tournament is played between winner offspring solution and First parent. The winner of second level of tournament is selected as best solution and becomes member of next generation population.

4 Experimental Setup and Results

The proposed framework for MOGA with three strategies for dying of solution is coded in MatLab 7.1. 30 independent runs have been taken for each problem with different strategies. Experimental parameter settings used are:

Table 1 Performance of ParRem1 on UF1-UF9 (N = 100 and max gen = 3,000)

Fun	UF1	UF2	UF3	UF4	UF5	UF6	UF7	UF8	UF9
DR	MeanIGD								
10	0.0673	0.0232	0.2011	0.0502	0.1538	0.0649	0.0834	0.8646	0.9094
30	0.0680	0.0241	0.2100	0.0591	0.1838	0.0669	0.1038	0.7743	0.8999
40	0.0622	0.0240	0.2331	0.0620	0.1938	0.0690	0.1036	0.7604	0.9001
50	0.0616	0.0192	0.2400	0.0623	0.2142	0.0721	0.0983	0.7553	0.9144
60	0.0699	0.0215	0.2467	0.0635	0.2292	0.0784	0.0990	0.7743	0.9802
90	0.0639	0.0323	0.2567	0.0643	0.2468	0.0821	0.1339	0.8359	0.9404
	SPREAD								
10	0.6785	0.4404	0.5266	0.6879	0.7962	1.0028	0.7821	0.9706	0.7191
30	0.6190	0.4685	0.5384	0.6118	0.7909	0.9134	0.7460	0.9180	0.7154
40	0.5984	0.4348	0.5654	0.5692	0.7904	0.8043	0.7261	0.9034	0.6914
50	0.5801	0.4081	0.5474	0.5046	0.6787	0.7006	0.6703	0.4640	0.6308
60	0.5998	0.4543	0.5224	0.5752	0.7463	0.7506	0.8066	0.6706	0.6574
90	0.6369	0.4318	0.5634	0.5619	0.7373	0.8442	0.8731	0.7860	0.6946

- Population size (N) = 100
- Maximum no. of generation (max gen) = 3,000
- Dying rate (DR) = 10, 30, 40, 50, 60, 90 %
- Probability Distribution index (μ) for MPX crossover operator = 1
- Number of Parents = 2
- Number of offspring = 2.

Test functions: The test problems (UF1-UF9) in this work are taken from IEEE CEC2009 special session and competition [15]. Functions UF1-UF7 has two objectives and UF8 & UF9 are three objective functions.

Metric used for Performance measure:

- Inverted Generational Distance (IGD): The Inverted Generational Distance (IGD) metric is used as performance indicator to quantify the quality of the obtained results. The IGD metric measures “how well is the Pareto-optimal front represented by the obtained solution set”. [13]
- SPREAD (Δ): This metric is used to measure diversity among solution. It takes care of uniform distribution and extent of distribution of obtained solution [5].

Results obtained by MOGA with three strategies of dying of solution are given as:

ParRem1: The Performance of dying strategy ParRem1 in terms of MeanIGD & SPREAD on function UF1-UF9 is shown in Table 1. For functions UF1, UF2 & UF8 best MeanIGD value is obtained for 50 % dying rate. For rest of the functions best MeanIGD is for low dying rate. It is observed that a moderate dying rate (50 %) has given best spread for all functions except UF3 (60 %).

ParRem2: The Performance of dying strategy ParRem2 in terms of MeanIGD & SPREAD on function UF1-UF9 is shown in Table 2. For functions UF2 & UF8

Table 2 Performance of ParRem2 on UF1-UF9 (N = 100 and max gen = 3,000)

Fun	UF1	UF2	UF3	UF4	UF5	UF6	UF7	UF8	UF9
DR	MeanIGD								
10	0.0607	0.0243	0.1702	0.0543	0.2081	0.0704	0.1232	0.8408	0.9045
30	0.0621	0.0256	0.1882	0.0543	0.2180	0.0714	0.1138	0.8294	0.9193
40	0.0638	0.0269	0.2076	0.0669	0.2199	0.0710	0.1589	0.8041	0.9297
50	0.0662	0.0238	0.2001	0.0694	0.2284	0.0725	0.1439	0.7389	0.9344
60	0.0671	0.0251	0.2422	0.0713	0.2412	0.0734	0.1590	0.7608	0.9649
90	0.0650	0.0269	0.2512	0.0709	0.3099	0.0740	0.1611	0.7760	1.0890
	SPREAD								
10	0.6227	0.4858	0.6106	0.7156	0.8611	0.8263	0.8827	0.7238	0.6871
30	0.6043	0.4601	0.5137	0.6698	0.8512	0.8163	0.8767	0.7108	0.5891
40	0.6043	0.4680	0.5937	0.6037	0.8100	0.8263	0.8785	0.7038	0.5891
50	0.5593	0.4503	0.4709	0.5348	0.7842	0.7688	0.7903	0.6806	0.6423
60	0.5613	0.4639	0.6136	0.7559	0.8024	0.8454	0.8066	0.7381	0.6806
90	0.5600	0.4843	0.5546	0.6886	0.8234	0.9547	0.9371	0.7313	0.6806

Table 3 Performance of ParRem3 on UF1-UF9 (N = 100 and max gen = 3,000)

Fun	UF1	UF2	UF3	UF4	UF5	UF6	UF7	UF8	UF9
DR	MeanIGD								
10	0.0676	0.0223	0.1801	0.0540	0.2012	0.0596	0.1102	0.7773	0.9685
30	0.0689	0.0241	0.1892	0.0601	0.2017	0.0592	0.1138	0.7386	0.9596
40	0.0689	0.0248	0.2098	0.0648	0.2012	0.0609	0.1380	0.7567	0.9610
50	0.0691	0.0257	0.2101	0.0693	0.2344	0.0642	0.1594	0.8366	0.9656
60	0.0704	0.0261	0.2301	0.0631	0.2389	0.0662	0.1619	0.9225	0.9676
90	0.0748	0.0273	0.2599	0.0641	0.2478	0.0672	0.1711	0.9324	0.9672
	SPREAD								
10	0.5810	0.4604	0.5417	0.7125	0.9315	0.8229	0.8072	0.6502	0.6820
30	0.6135	0.4270	0.5001	0.5524	0.9001	0.8209	0.8240	0.6840	0.7231
40	0.5652	0.4310	0.5044	0.6328	0.8213	0.8023	0.8560	0.6986	0.7316
50	0.5565	0.4307	0.5470	0.6310	0.6387	0.6796	0.7043	0.8324	0.6345
60	0.6745	0.4525	0.5024	0.5617	0.6897	0.9221	0.7601	0.8025	0.7214
90	0.5900	0.4584	0.4769	0.5765	0.8010	0.9640	0.7930	0.7818	0.7334

best MeanIGD value is obtained at 50 % dying rate. For rest of the functions best MeanIGD is for low dying rate. It is also observed that a moderate DR (50 %) has given best spread for all functions except UF9 (40 %).

ParRem3: The Performance of dying strategy ParRem3 in terms of MeanIGD & SPREAD on function UF1-UF9 is shown in Table 3. The best MeanIGD value for all the function is obtained at low dying rate. Also as dying rate increases it increases MeanIGD value that reflects the degradation of performance of algorithm. Best SPREAD on UF1 & UF5-UF7 is for dying rate = 50 %, function UF2-UF4 is for dying rate = 30 and on function UF8 & UF9 is for dying rate = 10 %.

Table 4 shows best performance of the three dying strategies. It is observed that for all three strategies on most of the functions best results for SPREAD have been

Table 4 Performance comparison of ParRem1, ParRem2 and ParRem3 on Function UF1-UF9

Scheme Function	ParRem1		ParRem2		ParRem3		Sign test (IGD, SPRD)
	MeanIGD (DR)	SPREAD (DR)	MeanIGD (DR)	SPREAD (DR)	MeanIGD (DR)	SPREAD (DR)	
UF1	0.06165 (50 %)	0.58012 (50 %)	0.060779 (10 %)	0.55635 (50 %)	0.067689 (10 %)	0.55652 (50 %)	(-, -)
UF2	0.019220 (50 %)	0.40815 (50 %)	0.023877 (50 %)	0.45033 (50 %)	0.022364 (10 %)	0.42707 (30 %)	(+, +)
UF3	0.201176 (10 %)	0.52668 (10 %)	0.170225 (10 %)	0.47090 (50 %)	0.180125 (10 %)	0.50012 (30 %)	(+, +)
UF4	0.050215 (10 %)	0.50465 (50 %)	0.0543763 (10 %)	0.53480 (50 %)	0.054034 (10 %)	0.55244 (30 %)	(-, +)
UF5	0.153826 (10 %)	0.67871 (50 %)	0.208118 (10 %)	0.78428 (50 %)	0.201256 (10 %)	0.63878 (50 %)	(+, +)
UF6	0.064901 (10 %)	0.70065 (50 %)	0.0704405 (10 %)	0.76884 (50 %)	0.059221 (30 %)	0.67963 (50 %)	(+, +)
UF7	0.083477 (10 %)	0.67037 (50 %)	0.113827 (30 %)	0.79031 (50 %)	0.110274 (10 %)	0.70431 (50 %)	(+, +)
UF8	0.755344 (50 %)	0.46400 (50 %)	0.738952 (50 %)	0.68068 (50 %)	0.738637 (30 %)	0.65029 (10 %)	(-, +)
UF9	0.899907 (30 %)	0.63089 (50 %)	0.9045648 (10 %)	0.58913 (40 %)	0.959632 (30 %)	0.68203 (10 %)	(+, +)

Table 5 Statistical sum of problems for which each strategy obtains significantly better results

Metric	ParRem1	ParRem2	ParRem3
MeanIGD	04	01	01
SPREAD	04	02	02
Total	08	03	03

Table 6 Performance comparison of MOGA-ParRem1 with MOGA-WD and NSGA-II-MPX

Scheme	MOGA-WD		MOGA-ParRem1 (DR = 50 %)		NSGA-II-MPX		Sign test (MeanIGD SPREAD)
	MeanIGD	SPREAD	MeanIGD	SPREAD	MeanIGD	SPREAD	
UF1	0.061892	0.615871	0.061657	0.580122	0.068187	0.697345	(-, +)
UF2	0.057764	0.644613	0.019220	0.408153	0.024311	0.540971	(+, +)
UF3	0.245993	0.583918	0.240011	0.547451	0.135773	0.666309	(+, +)
UF4	0.066059	0.550369	0.062340	0.504653	0.019737	0.682081	(+, +)
UF5	1.046338	0.753780	0.214201	0.678715	0.211415	0.658579	(+, -)
UF6	0.230648	0.695138	0.072117	0.700656	0.152526	0.694642	(+, -)
UF7	0.101259	0.715222	0.098397	0.670374	0.031356	0.626167	(+, +)
UF8	0.836721	0.686384	0.755344	0.464006	0.536727	0.616086	(+, +)
UF9	0.975251	0.883306	0.914492	0.630897	0.375251	0.703306	(+, +)

obtained for dying rate at or below 50 %. Hence we can say that excessive dying of solutions affects diversity in Pareto front. It is also observed that dying strategy ParRem1 with dying rate 50 % has outperformed ParRem2 and ParRem3 in terms of SPREAD on four functions (UF2, UF4, UF7 and UF8) out of nine functions. ParRem2 has outperformed ParRem1 and ParRem3 on three functions (UF1 & UF3 with dying rate 50 % and UF9 with dying rate 40 %). ParRem3 has outperformed ParRem1 and ParRem2 on two functions (UF5 and UF6 with dying rate 50 %). Table 5 shows Statistical Sum of Problems for which each strategy obtains significantly better results. Thus dying strategy ParRem1 is the best among the three dying strategies and has given better convergence and maintained diversity among solutions.

5 Comparison with Other Algorithms

The performance of proposed MOGA with dying strategy ParRem1 (MOGA-ParRem1) is compared with proposed MOGA without dying (MOGA-WD) and NSGA-II-MPX. NSGA-II is coded in MatLab 7.1 and in place of SBX operator MPX crossover operator is used. Also parameter setting used for MOGA-WD and NSGA-II-MPX is same as given in Sect. 4. Table 6 shows MeanIGD and

Table 7 Statistical sum of problems for which each strategy obtains significantly better results

Metric	MOGA-WD	MOGA-ParRem1 (DR = 50 %)	NSGA-II-MPX
MeanIGD	0	02	06
SPREAD	0	06	01
Total	0	08	07

Fig. 1 Best approximate of UF2 with MOGA-WD

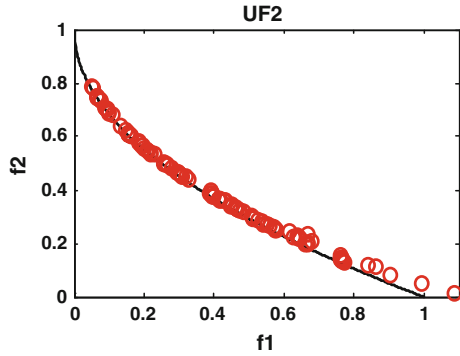


Fig. 2 Best approximate of UF2 with MOGA-ParRem1

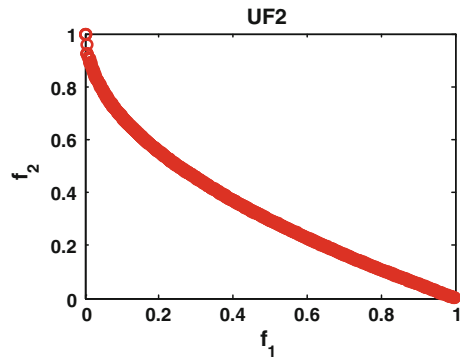


Fig. 3 Best approximate of UF2 with NSGA-II-MPX

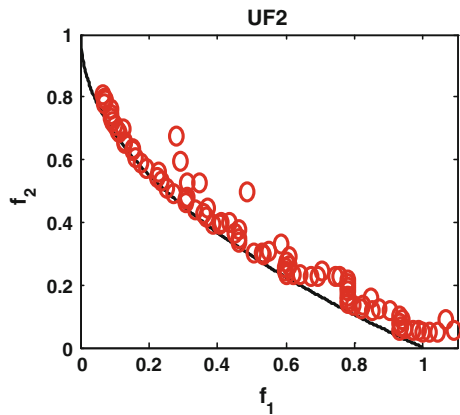


Fig. 4 Best approximate of UF4 with MOGA-WD

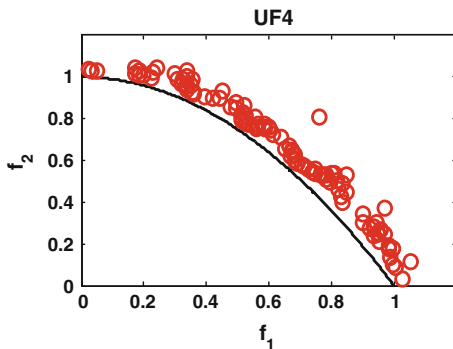


Fig. 5 Best approximate of UF4 with MOGA-ParRem1

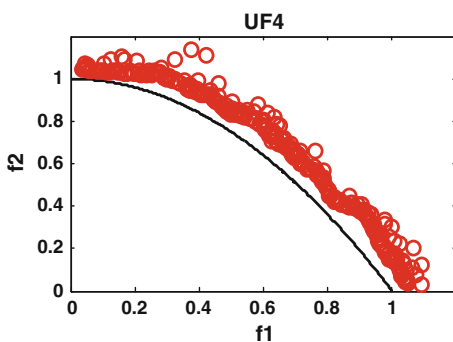
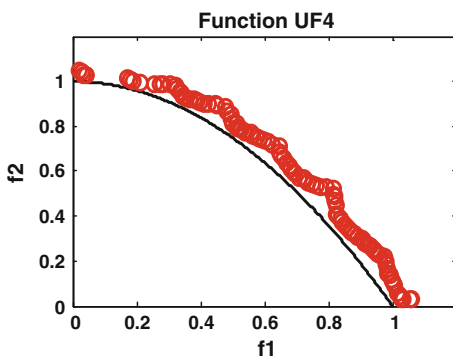


Fig. 6 Best approximate of UF4 with NSGA-II-MPX



SPREAD of MOGA-WD, MOGA-ParRem1 and NSGA-II-MPX. SPREAD values reported in the Table 6 indicate that MOGA-ParRem1 has outperformed MOGA-WD on functions UF1-UF5 and UF7-UF9. Also MOGA-ParRem1 has outperformed NSGA-II-MPX by giving better SPREAD on functions UF1-UF4 and UF8-UF9. IGD measure taken for MOGA-ParRem1 shows that it has given good performance on functions UF1, UF2 and UF6 functions. NSGA-II-MPX has

Fig. 7 Best approximate of UF7 with MOGA-WD

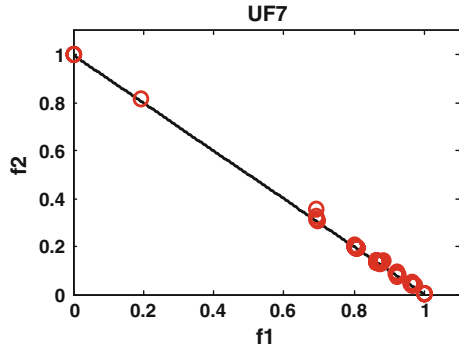


Fig. 8 Best approximate of UF7 with MOGA-ParRem1

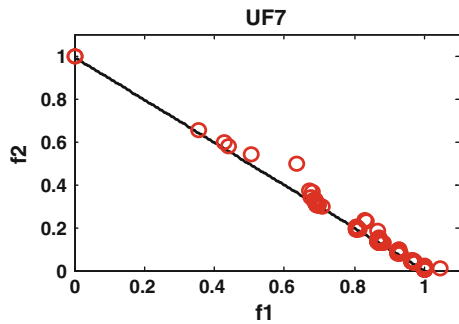
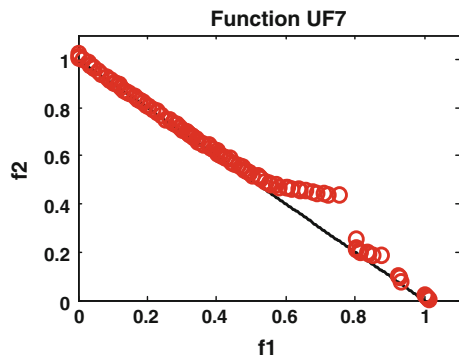


Fig. 9 Best approximate of UF7 with NSGA-II-MPX



shown good IGD values on functions UF3-UF5 and UF7-UF9. Table 7 shows Statistical Sum of Problems for which each algorithm obtains significantly better results.

Figures 1, 2, 3, 4, 5, 6, 7, 8, 9 shows plots of non dominated set of solutions having best MeanIGD obtained on functions UF2, UF4 and UF7 with MOGA-WD, MOGA-ParRem1 and NSGA-II-MPX. From Figs. 2 and 5 it is clear that MOGA-ParRem1 has given better diversity and convergence than MOGA-WD and NSGA-II-MPX on

test functions UF2 & UF4. Poor performance of MOGA-ParRem1 is seen in Fig. 8 on function UF7. MOGA-WD has shown worst performance on all three test functions as shown in Figs. 1, 4 and 7.

6 Conclusion

In this work GA procedure is made more close to natural evolutionary process which incorporate Birth-Reproduction-death cycle. Dying mechanism is made explicit part of GA. We have proposed and implemented strategies for dying of solution in specific generations. Experimental results indicate that the proposed MOGA with three strategies of dying of solutions is able to guide the search process towards the optimum for the seven bi-objective and the two 3-objective test functions. Among the three strategies of dying of solutions, strategy ParRem1 has shown best performance in terms of better diversity among solutions. Also, as dying rate increases diversity among solution increases but diversity deteriorates when dying rate exceeds 50 %. So we can conclude that a moderate dying rate (around 50 %) gives the better spread.

When MOGA-ParRem1 compared with MOGA-WD it is found that MOGA-ParRem1 has outperformed MOGA-WD on all functions. MOGA-ParRem1 has also outperformed NSGA-II-MPX on many functions. Thus we have successfully implemented a simple and efficient MOGA and proved that excessive dying of solution affects performance of GA. A controlled dying of solution can improve performance of genetic algorithm.

Future work will be on fine tuning the three strategies of dying and design of new algorithm which adaptively selects the dying strategy to solve MOPs having complex Pareto fronts and Pareto sets.

References

1. en.wikipedia.org/wiki/Death
2. en.wikipedia.org/wiki/Extinction
3. Goldberg, D.E.: Genetic Algorithms in Search, Optimisation and Machine Learning. Addison Wesley, Reading (1989)
4. Deb, K.: Multi-objective Optimization using Evolutionary Algorithms. Wiley, West Sussex (2001)
5. Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast elitist nondominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Proceedings of 6th International Conference, PPSN VI, LNCS, vol. 1917, Paris, France, pp. 849–858 (2000)
6. Corne, D., Knowles, J., Oates, M.: The Pareto envelope-based selection algorithm for multi-objective optimization. In: Proceedings of International Conference on PPSN VI, LNCS, vol. 1917, pp. 839–848 (2000)
7. Horn, J., Nafpliotis, N., Goldberg, D.E.: A niched Pareto genetic algorithm for multi-objective optimization. In: Proceedings of the First IEEE CEC, USA, pp. 82–87 (1994)

8. Knowles, J., Corne, D.: The Pareto archived evolution strategy: a new baseline algorithm for Pareto multi-objective optimisation. In: Proceedings of CEC99, USA, pp. 98–105 (1999)
9. Schaffer, J.D.: Multiple objective optimization with vector evaluated genetic algorithms. In: Proceedings of First International Conference on Genetic Algorithms and Their Applications, pp. 93–100 (1985)
10. Patel, R., Raghuvanshi, M., Malik L.: An improved ranking scheme for selection of parents in multi-objective genetic algorithm. In: Proceedings of IEEE International Conference on CSNT 2011, SMVDU (J&K), pp. 734–739 (2011)
11. Al-Qunaieer, F.S., Tizhoosh, H.R., Rahnamayan, S.: Opposition based computing—a survey. In: Proceedings of IEEE Transaction on Evolutionary Computation (2010)
12. Raghuvanshi, M., Kakde, O.: Multi-parent recombination operator with polynomial or lognormal distribution for real coded genetic algorithm. In: Proceedings of 2nd Indian International Conference on Artificial Intelligence (IICAI), pp. 3274–3290 (2005)
13. Zhang, Q., Zhou, A., Zhao, S.Z., Suganthan, P.N., Liu, W., Tiwari, S.: Multi-objective optimization test instances for the CEC 2009 special session and competition. Technical report, Nanyang Technological University, Singapore (2008)

A Closed Loop Supply Chain Inventory Model for the Deteriorating Items with JIT Implementation

S. R. Singh and Neha Saxena

Abstract In the past recent years, a growing environmental consciousness has been shaping the way society looks on green. Society's attitude towards environmental issues has been changing and hence recoverable product environments are becoming an increasingly important segment of the overall push in industry. We have proposed a model for the design of a closed loop supply chain with green components. We have investigated a joint economic production quantity model for a single vendor, single buyer system considering lot-splitting. The effect of deterioration is taken into consideration. Here we have assumed that the vendor fulfils the buyer's demand with the produced and remanufactured units, where the remanufactured items are considered as good as those of new items. Mathematical and numerical analysis are presented to describe the situation.

Keywords Inventory model · Production · Remanufacturing · Deterioration · JIT implementation

1 Introduction

Owing to increased public concern about the environment, most developed countries have made legislations, mandating manufacturers and importers to take back used products at the end of their useful lives. Consumers can now return goods within warranty period as part of the after-sales service if the products fail to meet their needs or when the products have reached the end of their useful lives.

S. R. Singh (✉) · N. Saxena
Department of Mathematics, D. N. College, Meerut, India
e-mail: shivrajpundir@gmail.com

N. Saxena
e-mail: nancineha.saxena@gmail.com

The returned products may then be refurbished or remanufactured to extend their periods of usage or recycled to recapture value. For sustainable development, a good supply chain strategy enables manufacturing plant to rescue and recover many parts and components from used products through reverse logistics activities of remanufacturing and reuse. It considers impacts of environmental protection and green image. Closed-loop supply chains comprising forward and reverse logistics can be combined to achieve more sustainable production and consumption.

There have been numerous studies and research on reverse logistics. The initial approach in the field of reverse logistics is made by Schrady [18]. He analyzed a traditional Economic Order Quantity (EOQ) model for repairable items with the joint coordination of reverse manufacturing with the forward supply chain by assuming that the manufacturing and recovery (repair) rates are instantaneous. His work was generalized by Nahmias and Rivera [12] with finite repair rates. This model is further extended by Koh et al. (2002) for limited repair capacity. Richter [13–15], and Richter and Dobos [16] have investigated the reverse logistics model. In these problems the return rate has been taken as a decision variable. Authors have examined problem with pure or bang-bang policy (total repair or total wastage disposal). Richter [15] concluded in his paper that the bang-bang policy is optimal when compared to mixed (production + remanufacturing) policy, while in later study (El Saadany and Jaber [5] it was observed that mixed strategy is optimal, rather than a pure strategy; either pure remanufacturing or pure production. The model developed by Ishii et al. [6] confirms that life-cycle design seeks to maximize the life-cycle value of a product at the initial stages of design. Savaskan et al. [17] developed a RL model by assuming the returned rate dependent to the Collection investment. In this model they investigated a closed-loop supply chain that includes reverse logistics in times of product recovery of the retailer. Dekker et al. [4] proposed a quantitative model for closed loop supply chain. He analyzed that the amount of returns is highly uncertain and this uncertainty greatly affect the collection and inventory decisions. El Saadany and Jaber [5] developed a model assuming that the collection rate of returned items is dependent on the purchasing price and the acceptance quality level of these returns. Konstantaras and Skouri [8] developed a model of supply chain by considering a general cycle pattern in which a variable number of reproduction lots of equal size are followed by a variable number of manufacturing lots of equal size. They also have studied the case where shortages are allowed in each manufacturing and reproduction cycle. Alamri [1] investigate the optimal returned quantity for a general Reverse Logistics inventory model for the deteriorated items. Green supply chain inventory model with short life cycle product is developed by Chung and Wee [3]. Singh and Saxena [20] investigated a reverse logistics inventory model with time dependent rates under shortages considering the returned rate and holding capacity as a decision variable. Along the same line as Singh and Saxena [20], Singh et al. [19] developed there model for the flexible manufacturing under the stock out situation.

With growing focus on supply chain management firms realize that inventories across the entire supply chain can be more sufficiently managed through great cooperation and better coordination in just in time production system, multiple deliveries is the one way to reduce the inventory level and the total cost. The main goal of this study is to analyse under what conditions (capacity, return rate, remanufacturing rate) a reverse-logistics system should be introduced at a company that uses a JIT production-management system. Numerous researches have been done on the integrated inventory models implementing JIT. Lu [11] proposed a lot splitting model implementing JIT deliveries assuming equal sized multiple shipments. A joint economic lot size model for a single vendor single buyer system has been developed by Benerji [2]. Kim and Ha [9] have developed a lot splitting model and discussed how and when the optimal policy for buyer and supplier can be achieved. Chang and Wee [3] have investigated a green supply chain model implementing JIT delivery.

In this paper we determined the coordination of reverse manufacturing with the forward supply chain in the inventory management. Reverse logistics operations deals with the collection of returns, cleaning of the collected returns and remanufacturing of the reusable collected items, while in the forward supply chain we have taken the single setup and multiple JIT deliveries (SSMD) strategy. The vendor fulfilled buyer's demand with newly produced and remanufactured items of quality standard "as good as those of new products". The produced/remanufactured items are passing through the process of inspection. After inspection the innocent products are delivered to the buyer in $M + N$ deliveries and the rest (deteriorated items or unacceptable for the primary market) are shipped to a secondary market in as-in condition. A general framework of such a system is shown in Fig. 1. In the next Sect. 2, assumptions and notation are provided for model development. In Sect. 3, the study develops an integrated buyer-vendor model with single setup and multiple deliveries. The model considers JIT deliveries, reverse-manufacturing costs and other costs, and derives the optimal replenishment. A numerical example is presented in Sect. 4. Concluding remarks are shown in Sect. 5.

2 Notations and Assumptions

2.1 Notations

C_{if}	Fixed inspection cost
C_{iv}	Variable inspection cost per setup
C_{ins}	The unit variable inspection cost
L_{fs}	The cost of less flexibility
H_B	Unit holding cost for the buyer
T_B	Delivery cycle time for the buyer

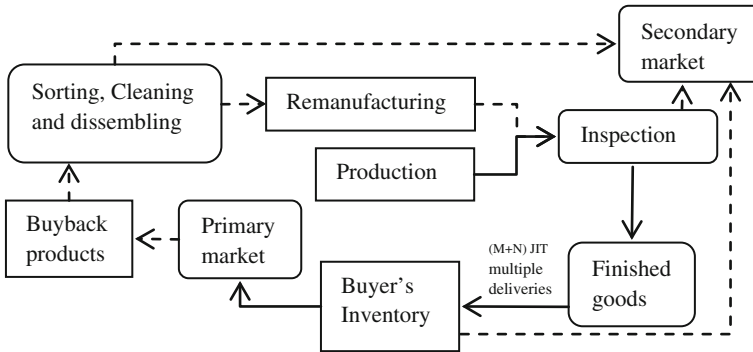


Fig. 1 Material flow in the green supply chain

- q Delivery size per delivery
- A Ordering cost per cycle
- P_m The production rate
- D The demand rate
- θ Deterioration rate
- M Number of deliveries delivered by the producer
- U_m Unit item cost
- K_m Fixed unit production cost
- C_m Variable unit production cost
- H_m Unit holding cost for the produced material
- C_s Selling price for the secondary market
- S_s Selling price for the primary market
- P_r Remanufacturing rate
- R Returned rate
- N The number of deliveries delivered by the remanufacturer
- η Scaling parameter, returned formulation
- K_r Fixed unit reproduction cost
- C_r Variable unit reproduction cost
- H_r Unit holding cost for the remanufactured items
- γ The scaling parameter, returned rate formulation
- U_R Unit returned item cost
- H_R Unit holding cost for the returned items
- CI Collection investment
- F_{cl} Fixed cost including cleaning and disassembly cost during the collecting process
- C_{cl} Variable cost including cleaning and disassembly cost during the collecting process

2.2 Assumptions

- Vendor fulfilled buyer's demand by the newly produced and remanufactured items.
- The buyback products are subject to the cleaning and disassembling process, where the products that confirm to certain quality standard (At a rate of $\left(\frac{1}{\eta}\right)R$) is delivered to remanufacturing unit and the rest $\left\{1 - \left(\frac{1}{\eta}\right)R\right\}$ are shipped for the secondary market in as-in condition.
- Items deteriorate while they are in storage, and production, remanufacturing and demand.
- After the production/remanufacturing products though the inspection process and the deteriorated items are shipped to the secondary market.
- The returned rate R is determined by the collection investment CI. The collection investment represents the monetary amount of effort (e.g., promotion, marketing) that the recycled-material vendor applies to the end-user market to create the necessary incentive to receive targeted returns. We model the return rate similar to the work by Savaskan et al. [17] in which, $R = \left(\sqrt{\frac{CI}{\gamma}}\right)$, where γ is a scaling parameter and $\left(\sqrt{\frac{CI}{\gamma}}\right) < D$
- We shall require that

$$P_r > D, P_m > D, P_r > R, D \neq 0, R \neq 0$$

3 Mathematical Modelling and Analysis

The items are ordered by buyer in $M + N$ shipments (equal lots of size q). The model assumes that the successive delivering batch arrives at the store as soon as the previous batch has been depleted. For each cycle, the system starts operating at time zero by which the reproduction process starts and the inventory level increases until time T_1 , where the stock-level reaches its maximum, and the reproduction process stopped. At the same instant of time production starts and the stock level of remanufactured items decreases continuously due to the demand and deterioration, while production is in process by the time T_2 where it reaches its maximum and the production process stopped. Then the inventory level decrease continuously by the time T_3 . The inventory level in returned process decreases until time T_1 where the inventory level becomes zero. After that the inventory level increases until the time $T_1 + T_2 + T_3$ where, the inventory level reaches its maximum. The process is repeated (Fig. 2).

The changes in the inventory levels depicted in Fig. 2 are governed by the following differential equations:

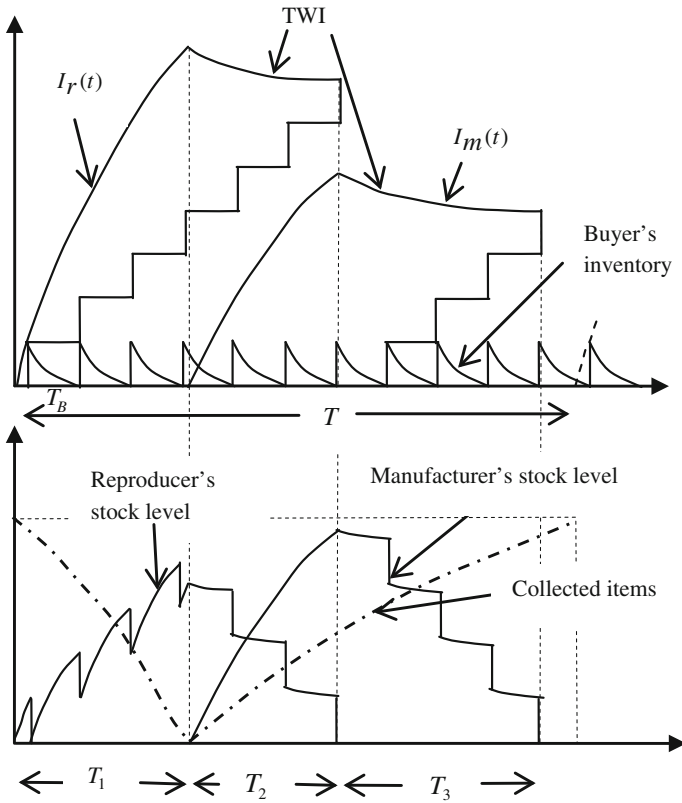


Fig. 2 Inventory variation of an inventory model for reverse logistics systems

$$I'_B(t) = -d - \theta I_B(t), \quad 0 \leq t \leq T_B \quad I_B(T_B) = 0 \tag{1}$$

$$I'_{r1}(t_{r1}) = P_r - \theta I_{r1}(t_{r1}), \quad 0 \leq t_{r1} \leq T_1 \quad I_{r1}(0) = 0 \tag{2}$$

$$I'_{r2}(t_{r2}) = -\theta I_{r2}(t_{r2}), \quad 0 \leq t_{r2} \leq T_2 \quad I_{r2}(T_2) = Nq \tag{3}$$

$$I'_{m1}(t_{m1}) = P_m - \theta I_{m1}(t_{m1}), \quad 0 \leq t_{m1} \leq T_2 \quad I_{m1}(0) = 0 \tag{4}$$

$$I'_{m2}(t_{m2}) = -\theta I_{m2}(t_{m2}), \quad 0 \leq t_{m2} \leq T_3 \quad I_{m2}(T_3) = Mq \tag{5}$$

$$I'_R(t_R) = \frac{1}{\eta}R - P_r, \quad 0 \leq t_R \leq T_1 \quad I_R(T_1) = 0 \tag{6}$$

$$I'_R(t_R) = \frac{1}{\eta}R \quad T_1 \leq t_R \leq (N + M)T_B \quad I_R(T_1) = 0 \tag{7}$$

The solutions of the above differential equations are

$$I_B(t) = \left\{ \frac{d}{\theta} \right\} (e^{\theta(T_B-t)} - 1) \quad 0 \leq t \leq T_B \tag{8}$$

$$I_{r1}(t_{r1}) = \left\{ \frac{P_r}{\theta} \right\} (1 - e^{-\theta t_{r1}}) \quad 0 \leq t_{r1} \leq T_1 \quad (9)$$

$$I_{r2}(t_{r2}) = qNe^{\theta(T_2 - t_{r2})} \quad 0 \leq t_{r2} \leq T_2 \quad (10)$$

$$I_{m1}(t_{m1}) = \left\{ \frac{P_m}{\theta} \right\} (1 - e^{-\theta t_{m1}}) \quad 0 \leq t_{m1} \leq T_2 \quad (11)$$

$$I_{m2}(t_{m2}) = qMe^{\theta(T_3 - t_{m2})} \quad 0 \leq t_{m2} \leq T_3 \quad (12)$$

$$I_R(t_R) = \left(P_r - \frac{1}{\eta} R \right) (T_1 - t_R), \quad 0 \leq t_R \leq T_1 \quad (13)$$

$$I_R(t_R) = \left\{ \frac{R}{\eta} \right\} (t_R - T_1), \quad T_1 \leq t_R \leq (N + M)T_B \quad (14)$$

Respectively

Using the boundary conditions $I_B(0) = q$ so that from Eq. (1), when $\theta \ll 1$ the delivery size is

$$q = dT_B \left(1 + \frac{\theta T_B}{2} \right) \quad (15)$$

As depicted in Fig. 2 we have

$$T_3 = MT_B \quad (16)$$

and

$$T_1 + T_2 = (N - 1)T_B + \frac{q}{P_r} \quad (17)$$

By which, we get

$$T_1 + T_2 + T_3 = (M + N - 1)T_B + \frac{q}{P_r} \quad (18)$$

Assuming the cycle time T so we have

$$T = (M + N)T_B \quad (19)$$

Using the boundary conditions, from Eqs. (13) and (14) we get

$$\left(P_r - \frac{1}{\eta} R \right) T_1 = \frac{1}{\eta} R ((M + N)T_B - T_1)$$

By which the remanufacturing period is

$$T_1 = \frac{R(M + N)T_B}{\eta P_r} \quad (20)$$

From Eq. (17) the production period

$$T_2 = (N - 1)T_B + \frac{q}{P_r} - \frac{R(M + N)T_B}{\eta P_r} = \left((N - 1) - \frac{R(M + N)}{\eta P_r} \right) T_B + \frac{q}{P_r} \quad (21)$$

Now the per cycle cost components for the given inventory system are as follows.

Sales Revenue from the secondary market

$$= c_s \left[(q - dT_B) + \{P_r T_1 + P_m T_2 - q(N + M)\} + \left\{ (M + N) \left(1 - \frac{1}{\eta} \right) \sqrt{\frac{CI}{\gamma}} \right\} T_B \right]$$

Sales Revenue from the primary market

$$= S_s d(M + N)T_B$$

Buyer's ordering cost and holding cost is as follows

$$= A + (M + N)H_B \int_0^{T_B} I_B(u) du$$

$$\text{Production and remanufacturing cost} = K_m + K_r + C_m \int_0^{T_2} P_m du + C_r \int_0^{T_1} P_r du = K_r + K_m + C_r T_1 P_r + C_m P_m T_2$$

Procurement and acquisition cost

$$= U_m \int_0^{T_2} P_m du + U_R \int_0^{(M+N)T_B} R du = U_m P_m T_2 + U_R (M + N) \sqrt{\frac{CI}{\gamma}} T_B$$

$$\text{Inspection cost} = \frac{C_{if}}{M+N} + (M + N)C_{iv} + C_{ins}(P_r T_1 + P_m T_2)$$

$$\text{The Vendor's cost of less flexibility of implementing JIT delivery is} = (M + N)L_{fs}$$

As depicted in Fig. 2 the inventory holding cost is

$$\begin{aligned} &= H_m \left[\int_0^{T_2} \left\{ \frac{P_m}{\theta} \right\} (1 - e^{-\theta u}) du + \int_0^{T_3} q M e^{\theta(T_3 - u)} du - T_B (q + 2q \dots + (M - 1)q) \right] \\ &+ H_r \left[\int_0^{T_1} \left\{ \frac{P_r}{\theta} \right\} (1 - e^{-\theta u}) du + \int_0^{T_2} q N e^{\theta(T_2 - u)} du - T_B (q + 2q \dots + (N - 1)q) \right] \\ &+ H_R \left[\int_0^{T_1} \left\{ P_r - \frac{1}{\eta} \sqrt{\frac{CI}{\gamma}} \right\} (T_1 - u) du + \int_{T_1}^{(M+N)T_B} \frac{1}{\eta} \sqrt{\frac{CI}{\gamma}} (u - T_1) du \right] \\ &+ H_B \int_0^{T_B} \left\{ \frac{d}{\theta} \right\} (e^{\theta(T_B - t)} - 1) du \end{aligned}$$

$$\text{Cleaning cost} = F_{cl} + C_{cl} \sqrt{\frac{CI}{\gamma}} T_B (M + N)$$

Total profit is calculated here,

$$\begin{aligned}
TP = \frac{1}{T} & \left\{ S_s D(M+N)T_B + c_s \left[(q - DT_B) + \{P_r T_1 + P_m T_2 - q(N+M)\} + (M+N) \left(1 - \frac{1}{\eta}\right) \sqrt{\frac{CI}{\gamma}} \right] T_B \right. \\
& - (M+N)H_B \left[\frac{DT_B^2}{2} \left(1 + \frac{\theta T_B}{3}\right) \right] \\
& - A - (U_m + C_m)P_m T_2 - \frac{C_{if}}{M+N} - (M+N)C_{iv} - C_{ins}(P_r T_1 + P_m T_2) - (M+N)L_{fs} \\
& - U_R(M+N) \sqrt{\frac{CI}{\gamma}} T_B - C_r T_1 P_r - F_{cl} - C_{cl} \sqrt{\frac{CI}{\gamma}} T_B (M+N) \\
& - H_m \left[\frac{P_m T_2^2}{2} \left(1 - \frac{\theta T_2}{3}\right) + MDT_B T_3 \left(1 + \frac{\theta T_B}{2} - \frac{\theta T_3}{2}\right) - \frac{1}{2} \left(DT_B^2 \left(1 + \frac{\theta T_B}{2}\right) M(M-1) \right) \right] \\
& - H_r \left[\frac{P_r T_1^2}{2} \left(1 - \frac{\theta T_1}{3}\right) + NDT_B T_3 \left(1 + \frac{\theta T_B}{2} - \frac{\theta T_2}{2}\right) \right. \\
& \left. - \frac{1}{2} \left(DT_B^2 \left(1 + \frac{\theta T_B}{2}\right) N(N-1) \right) \right] \\
& \left. - H_R \left[\left(P_r - \frac{1}{\eta} \sqrt{\frac{CI}{\gamma}} \right) \frac{T_1^2}{2} + \frac{1}{\eta} \sqrt{\frac{CI}{\gamma}} \left(\frac{1}{2} (N+M)^2 T_B^2 - \frac{T_1^2}{2} - (N+M)T_B T_1 + T_1^2 \right) \right] - CI - K_m - K_r \right\} \quad (22)
\end{aligned}$$

3.1 Solution Procedure

From Eqs. (15)–(22), the total profit function can be determined in terms of N , M and T_B .

Hence the purpose of this study is to derive the optimal number of deliveries and the replenishment cycle time by determining the optimal values of N , M and T_B that maximize the total profit. The model has been solved using Algorithm given below.

3.2 Algorithm

Step 1: Taking the first derivative of the total profit function w.r.t. to T_B , equate it equals to zero and then determine the value of T_B

Step 2: Putting the value of T_B in Eq. (25) and then derive the value of N and M by using the following conditions for maximizing the supplier's profit.

$$\begin{aligned}
TP(N, (M-1)|T_B) & \leq TP(N, M|T_B) \geq TP(N, (M+1)|T_B) \\
TP((N-1), M|T_B) & \leq TP(N, M|T_B) \geq TP((N+1), M|T_B)
\end{aligned}$$

Step 3: Using the optimum values of N , M and T_B determine the maximum profit for the system.

Table 1 Optimal results for integrated cost minimization problem

M	N	T_B	q	T_1	T_2	T_3	TC
4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98

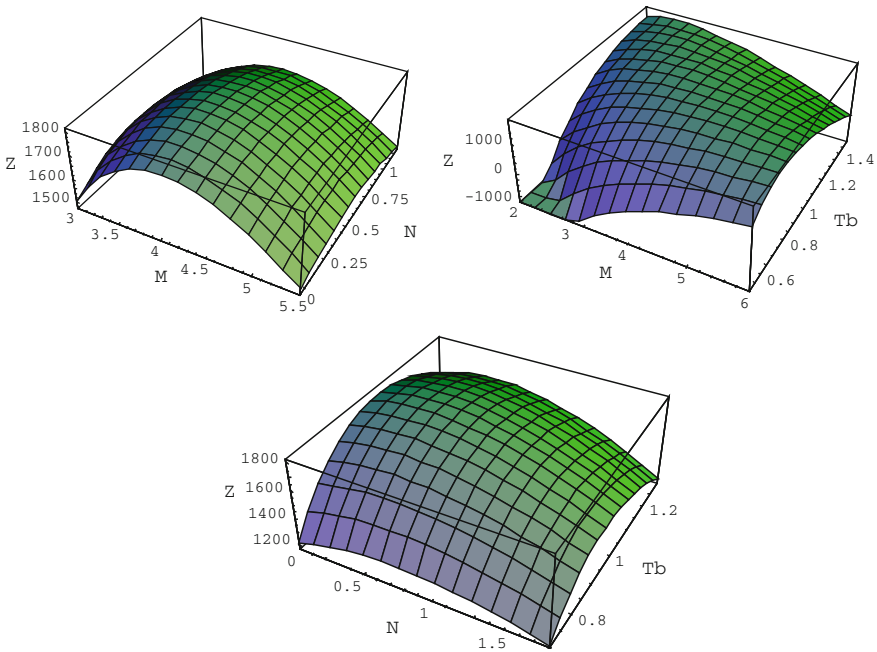


Fig. 3 Convexity of the closed loop supply chain

4 Numerical Analysis and Sensitivity Analysis

Example 1 The above theoretical results are illustrated through the numerical verification, to illustrate the proposed model we have considered the following input parameters in appropriate units

$P_r = 2500, P_m = 1200, d = 1000, \theta = 0.01, \eta = 5, CI = 8000, \gamma = 0.01, A = 500, H_B = 1, H_r = 1, H_R = 0.8, H_m = 1, K_m = 1000, K_R = 1000, C_m = 6, C_r = 3, U_R = 3, U_m = 4, C_S = 6, S_S = 12, C_{if} = 3000, C_{iv} = 500, C_{ins} = 0.5, L_{fs} = 100, C_{cl} = 0.2, F_{cl} = 1500.$

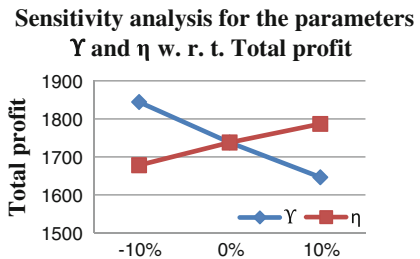
Mathematica 8.0 is used to derive the optimal solution and results are presented in Table 1.

The convexity of the reverse logistics inventory model is shown in Fig. 3. The three dimensional graph shows that the integrated expected total profit is concave, and that there exists a unique solution maximizing the integrated expected total profit.

Table 2 Sensitivity analysis for the production, remanufacturing and demand rate parameters F_{cl} , C_{iv} , A , η , γ , P_m , P_r and d

		M	N	T_B	q	T_1	T_2	T_3	TP
γ	-10 %	4	1	1.00837	1013.45	0.38028	0.0251014	4.03348	1844.32
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01315	1018.28	0.345606	0.0617059	4.05259	1646.42
η	-10 %	4	1	1.00566	1010.72	0.399773	0.00451352	4.02264	1678.06
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01525	1020.41	0.330207	0.0779554	4.06101	1786.73
P_m	-10 %	4	1	1.01098	1016.09	0.361698	0.0447368	4.04391	1742.78
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.0109	1016.01	0.36167	0.0447332	4.0436	1733.18
P_r	-10 %	4	1	1.01061	1015.72	0.401741	0.0496887	4.04244	1730.9
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01121	1016.32	0.328891	0.0406795	4.04483	1743.76
d	-10 %	4	1	1.05883	957.989	0.378817	0.00437826	4.23531	1459.25
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	0.968789	1070.83	0.346605	0.0817275	3.87516	2027.14
A	-10 %	4	1	1.00936	1014.45	0.361118	0.0446618	4.03743	1747.88
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01252	1017.64	0.362249	0.0448081	4.05007	1728.09
C_{iv}	-10 %	4	1	1.003	1008.03	0.358845	0.044368	4.01201	1787.63
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01881	1024.00	0.364499	0.0450992	4.07523	1688.71
F_{cl}	-10 %	4	1	1.00619	1011.25	0.359984	0.0445151	4.02474	1767.72
	0 %	4	1	1.01094	1016.05	0.361684	0.044735	4.04375	1737.98
	+10 %	4	1	1.01567	1020.83	0.363376	0.0449539	4.06267	1708.37

Fig. 4 Effect of returned rate on total profit



Example 2 We now study the effects of changes in the values of the system parameters F_{cl} , C_{iv} , A , η , γ , P_m , P_r and d on the optimal total profit and number of reorder. The sensitivity analysis is performed by changing each of the parameters by -10 , 0 and 10 %, taking one parameter at a time and keeping the remaining parameters the same as in Example 1. The sensitivity analyses of the different parameters are shown in Table 2. The graphical representation of the sensitivity of decision variables is provided in Figs. 4, 5 and 6.

Fig. 5 Effect of production, remanufacturing and demand on total profit

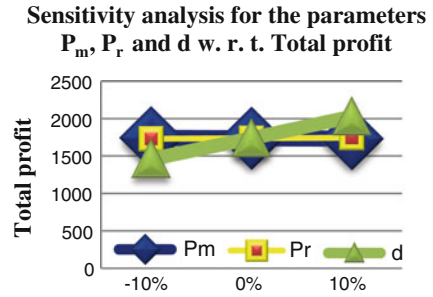
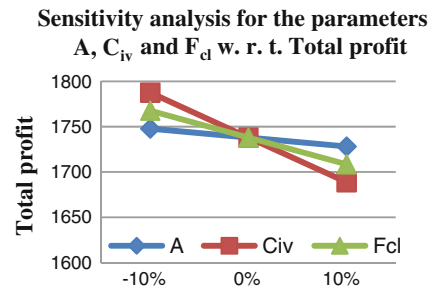


Fig. 6 Effect of ordering, inspection and cleaning cost on total profit



4.1 Observations

- 1) Table 2 reveals that the total profit is positive sensitive to the changes in parameter η while negative sensitive to the changes in parameter γ hence it seems that to procure the used products from the user is profitable but resell them in the secondary market in as-is condition is better than the remanufacturing.
- 2) From the table it is observed that the total profit is slightly negative sensitive to the changes in parameter P_m while slightly positive sensitive to the changes in parameter P_r . Hence it can be said that remanufacturing is more profitable than the production.
- 3) Table 2 reveals that the total profit is moderately positive sensitive to the changes in demand parameter. This is a logical tendency since it allows the vendor and buyer to get more cost savings from accumulated revenue.
- 4) From the table it is observed that the total profit is moderately positive sensitive to the changes in parameters F_{cl} , C_{iv} and A which is obvious. But if we compare the behaviour of these three parameters it is noted that the cleaning cost is more sensitive than the ordering cost while less sensitive than the inspection cost.

5 Conclusion

In the presented paper we have discussed the effect of JIT on a green supply chain inventory model. Here we have determined the model for the deteriorating products where the deteriorated products are delivered to the secondary market. In

this article we derived a profit function for the mathematical model developed here and hence determine the optimal results with the help of a numerical example. The results shows that the reselling the used products in the secondary market is profitable than the remanufacturing. In the further study the model can be developed for the remanufactured items whose quality standard is not as good as those of new products.

Acknowledgments The second author would like to express her thanks to University Grants Commission India for providing the financial help in the form of Junior Research Fellowship.

References

1. Alamri, A.A.: Theory and methodology on the global optimal solution to a general Reverse logistics Inventory Model for deteriorating items and time-varying rates. *Comput. Ind. Eng.* **60**, 236–247 (2010)
2. Banerjee, A.: A joint economic-lot-size model for purchaser and vendor. *Decis. Sci.* **17**, 292–311 (1986)
3. Chung, C.J., Wee, H.M.: Short life-cycle deteriorating product remanufacturing in a green supply chain inventory control system. *Int. J. Prod. Econ.* **129**, 195–203 (2011)
4. Dekker, R., Fleischmann, M., Inderfurth, K.: *Reverse logistics: quantitative models for closed-loop supply chains*. Springer-Verlag, Heidelberg (2004)
5. El Saadany, A.M.A., Jaber, M.Y.: A production/remanufacturing inventory model with price and quality dependant return rate. *Comput. Ind. Eng.* **58**(3), 352–362 (2010)
6. Ishii, K., Eubanks, C.F., Marco, P.D.: Design for product retirement and material life-cycle. *J. Mate. Des.* **15**(4), 225–233 (1994)
7. Jaber, M.Y., El Saadany, A.M.A.: The production, remanufacture and waste disposal model with lost sales. *Int. J. Prod. Econ.* **120**(1), 115–124 (2009)
8. Konstantaras, I., Skouri, K.: Lot sizing for a single product recovery system with variable setup numbers. *Eur. J. Oper. Res.* **203**(2), 326–335 (2010)
9. Kim, S.L., Ha, D.: A JIT lot-splitting model for supply chain management: enhancing buyer–supplier linkage. *Int. J. Prod. Econ.* **86**, 1–10 (2003)
10. King, A.M., Burgess, S.C., Ijomah, W., McMahon, C.A.: Reducing waste: repair, recondition, remanufacture or recycle? *Sustain. Dev.* **14**(4), 257–267 (2006)
11. Lu, L.: A one-vendor multi-buyer integrated inventory model. *Eur. J. Oper. Res.* **81**, 312–323 (1995)
12. Nahmias, N., Rivera, H.: A deterministic model for a repairable item inventory system with a finite repair rate. *Int. J. Prod. Res.* **17**(3), 215–221 (1979)
13. Richter, K.: The EOQ repair and waste disposal model with variable setup numbers. *Eur. J. Oper. Res.* **95**(2), 313–324 (1996)
14. Richter, K.: The extended EOQ repair and waste disposal model. *Int. J. Prod. Econ.* **45**(1–3), 443–447 (1996)
15. Richter, K.: Pure and mixed strategies for the EOQ repair and waste disposal problem. *OR Spektrum* **19**(2), 123–129 (1997)
16. Richter, K., Dobos, I.: Analysis of the EOQ repair and waste disposal model with integer setup numbers. *Int. J. Prod. Econ.* **59**(1–3), 463–467 (1999)
17. Savaskan, R.C., Bhattacharya, S., van Wassenhove, L.N.: Closed-loop supply chain models with product remanufacturing. *Manage. Sci.* **50**(2), 239–252 (2004)
18. Schrady, D.A.: A deterministic inventory model for repairable items. *Naval Res. Logist. Q.* **14**, 391–398 (1967)

19. Singh, S.R., Prasher, L., Saxena, N.: A centralized reverse channel structure with flexible manufacturing under the stock out situation. *Int. J. Ind. Eng. Comput.* **4**, 559–570 (2013)
20. Singh, S.R., Saxena, N.: An optimal returned policy for a reverse logistics inventory model with backorders. *Adv. Decis. Sci.* **2012**(386598), 21 (2012)

Transient Stability Enhancement of a Multi-Machine System Using BFOA

M. Jagadeesh Kumar, S. S. Dash, C. Subramani, M. Arun Bhaskar and R. G. Akila

Abstract The Static Synchronous Compensator (STATCOM) is the typical Flexible AC Transmission System (FACTS) devices playing a vital role as a stability aid for the large transient disturbances in a multi-machine power system. This paper deals with the design of STATCOM with two different controllers installed in a multi-machine power system. The disturbances are created in the multi-machine system and the results obtained are analyzed. The results proved the supremacy of the feedback linearizing controller tuned by Bacterial Foraging Optimization Algorithm (BFOA) of STATCOM over the power system stabilizer (PSS) equipped with PI Proportional-Integral controller of STATCOM.

Keywords STATCOM · FACTS · PI · PSS · AV · FBLC · BFOA

1 Introduction

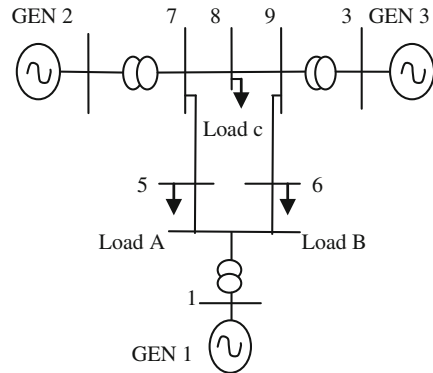
Transient stability refers to the ability of a power system to maintain synchronism when subjected to a severe and sudden transient disturbance [1]. Stabilization of a synchronous generator is undoubtedly one of the most important problems in power system control. Automatic voltage regulators (AVR) with exciter and power system stabilizer (PSS) are normally employed for the purpose of damping the electromechanical oscillations and also for the post fault bus-voltage recovery.

However, it is well known that the performances of PSS and AVR are limited since their designs are primarily based on linear control algorithms. In the event of

M. Jagadeesh Kumar (✉) · M. Arun Bhaskar · R. G. Akila
Department of EEE, Velammal Engineering College, Chennai 600066, India
e-mail: mp.jagadeesh@gmail.com

S. S. Dash · C. Subramani
Department of EEE, SRM University, Kattankulathur, Chennai, India

Fig. 1 Multi-machine power system



large faults, the nonlinearities of the system become very severe, thereby putting limitations on the performances of classical control designs. Thus, the most appropriate approach for controller design for a power system is the use of non-linear control theory, i.e., multivariable feedback linearization scheme.

The feedback linearization technique is based on the idea of cancelling the nonlinearities of the system and imposing a desired linear dynamics to control the system. Successful applications of FACTS equipment for power flow control, voltage control and transient stability improvement have been reported in the literature [2, 3]. During the last decade, a technology called Flexible AC Transmission Systems (FACTS) have been proposed and implemented.

2 IEEE Standard Multi-Machine Test System

Figure 1 shows the multi-machine power system considered in this paper. It consists of nine bus and three synchronous generators. The FACTS device is connected at the eighth bus.

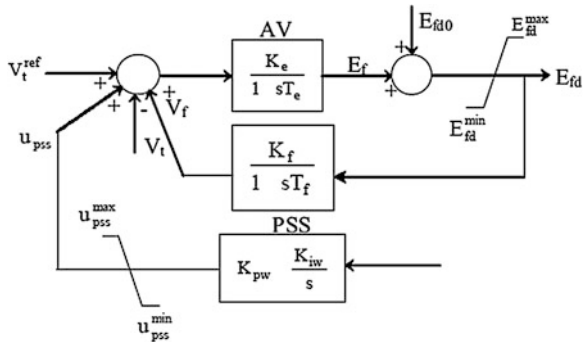
3 Mathematical Models and Conventional Control Schemes

3.1 Synchronous Generator and Speed Governor

The synchronous generator is described by a third order non-linear model [1, 4]

$$\frac{d\delta}{dt} = \Delta\omega \quad (1)$$

Fig. 2 AVR, exciter and PSS control system of generating station



$$\frac{d\Delta\omega}{dt} = \frac{1}{M} [P_m - E'_q i_q - (x_q - x'_d) i_d i_q] \quad (2)$$

$$\frac{dE'_q}{dt} = \frac{1}{T'_{do}} [E_{fd} - E'_q - (x_d - x'_d) i_d] \quad (3)$$

$$\frac{d\Delta P_v}{dt} = [-\Delta P_v - K_g \Delta\omega + u_g] \quad (4)$$

$$\frac{d\Delta P_t}{dt} = [\Delta P_v - \Delta P_t] \quad (5)$$

Where $\Delta\delta = \delta - \delta_o$ and $\Delta\omega = \omega - \omega_o$.

3.2 AVR, Exciter and PSS

The excitation system of the generator consists of a simple automatic voltage regulator (AVR) along with exciter and a supplementary power system stabilizer (PSS). The complete AVR, exciter and PSS is shown in Fig. 2.

3.3 Modeling of STATCOM

The dynamic equations governing the instantaneous values of the three-phase voltages across the two sides of shunt converter [5] and the current flowing into it are given by:

$$\left(R_p + L_p \frac{d}{dt} \right) i_p = V_s - V_p \quad (6)$$

where

$$\begin{aligned}
i_p &= [i_{pa} i_{pb} i_{pc}]^T \\
V_s &= [V_{sa} V_{sb} V_{sc}]^T \\
V_p &= [V_{pa} V_{pb} V_{pc}]^T \\
R_p &= \begin{bmatrix} R_p & 0 & 0 \\ 0 & R_p & 0 \\ 0 & 0 & R_p \end{bmatrix} \quad \text{and} \quad L_p = \begin{bmatrix} L_p & 0 & 0 \\ 0 & L_p & 0 \\ 0 & 0 & L_p \end{bmatrix}
\end{aligned} \tag{7}$$

Let us assuming that the system has no zero sequence components, all currents and voltages can be uniquely represented by equivalent space phasors and then transformed into the synchronous d - q - o frame by applying the following transformation (θ is the angle between the d -axis and reference phase axis):

Thus the transformed dynamic equations are:

$$\frac{di_{pd}}{dt} = -\frac{R_p}{L_p} i_{pd} + \omega i_{pq} + \frac{1}{L_p} (V_{sd} - V_{pd}) \tag{9}$$

$$\frac{di_{pq}}{dt} = -\omega i_{pd} - \frac{R_p}{L_p} i_{pq} + \frac{1}{L_p} (V_{sq} - V_{pq}) \tag{10}$$

where ω is the angular frequency of the AC-bus voltage.

The d - and q -axis components of the VSI voltage, i.e., V_{pd} and V_{pq} , all harmonics, which are near to or above the VSI switching frequency are neglected. In the real-time implementation of these quantities, they are converted into modulation index (m) and phase angle (ϕ):

$$\begin{aligned}
M &= \frac{\sqrt{V_{pd}^2 + V_{pq}^2}}{KV_{dc}} \\
\Phi &= \tan^{-1} \left(\frac{V_{pq}}{V_{pd}} \right)
\end{aligned} \tag{11}$$

where k is a constant whose amplitude depends upon the adopted modulation technique.

On an instantaneous basis, for an effective DC-voltage control, the input power should be equal to the sum of load power and the charging rate of capacitor voltage. Thus, by power balance between the AC input and the DC output is given by:

$$P = [V_{sd} i_{pd} + V_{sq} i_{pq} - (i_{pd}^2 + i_{pq}^2) R_p] = CV_{dc} \frac{dV_{dc}}{dt} + \frac{V_{dc}^2}{R_{dc}} \tag{12}$$

Hence:

$$\frac{dV_{dc}}{dt} = \frac{V_{sd} i_{pd} + V_{sq} i_{pq} - (i_{pd}^2 + i_{pq}^2) R_p}{CV_{dc}} - \frac{V_{dc}}{CR_{dc}} \tag{13}$$

Equation (13) models the dynamic behaviour of the dc -side capacitor voltage. Thus Eqs. (9), (10) and (13) together describe the dynamic model of the converter which is summarized in the following equation:

$$\frac{d}{dt} \begin{bmatrix} i_{pd} \\ i_{pq} \\ V_{dc} \end{bmatrix} = \begin{bmatrix} -\frac{R_p}{L_p} i_{pd} + \omega i_{pq} \\ -\frac{R_p}{L_p} i_{pq} - \omega i_{pd} \\ \frac{1}{CV_{dc}} [V_{sd} i_{pd} + V_{sq} i_{pq} - (i_{pd}^2 + i_{pq}^2) R_p] - \frac{V_{dc}}{CR_{dc}} \end{bmatrix} \begin{pmatrix} \frac{1}{L_p} & 0 \\ 0 & \frac{1}{L_p} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_{sd} & -V_{pd} \\ V_{sq} & -V_{pq} \end{pmatrix} \quad (14)$$

3.3.1 Cascade Control Scheme for Series Converter

The conventional control strategy for converter mainly deals with the control of ac -bus and dc -bus voltage on both sides of converter. The dual control objectives have been met by generating appropriate current reference (for d - and q -axis) and then by regulating those currents in the converter. While attempting to decouple the d - and q -axis current regulators, PI controllers are conventionally employed. The converter current (ip) is split into real (in phase with ac -bus voltage) and reactive components. The reference value for the real current is decided so that the capacitor voltage is regulated by power balance. The reference for reactive component is determined by ac -bus voltage regulator. As per the strategy, the original currents in d - q frame (ipd and ipq) are transformed into another frame called d' - q' frame, where d' -axis coincides with the ac -bus voltage (V_s), as shown in Fig. 3.

Thus, in d' - q' frame, the currents ipd' and ipq' represent the real and reactive currents and they are given by:

$$i_{pd'} = i_{pd} \cos \delta_s + i_{pq} \sin \delta_s \quad (15)$$

$$i_{pq'} = i_{pq} \cos \delta_s - i_{pd} \sin \delta_s \quad (16)$$

Now, for current control, the differential Eqs. (9) and (10) are re expressed as:

$$\frac{di_{pd'}}{dt} = -\frac{R_p}{L_p} i_{pd'} + \bar{\omega} i_{pq'} + \frac{1}{L_p} (v_s - v_{pd'}) \quad (17)$$

$$\frac{di_{pq'}}{dt} = -\bar{\omega} i_{pd'} - \frac{R_p}{L_p} i_{pq'} + \frac{1}{L_p} (-v_{pq'}) \quad (18)$$

where

$$v_{pd'} = v_{pd} \cos \delta_s + v_{pq} \sin \delta_s \quad (19)$$

$$v_{pq'} = v_{pq} \cos \delta_s - v_{pd} \sin \delta_s \quad (20)$$

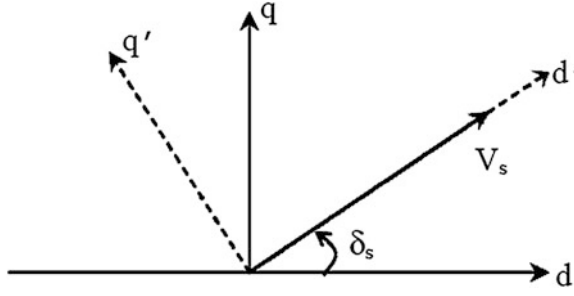


Fig. 3 Phasor diagram showing d - q and d' - q' frames

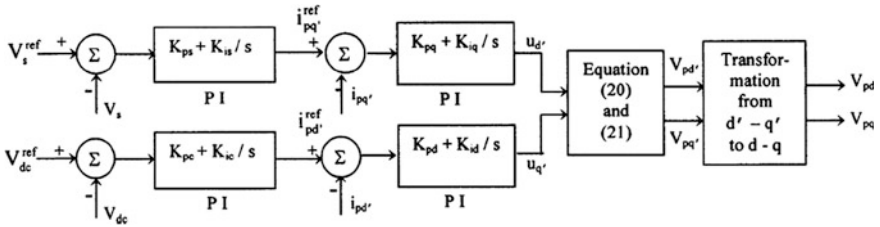


Fig. 4 Cascade control architecture

$$\bar{\omega} = \omega + \frac{d\delta_s}{dt} \quad (21)$$

The VSI voltages are controlled as follows:

$$v_{pq'} = -(\bar{\omega}L_{pi_{pd'}} + L_{pu_{q'}}) \quad (22)$$

$$v_{pd'} = \bar{\omega}L_{pi_{pq'}} + V_{s^-} + L_{pu_{d'}} \quad (23)$$

By putting the above expressions for $V_{pd'}$ and $V_{pq'}$ in Eqs. (17) and (18), the following set of decoupled equations are obtained:

$$\frac{di_{pd'}}{dt} = -\frac{R_p}{L_p}i_{pd'} + u_{d'} \quad (24)$$

$$\frac{di_{pq'}}{dt} = -\frac{R_p}{L_p}i_{pq'} + u_{q'} \quad (25)$$

Also, the dc -bus voltage dynamic equation is now modified as:

$$\frac{dV_{dc}}{dt} = \frac{V_s i_{pd'} - (i_{pd'}^2 + i_{pq'}^2)R_p}{CV_{dc}} - \frac{V_{dc}}{CR_{dc}} \quad (26)$$

Now, the control signals ud' and uq' can be easily determined by linear PI controllers. The complete cascade control architecture is shown in Fig. 4 where Kps , Kis , Kpc , Kic , Kpq , Kiq , Kpd and Kid are the respective gains of the PI controllers.

3.4 Feedback Linearising Non-linear Control of STATCOM

In this section, the design steps for the feedback linearizing control of STATCOM have been presented followed by simulation results under various transient disturbances [6].

3.4.1 Non-linear Control Design

In the STATCOM control, there are two broad objectives, i.e., *ac*-bus voltage (V_s) and *dc*-bus voltage (V_{dc}) control. In the following control design, V_s is taken as an additional state in addition to the other three states (ipd' , ipq' and V_{dc}) in the STATCOM modeling [7]. The dynamic equation for V_s is obtained as follows with reference to Fig. 4 (in the d' - q' frame):

$$V_s = V_{td'} + x_{t1}(i_{pq'} + i_{bq'} + i_{lq'}) \quad (27)$$

$$\frac{dV_s}{dt} = \frac{dy_{td'}}{dt} + x_{t1} \frac{d}{dt}(i_{lq'} + i_{bq'}) - \left(\frac{x_{t1}R_p}{L_p}\right)i_{pq'} - x_{t1}\omega i_{pd'} - \left(\frac{x_{t1}}{L_p}\right)V_{pq'} \quad (28)$$

Consider a MIMO system, the state equation of this non-linear system is given by,

$$\begin{aligned} \dot{x} &= f(x) + g(x)u \\ y &= h(x) \end{aligned}$$

where $x(\in \mathfrak{R}^n)$ is state vector, $u(\in \mathfrak{R}^m)$ represents control inputs, $y(\in \mathfrak{R}^m)$ stands for output, f and g are smooth vector fields, and h is a smooth scalar function. The input–output linearization of the above MIMO system is achieved by differentiating y of the system until the outputs appear explicitly.

Now, for the control design, the complete state space model is expressed in the form of above two equations follows:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} i_{pd'} \\ i_{pq'} \\ V_{dc} \\ V_s \end{bmatrix} \quad U = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -V_{pd'} \\ -V_{pq'} \end{bmatrix} \quad (29)$$

$$\dot{x}_1 = f_1(x) + g_1u_1 \quad (30)$$

$$x_2 = f_2(x) + g_2 u_2 \quad (31)$$

$$x_3 = f_3(x) \quad (32)$$

$$x_4 = f_4(x) + g_{41} u_1 + g_{42} u_2 \quad (33)$$

$$f_1(x) = -\frac{R_p}{L_p} x_1 + \omega x_2 + \frac{1}{L_p} x_4 \quad (34)$$

$$g_1 = \frac{1}{L_p} = g_2 \quad (35)$$

$$f_2(x) = -\omega x_1 - \frac{R_p}{L_p} x_2 \quad (36)$$

$$f_3(x) = \frac{1}{Cx_3} (x_1 x_4 - R_p (x_1^2 + x_2^2)) - \frac{x_3}{CR_{dc}} \quad (37)$$

$$f_4(x) = \left[-\frac{x_{t1} R_p}{L_p} \right] x_2 - \bar{\omega} x_{t1} x_1 + \frac{dV_{td'}}{dt} + x_{t1} \frac{d(i_{lq'} + i_{ld'})}{dt} \quad (38)$$

$$g_{41} = 0, \quad g_{42} = \frac{x_{t1}}{L_p} \quad (39)$$

The outputs of the system are V_s and V_{dc} .

Thus, $y_1 = V_s$ and $y_2 = V_{dc}$.

$$\begin{aligned} \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \end{bmatrix} &= \begin{bmatrix} f_4(x) \\ a_{11}f_1(x) + a_{12}f_2(x) + a_{13}f_3(x) + a_{14}f_4(x) \end{bmatrix} + \begin{bmatrix} g_{41} & g_{42} \\ a_{11}g_1 + a_{14}g_{41} & a_{12}g_2 + a_{14}g_{42} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \\ &= A(x) + E(x) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \end{aligned} \quad (40)$$

where,

$$a_{11} = \frac{1}{2Cx_3} (x_4 - 2R_p x_1) \quad (41)$$

$$a_{12} = -\frac{2R_p x_2}{Cx_3} \quad (42)$$

$$a_{13} = \frac{1}{x_3^2} [R_p (x_1^2 + x_2^2) - x_1 x_4] - \frac{1}{CR_{dc}} \quad (43)$$

$$a_{14} = \frac{x_1}{Cx_3} \quad (44)$$

Thus:

$$\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = E^{-1}(x) \left[-A(x) + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \right] \quad (45)$$

The non-singularity of $E(x)$ can be observed by computing the determinant of $E(x)$, which is:

$$|E(x)| = -\frac{x_{t1}}{CL_p^2} \left(\frac{V_s - 2R_p i_{pd'}}{V_{dc}} \right) \quad (46)$$

It is known that the magnitude of current $i_{pd'}$ is very small such that $2R_p i_{pd'}$ is almost negligible compared to V_s . Now it is readily seen that $E(x)$ is nonsingular in the operating ranges of V_s and V_{dc} . For tracking of V_s and V_{dc} , the new control inputs v_1 and v_2 are selected as (by both proportional and integral control):

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} \dot{Y}_{1ref} + K_{11} \dot{e}_1 + K_{12} \int e_1 dt \\ \ddot{Y}_{2ref} + K_{21} \dot{e}_2 + K_{22} e_2 + K_{23} \int e_2 dt \end{bmatrix} \quad (47)$$

where y_{1ref} is the ac -bus reference voltage (V_s^{ref}) and y_{2ref} is the dc -bus reference voltage (V_{dc}^{ref}) and e_1 and e_2 are error variables defined by:

$$\begin{aligned} e_1 &= V_s^{ref} - V_s \text{ and,} \\ e_2 &= V_{dc}^{ref} - V_{dc} \end{aligned} \quad (48)$$

From the above equations the error dynamics are given by:

$$\ddot{e}_1 + K_{11} \dot{e}_1 + K_{12} e_1 = 0 \quad (49)$$

$$\ddot{e}_2 + K_{21} \dot{e}_2 + K_{22} e_2 + K_{23} \int e_2 dt = 0 \quad (50)$$

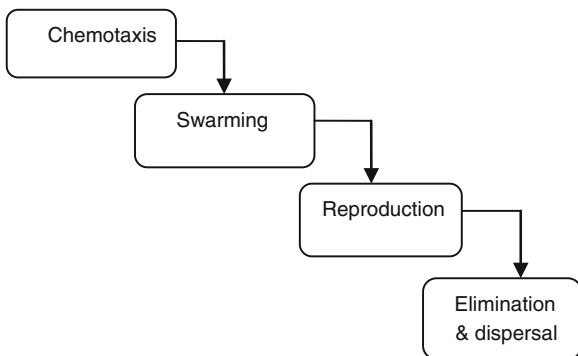
The gain parameters K_{11} , K_{12} , K_{21} , K_{22} and K_{23} are determined by assigning desired poles on the left half s -plane and, thus, asymptotic tracking control to the reference can be achieved. From u_1 and u_2 , the control signals in d' - q' frame are determined by:

$$V_{pd'} = -u_1 \quad \text{and} \quad V_{pq'} = -u_2 \quad (51)$$

Again, from $V_{pd'}$ and $V_{pq'}$, the control signals in d - q frame i.e., V_{pd} and V_{pq} , are obtained by making use of Eqs. (19) and (20).

In the computer simulation studies presented in the followings, the derivative $dv_{td'}/dt$ appearing in the control design, i.e., Eqs. (29) and (39), is neglected in the control computation. This leads to the assumption that the generator bus voltage V_t is treated as a constant only for the control design. In the complete simulation of the nonlinear system, V_t varies as per the system conditions. It is emphasized that, by neglecting $dv_{td'}/dt$ exclusively for the control computation, any unusual rise in the control signal is avoided under severe transient conditions, i.e., 3-phase fault, which may lead to instability depending upon the operating point. Further, because

Fig. 5 Four process in BFOA



of the direct non-availability of rotor speed deviation ($\Delta\omega$) at the load bus, the nominal frequency ω_0 has been used in the control calculations.

3.5 Bacterial Foraging Optimisation Algorithm

Bacterial foraging optimization algorithm (BFOA) has been widely accepted as a global optimization algorithm of current interest for distributed optimization and control. BFOA is inspired by the social foraging behavior of *Escherichia coli*.

It possesses the following 4 process (Fig. 5).

In order to get better output, the control parameters of FBLC are tuned by using Bacterial Foraging Optimization Algorithm (BFOA). There are five control parameters namely K11, K12, K21, K22 & K23.

These control parameters are determined by using two poles namely Pole1 and Pole2. The formula for calculating the control parameters are given below:

- $K11 = 2 \times \text{Pole1}$
- $K12 = \text{Pole1} \times \text{Pole1}$
- $K21 = 3 \times \text{Pole2}$
- $K22 = 3 \times \text{Pole2} \times \text{Pole2}$
- $K23 = \text{Pole2} \times \text{Pole2} \times \text{Pole2}$

These two poles are determined by using BFOA.

The input parameters are given as follows:

- $P = 2$: Dimension of the search space
- $S = 26$: Total number of bacteria
- $N_c = 50$: The number of chemotactic steps
- $N_s = 4$: The swimming length
- $N_{re} = 4$: The number of reproduction steps
- $N_{ed} = 2$: The number of elimination-dispersal events
- $P_{ed} = 0.25$: Elimination-dispersal probability

Fig. 6 BFOA output waveform for pole 1

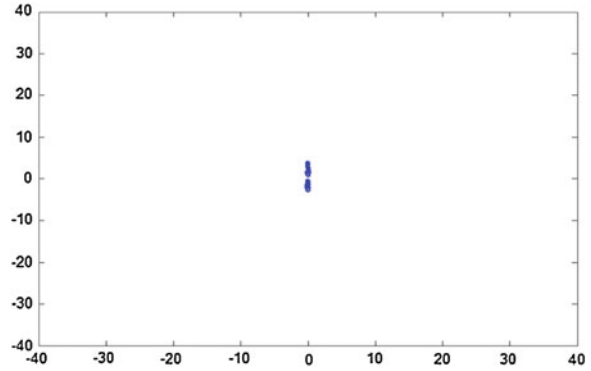
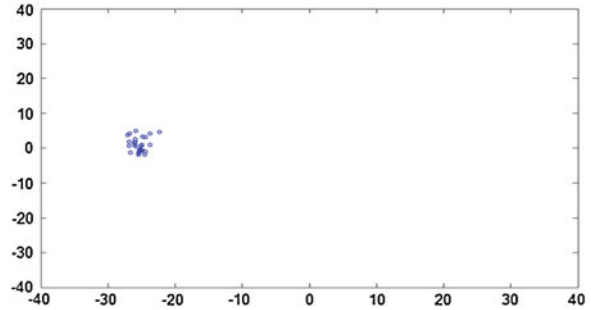


Fig. 7 BFOA output waveform for pole 2



From Figs. 6 and 7 the output waveforms obtained are:

- Pole 1 = 0.03
- Pole 2 = 28.

3.6 Simulation Results

In a three machine nine bus system, the 3 phase fault is created in the bus 7 for the duration of 0.25 s and the line connecting the buses 5 and 7 is removed for the post-fault condition. The 3 phase fault is created at 1 s and removed at 1.25 s. The FACTS device is connected at the 8th bus. The waveforms in Figs. 8, 9 and 10 show the comparison of transient responses of multi-machine system with STATCOM in presence of two different controllers.

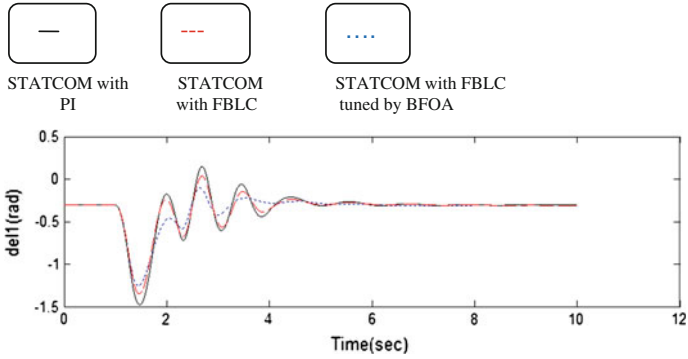


Fig. 8 The rotor angle curves of generator 1

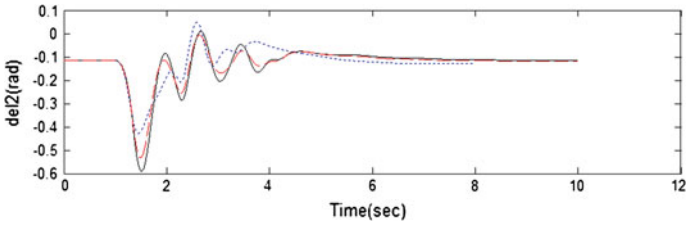


Fig. 9 The rotor angle curves of generator 2

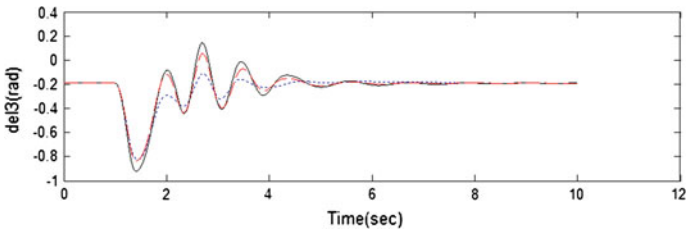


Fig. 10 The rotor angle curves of generator 3

3.7 Comparison of Transient Responses of STATCOM with PI & STATCOM with FBLC

	STATCOM with PI	STATCOM with FBLC	STATCOM with FBLC tuned by BFOA
Fault starting time (s)	1.0	1.0	1.0
Fault duration (s)	0.25 (1.0–1.25)	0.25 (1.0–1.25)	0.25 (1.0–1.25)

(continued)

(continued)

	STATCOM with PI	STATCOM with FBLC	STATCOM with FBLC tuned by BFOA
Fault clearing time for generator 1 (s)	6.0	5.5	4.5
Fault clearing time for generator 2 (s)	5.0	4.5	4.0
Fault clearing time for generator 3 (s)	6.5	6.0	5.0

4 Conclusion

From the comparison table of transient response of PI and FBLC, the fault clearing time for STATCOM with FBLC tuned by BFOA is much less than that of STATCOM with FBLC and STATCOM with PI. Thus FBLC tuned by BFOA will damp out the oscillation at much faster rate than PI and hence enhance the transient stability of the multi-machine system. The mathematical model for STATCOM with FBLC and STATCOM with PI has been independently developed and the MATLAB experimental results proved the supremacy of the FBLC over the PI. For all type of transient disturbances created in the Multi-Machine Infinite Bus (MMIB) system, the Feedback linearizing controller tuned by BFOA of STATCOM damped the electromechanical oscillations faster than the Proportional Integral (PI) controller equipped with Power System Stabilizers (PSS) of STATCOM.

References

1. Anderson, P.M., Fouad, A.A.: Power System Control and Stability. IEEE Press, New York (1994)
2. Mihaliã, R., Zunko, P., Papiã, I., Povh, D.: Improvement of transient stability by insertion of FACTS devices. In: IEEE/NTUA Athens Power Tech. Conference Proceedings, pp. 521–525 (1993)
3. Ghosh, A.; Chatterjee, D.: Transient stability assessment of power system containing series and shunt compensators. IEEE Trans. Power Syst. **22**(3) (2007)
4. Bergen, A.R.: Power System Analysis. Prentice-Hall, New Jersey (1986)
5. Dash, S.S., Panigrahi, B.K.: Nonlinear control of STATCOM for stabilization of synchronous generator. Eng. Model. **16**(3–4), 111–120 (2003)
6. Akhrif, O., Okou, F.-A., Dessaint, L.-A., Champagne, R.: Application of a multivariable feedback linearization scheme for rotor angle stability and voltage regulation of power systems. IEEE Trans Power Syst **14**(2) (1999)
7. Slotine, J.E., Li, W.: Applied Nonlinear Control. Prentice-Hall International Editions, NJ, pp. 207 (1991)
8. Das, S., Biswas, A., Dasgupta, S., Abraham, A.: Bacterial foraging optimization algorithm: theoretical foundations, analysis and applications. IEEE Trans. Evol. Comput. **15**(1) (2011)

9. Vishwakarma, S.; Tripathi, R.K.: Transient energy dissipation and damping improvement using STATCOM & SSSC. In: IEEE Conference Publications on Power, Control and Embedded Systems (ICPCES) (2010)
10. Amara, S., Hsan, H.A.: Power system stability improvement by FACTS devices: a comparison between STATCOM, SSSC and UPFC. In: IEEE Conference Publications on Renewable Energies and Vehicular Technology (REVET) (2012)
11. Suul, J.A., Molinas, M., Undeland, T.: STATCOM-based indirect torque control of induction machines during voltage recovery after grid faults. *IEEE Trans. Power Electron.* **25**(5) (2010)

Common Fixed Points by Using E.A. Property in Fuzzy Metric Spaces

Vishal Gupta and Naveen Mani

Abstract In last some decades, Fuzzy topology has been extensively used in logic programming. It has been noticed by several researchers that, this theory was applied on various logical program to find more truthful result. The strength of fuzzy mathematics lies in its usefulness and having fruitful applications especially outside mathematics. In this paper, we prove some common fixed point theorem by using E.A. property in fuzzy metric spaces. We prove our results in fuzzy metric spaces in the sense of Kramosil and Michalek [1]. Our result generalize and extend relevant result of Mihet [2] and Vijayaraju [3]. An application of finite families of self mappings is given to support our result.

Keywords Fuzzy metric spaces · Common fixed point · Finite families of self mappings · E.A. property

1 Introduction and Preliminaries

A mathematical way to represent uncertainty in day to day life was initiated by Zadeh [4] in 1965, by introducing the concept of fuzzy set. After 1980s, this theory was studied by various authors because of their wide spread applications in diverse areas. Recently, Gregori et al. [5] employed the concept of fuzzy metric spaces to color image processing and also studied several fascinating examples of fuzzy metrics in the sense of George and Veeramani [6].

In topology and analysis, various researchers defined the concept of fuzzy metric spaces in several ways. In 1975, Kramosil and Michalek [1] have

V. Gupta (✉) · N. Mani

Department of Mathematics, Maharishi Markandeshwar University, Mullana,
Ambala 133207, Haryana, India
e-mail: vishal.gmn@gmail.com

N. Mani

e-mail: naveenmani81@gmail.com

introduced the concept of fuzzy metric space by using the notion of continuous triangular norm defined by Schweizer [7].

Definition 1 [7] A binary operation $*$: $[0,1] \times [0,1] \rightarrow [0,1]$ is a **continuous triangular norm** if for all $a, b, c, e \in [0, 1]$ the following condition satisfied:

- I. $*$ is Commutative and Associative
- II. $a * 1 = a$
- III. $*$ is continuous
- IV. $a * b \leq c * e$ whenever $a \leq c$ and $b \leq e$.

Kramosil and Michalek [1] defined fuzzy metric space as follows,

Definition 2 [1] The triplet $(X, M, *)$ is fuzzy metric space if X is an arbitrary set, $*$ is continuous t-norm, M is fuzzy set in $X^2 \times [0, \infty)$ satisfying the following condition.

- I. $M(x, y, 0) = 0$
- II. $M(x, y, t) = 1 \quad \forall t > 0$ iff $x = y$
- III. $M(x, y, t) = M(y, x, t)$
- IV. $M(x, y, t) * M(y, z, s) \leq M(x, z, t + s) \quad \forall x, y, z \in E$ and $t, s > 0$
- V. $M(x, y, \cdot) : [0, \infty) \rightarrow [0, 1]$ is left continuous
- VI. $\lim_{t \rightarrow \infty} M(x, y, t) = 1 \quad \forall x, y \in E$.

In 1988, Grabiec [8] extended the fixed point theorem of Banach [9] to fuzzy metric space,

Theorem 1 [8] Let $(X, M, *)$ be a complete fuzzy metric spaces such that for all $x, y, \in X$ where $0 < k < 1$ if

- I. $\lim_{t \rightarrow \infty} M(x, y, t) = 1$
- II. $M(Fx, Fy, kt) \geq M(x, y, t)$ then F has a unique fixed point.

Later, George and Veeramani [6] modified the concept of fuzzy metric space introduced by Karmosil and Michalek with the help of t-norm. With this introduction, they also succeeded in introducing a Hausdorff topology on such fuzzy metric spaces. Some more metric fixed point results were generalized to fuzzy metric spaces by various author such as Subrahmanyam [10], Vasuki [11], Saini et al. [12, 13], Vishal Gupta and Ashima Kanwar [14] and the references cited there in.

Now we define some important definition and lemmas.

Definition 3 [8] A sequence $\{x_n\}$ in a fuzzy metric space $(X, M, *)$ is said to be convergent to $x \in X$ if $\lim_{n \rightarrow \infty} M(x_n, x, t) = 1 \quad \forall t > 0$

Definition 4 [8] A sequence $\{x_n\}$ in a fuzzy metric space $(X, M, *)$ is called Cauchy sequence if $\lim_{n \rightarrow \infty} M(x_{n+p}, x_n, t) = 1 \quad \forall t > 0$ and each $p > 0$

Definition 5 [8] A fuzzy metric space $(X, M, *)$ is said to be complete if every Cauchy sequence in X converges in X

Example 1 Let $E = N$. Define $a * b = ab \quad \forall a, b \in [0, 1]$ and $t > 0$. Let

$$M(x, y, t) = \begin{cases} \frac{x}{y}, & \text{if } x \leq y \\ \frac{y}{x}, & \text{if } y \leq x \end{cases}$$

then $(E, M, *)$ is a fuzzy metric space.

Lemma 1 [7] *If $(X, M, *)$ is a fuzzy metric space and $\{x_n\}, \{y_n\}$ are sequences in X such that $x_n \rightarrow x, y_n \rightarrow y$, then $M(x_n, y_n, t) \rightarrow M(x, y, t)$ for every continuity point $t \in M(x, y)$.*

Definition 6 [2] *A pair (A, S) of self mappings of a fuzzy metric space $(X, M, *)$ is said to satisfy the E.A. property if there exist a sequence $\{x_n\}$ in X such that for all $t > 0$*

$$\lim_{n \rightarrow \infty} M(Ax_n, Sx_n, t) = 1$$

Clearly a pair of nontrivial compatible mappings enjoys the property (E.A.).

Definition 7 [15] *Two pairs (A, S) and (B, T) of self mappings of a fuzzy metric space $(X, M, *)$ are said to satisfy the common property (E.A.), if there exist two sequences $\{x_n\}, \{y_n\}$ in X such that for all $t > 0$*

$$\lim_{n \rightarrow \infty} M(Ax_n, Sx_n, t) = \lim_{n \rightarrow \infty} M(By_n, Ty_n, t) = 1$$

In our result, we define a class Φ of all mappings $\zeta: [0,1] \rightarrow [0,1]$ satisfying the following conditions:

- I. ζ is increasing on $[0,1]$
- II. $\zeta(t) > t \forall t \in (0, 1]$ and $\zeta(t) = t$ if and only if $t = 1$.

The purpose of this paper is to unify the role of the property (E.A.) and the common property (E.A.) in the existence of a common fixed point for contractive mappings in fuzzy metric spaces. We prove our main result in fuzzy metric spaces in the sense of Kramosil and Michalek [1].

2 Main Result

First we prove the following lemma.

Lemma 2 *Let A, B, S and T be four self mappings of a fuzzy metric space $(X, M, *)$ such that for each $x \neq y \in X$ and $t > 0$*

- I. (A, S) or (B, T) satisfies the E.A property
- II. $A(X) \subset T(X)$ or $B(X) \subset S(X)$
- III. $B(y_n)$ converges for every sequence $\{y_n\}$ in X whenever $T(y_n)$ converges (or $A(y_n)$ converges for every sequence $\{y_n\}$ in X whenever $S(y_n)$ converges)
- IV. for some $\zeta \in \Phi$

$$M(Ax, By, t) \geq \xi(\min\{M(Sx, Ty, t), M(Sx, By, t), M(Ty, By, t), \frac{M(Ax, Sx, t) + M(Ax, Ty, t)}{2}, \frac{M(By, Ty, t) \cdot M(Ax, Ty, t)}{M(Sx, Ty, t)}\}) \quad (1)$$

then the pair (A, S) and (B, T) share the common property (E.A.).

Proof Since the pair (A, S) satisfying the property (E.A.), therefore there exists a sequence $\{x_n\}$ in X such that

$$\lim_{n \rightarrow \infty} Ax_n = \lim_{n \rightarrow \infty} Sx_n = z$$

this implies that $\lim_{n \rightarrow \infty} M(Ax_n, Sx_n, t) = 1$. Since $A(X) \subset T(X)$, there exist a sequence $\{y_n\}$ in X such that $Ax_n = Ty_n = z$.

Now we assert that $\lim_{n \rightarrow \infty} M(By_n, z, t) = 1$. Suppose not, that is $\lim_{n \rightarrow \infty} By_n = v \neq z$ then by using (1), we have

$$M(Ax_n, By_n, t) \geq \xi(\min\{M(Sx_n, Ty_n, t), M(Sx_n, By_n, t), M(Ty_n, By_n, t), \frac{M(Ax_n, Sx_n, t) + M(Ax_n, Ty_n, t)}{2}, \frac{M(By_n, Ty_n, t) \cdot M(Ax_n, Ty_n, t)}{M(Sx_n, BTy_n, t)}\})$$

On letting $n \rightarrow \infty$ and making use of Lemma (1), we get

$$M(z, v, t) \geq \xi(M(z, v, t))$$

As $z \neq v$, we have $0 < M(z, v, t_0) < 1$ for some $t_0 > 0$. Since $M(z, l, \cdot)$ is left-continuous and also nondecreasing, it can have at the most countably many points of discontinuity. Let on contrary that t_0 is a point of continuity of $M(z, l, \cdot)$. Then, in view of condition ξ , we get $\xi(M(z, v, t_0)) > M(z, v, t_0)$, which is a contradiction, implying thereby $z = v$. This shows that the pairs (A, S) and (B, T) share the common property (E.A.).

Now we prove our main result by using common property (E.A.).

Theorem 2 Let A, B, S and T be four self mappings of a fuzzy metric space $(X, M, *)$ such that for each $x \neq y \in X$ and $t > 0$

- I. the pairs (A, S) and (B, T) share the common property (E.A.),
- II. $S(X)$ and $T(X)$ are closed subsets of X ,
- III. for some $\xi \in \Phi$,

$$M(Ax, By, t) \geq \xi(\min\{M(Sx, Ty, t), M(Sx, By, t), M(Ty, By, t), \frac{M(Ax, Sx, t) + M(Ax, Ty, t)}{2}, \frac{M(By, Ty, t) \cdot M(Ax, Ty, t)}{M(Sx, Ty, t)}\}) \quad (2)$$

then the pair (A, S) as well as (B, T) have a coincidence point. Moreover, A, B, S and T have a unique common fixed point in X provided both the pairs (A, S) and (B, T) are weakly compatible.

Proof Since the pairs (A, S) and (B, T) share the common property (E.A.), therefore there exist two sequences $\{x_n\}$ and $\{y_n\}$ in X such that

$$\lim_{n \rightarrow \infty} Ax_n = \lim_{n \rightarrow \infty} Sx_n = \lim_{n \rightarrow \infty} By_n = \lim_{n \rightarrow \infty} Ty_n = z$$

for some $z \in X$. Since $S(X)$ is a closed subset of X , therefore $\lim_{n \rightarrow \infty} Sx_n = z \in S(X)$. Hence there exist a point $u \in X$ such that $Su = z$. Now we Claim that $M(Au, z, t) = 1$. If not then using Condition 2, we get

$$M(Au, By_n, t) \geq \xi \left(\min \{ M(Su, Ty_n, t), M(Su, By_n, t), M(Ty_n, By_n, t), \right. \\ \left. \frac{M(Au, Su, t) + M(Au, Ty_n, t)}{2}, \frac{M(By_n, Ty_n, t) \cdot M(Au, Ty_n, t)}{M(Su, Ty_n, t)} \right)$$

for all $n \in N$. On taking $\lim n \rightarrow \infty$ and making use of Lemma (1), give rise

$$M(Au, z, t) \geq \xi \left(\min \{ M(z, z, t), M(z, z, t), M(z, z, t), \right. \\ \left. \frac{M(Au, z, t) + M(Au, z, t)}{2}, \frac{M(z, z, t) \cdot M(Au, z, t)}{M(z, z, t)} \right)$$

implies,

$$M(Au, z, t) \geq \xi(M(Au, z, t))$$

As $z \neq Au$, therefore $0 < M(z, Au, t_0) < 1$ for some $t_0 > 0$. Since $M(z, Au, \cdot)$ is left-continuous and also nondecreasing, it can have at the most countably many points of discontinuity. If we assume that t_0 is a continuity point of $M(z, Au, \cdot)$. Then, in view of condition ξ , we get $\xi(M(z, Au, t_0)) > M(z, Au, t_0)$, which is a contradiction, therefore $z = Au$. Thus $Au = Su = z$. This shows that u is a coincidence point of the pair (A, S) .

Also $T(X)$ is a closed subset of X , therefore $\lim n \rightarrow \infty Ty_n = z \in T(X)$ Also there exists a point $w \in X$ such that $Tw = z$.

Now we Claim that $M(Bw, z, t) = 1$. If not then using Condition 2, we get

$$M(Au, Bw, t) \geq \xi \left(\min \{ M(Su, Tw, t), M(Su, Bw, t), M(Tw, Bw, t), \right. \\ \left. \frac{M(Au, Su, t) + M(Au, Tw, t)}{2}, \frac{M(Bw, Tw, t) \cdot M(Au, Tw, t)}{M(Su, Tw, t)} \right)$$

or

$$M(z, Bw, t) \geq \xi(M(z, Bw, t))$$

As $z \neq Bw$, therefore $0 < M(z, Bw, t_0) < 1$ for some $t_0 > 0$. Since $M(z, Bw, \cdot)$ is left-continuous and also nondecreasing, it has only (at most) countably many points of discontinuity. Now we may suppose that t_0 is a continuity point of $M(z, Bw, \cdot)$. Then, we get $\xi(M(z, Bw, t_0)) > M(z, Bw, t_0)$, which is a contradiction, therefore $z = Bw$. Thus $Bw = Tw = z$. Hence w is a coincidence point of the pair (B, T) .

Since $Au = Su$ and the pair (A, S) is weakly compatible, therefore $Az = ASz = SAz = Sz$. Now we claim that z is the common fixed point of the pair (A, S) . For this, we show that $M(Az, z, t) = 1$. Suppose not, then by using Condition (2)

$$M(Az, Bw, t) \geq \xi(\min\{M(Sz, Tw, t), M(Sz, Bw, t), M(Tw, Bw, t), \\ \frac{M(Az, Sz, t) + M(Az, Tw, t)}{2}, \frac{M(Bw, Tw, t) \cdot M(Az, Tw, t)}{M(Sz, Tw, t)}\})$$

implies that

$$M(Az, z, t) \geq \xi(\min\{M(Az, z, t), M(Az, z, t), M(z, z, t), \\ \frac{M(Az, Az, t) + M(Az, z, t)}{2}, \frac{M(z, z, t) \cdot M(Az, z, t)}{M(Az, z, t)}\})$$

or

$$M(Az, z, t) \geq \xi\left(\min\left\{1, M(Az, z, t), \frac{1 + M(Az, z, t)}{2}\right\}\right).$$

Now here three cases arises

- I. if $1 < \min\left\{M(Az, z, t), \frac{1 + M(Az, z, t)}{2}\right\}$, then $M(Az, z, t) = 1$. Thus z is the common fixed point of (A, S) .
- II. if $M(Az, z, t) < \min\left\{1, \frac{1 + M(Az, z, t)}{2}\right\}$, then $M(Az, z, t) \geq \xi(M(Az, z, t))$.

As $z \neq Az$, therefore $0 < M(z, Az, t_0) < 1$ for some $t_0 > 0$. As $M(z, Az, \cdot)$ is left-continuous and also nondecreasing, it can have at the most countably many points of discontinuity. If we suppose that t_0 is a continuity point of $M(z, Az, \cdot)$. Then, we get $\xi(M(z, Az, t_0)) > M(z, Az, t_0)$, which is a contradiction, therefore $z = Az$. Thus $M(Az, z, t) = 1$. This shows that z is a common fixed point of the pair (A, S) .

- III. Now if $\frac{1 + M(Az, z, t)}{2} < \min\{1, M(Az, z, t)\}$, then $M(Az, z, t) \geq \xi\left(\frac{1 + M(Az, z, t)}{2}\right)$ then again as $z \neq Az$, therefore $0 < M(z, Az, t_0) < 1$ for some $t_0 > 0$. As $M(z, Az, \cdot)$ is left-continuous and also nondecreasing, it can have at the most countably many points of discontinuity. If we suppose that t_0 is a continuity point of $M(z, Az, \cdot)$. Then, we get

$$\xi \left(\frac{1 + M(Az, z, t)}{2} \right) > \frac{1 + M(Az, z, t)}{2},$$

or

$$M(Az, z, t) \geq \xi \left(\frac{1 + M(Az, z, t)}{2} \right) > \left(\frac{1 + M(Az, z, t)}{2} \right)$$

or

$$2M(Az, z, t) \geq (1 + M(Az, z, t))$$

Thus $M(Az, z, t) = 1$, therefore we can say that z is a common fixed point of the pair (A, S) .

Since $Bw = Tw$ and the pair (B, T) is weakly compatible, therefore $Bz = BTw = TBw = Tz$. Now we claim that z is the common fixed point of the pair (B, T) . For this, we show that $M(Bz, z, t) = 1$. if not, then by using Condition (2)

$$M(Au, Bz, t) \geq \xi \left(\min \{ M(Su, Tz, t), M(Su, Bz, t), M(Tz, Bz, t), \right. \\ \left. \frac{M(Au, Su, t) + M(Au, Tz, t)}{2}, \frac{M(Bz, Tz, t) \cdot M(Au, Tz, t)}{M(Su, Tz, t)} \} \right)$$

implies that

$$M(z, Bz, t) \geq \xi \left(\min \{ M(z, z, t), M(z, Bz, t), M(z, Bz, t), \right. \\ \left. \frac{M(z, z, t) + M(z, z, t)}{2}, \frac{M(Bz, z, t) \cdot M(z, z, t)}{M(z, z, t)} \} \right)$$

or

$$M(z, Bz, t) \geq \xi(\min\{M(z, Bz, t)\}).$$

As earlier we can obtain $z = Az$, which shows that z is a common fixed point of the pair (B, T) . Hence Z is a common fixed point of A, B, S and T . Uniqueness of the common fixed point of is an easy consequence of Condition (1) and condition of ξ . This concludes the proof.

Theorem 3 Let A, B, S and T be four self mappings of a fuzzy metric space $(X, M, *)$ such that for each $x \neq y \in X$ and $t > 0$

- I. (A, S) or (B, T) satisfies the E.A property
- II. $A(X) \subset T(X)$ or $B(X) \subset S(X)$
- III. $B(y_n)$ converges for every sequence $\{y_n\}$ in X whenever $T(y_n)$ converges (or $A(y_n)$ converges for every sequence $\{y_n\}$ in X whenever $S(y_n)$ converges)
- IV. for some $\xi \in \Phi$

$$M(Ax, By, t) \geq \xi \left(\min \{ M(Sx, Ty, t), M(Sx, By, t), M(Ty, By, t), \right. \\ \left. \frac{M(Ax, Sx, t) + M(Ax, Ty, t)}{2}, \frac{M(By, Ty, t) \cdot M(Ax, Ty, t)}{M(Sx, Ty, t)} \right) \quad (3)$$

V. $S(X)$ or $T(X)$ is a closed subset of X .

then the pair (A, S) as well as (B, T) have a coincidence point. Moreover, A, B, S and T have a unique common fixed point in X provided both the pairs (A, S) and (B, T) are weakly compatible.

Proof Following Lemma 2, since the pairs (A, S) and (B, T) share the common property (E.A.), therefore there exist two sequences $\{x_n\}$ and $\{y_n\}$ in X such that

$$\lim n \rightarrow \infty Ax_n = \lim n \rightarrow \infty Sx_n = \lim n \rightarrow \infty By_n = \lim n \rightarrow \infty Ty_n = z \quad (4)$$

for some $z \in X$. Since $S(X)$ is a closed subset of X , therefore $\lim n \rightarrow \infty Sx_n = z \in S(X)$. Hence there exist a point $u \in X$ such that $Su = z$. Now we Claim that $M(Au, z, t) = 1$. If not then using Condition 2, we get

$$M(Au, By_n, t) \geq \xi \left(\min \{ M(Su, Ty_n, t), M(Su, By_n, t), M(Ty_n, By_n, t), \right. \\ \left. \frac{M(Au, Su, t) + M(Au, Ty_n, t)}{2}, \frac{M(By_n, Ty_n, t) \cdot M(Au, Ty_n, t)}{M(Su, Ty_n, t)} \right)$$

for all $n \in N$. On taking $\lim n \rightarrow \infty$ and making use of Lemma (1), give rise

$$M(Au, z, t) \geq \xi \left(\min \{ M(z, z, t), M(z, z, t), M(z, z, t), \right. \\ \left. \frac{M(Au, z, t) + M(Au, z, t)}{2}, \frac{M(z, z, t) \cdot M(Au, z, t)}{M(z, z, t)} \right)$$

implies,

$$M(Au, z, t) \geq \xi(M(Au, z, t))$$

As $z \neq Au$, therefore $0 < M(z, Au, t_0) < 1$ for some $t_0 > 0$. Since $M(z, Au, \cdot)$ is left-continuous and also nondecreasing, it can have at the most countably many points of discontinuity. If we assume that t_0 is a continuity point of $M(z, Au, \cdot)$. Then, in view of condition ξ , we get $\xi(M(z, Au, t_0)) > M(z, Au, t_0)$, which is a contradiction, therefore $z = Au$. Thus $Au = Su = z$. This shows that u is a coincidence point of the pair (A, S) . Since $A(X) \subset T(X)$ and $Au \in A(X)$, there exist $w \in X$ such that $z = Au = Tw$.

Now we Claim that $M(Bw, z, t) = 1$. If not then using Condition 2, we get

$$M(Au, Bw, t) \geq \xi \left(\min \{ M(Su, Tw, t), M(Su, Bw, t), M(Tw, Bw, t), \right. \\ \left. \frac{M(Au, Su, t) + M(Au, Tw, t)}{2}, \frac{M(Bw, Tw, t) \cdot M(Au, Tw, t)}{M(Su, Tw, t)} \right)$$

or

$$M(z, Bw, t) \geq \zeta(M(z, Bw, t))$$

As $z \neq Bw$, therefore $0 < M(z, Bw, t_0) < 1$ for some $t_0 > 0$. Since $M(z, Bw, \cdot)$ is left-continuous and also nondecreasing, it has only (at most) countably many points of discontinuity. Now we may suppose that t_0 is a continuity point of $M(z, Bw, \cdot)$. Then, we get $\zeta(M(z, Bw, t_0)) > M(z, Bw, t_0)$, which is a contradiction, therefore $z = Bw$. Thus $Bw = Tw = z$. Hence w is a coincidence point of the pair (B, T) .

Now rest of the proof can be completed by following the same arguments as in Theorem 1.

As an application of Theorem 2, we have the following result for four finite families of self mappings.

Theorem 4 *Let $A_1, A_2, \dots, A_m, B_1, B_2, \dots, B_m, S_1, S_2, \dots, S_m$ and T_1, T_2, \dots, T_m be four finite families of self mappings of a fuzzy metric space (X, M, \cdot) with $A = A_1A_2\dots A_m, B = B_1B_2\dots B_m, S = S_1S_2\dots S_m$ and $T = T_1T_2\dots T_m$ satisfying Condition 2 and pairs (A, S) and (B, T) share the common property (E.A.). If $S(X)$ and $T(X)$ are closed subsets of X , then*

I. the pairs (A, S) and (B, T) have a coincidence point each.

Moreover, A_i, S_k, B_r and T_t have a unique common fixed point provided the pairs of families $(\{A_i\}, \{S_k\})$ and $(\{B_r\}, \{T_t\})$ commute pairwise, where $i \in \{1, 2, \dots, m\}, k \in \{1, 2, \dots, n\}, r \in \{1, 2, \dots, p\}$ and $t \in \{1, 2, \dots, q\}$.

Proof Proof follows on the lines of corresponding result given in [16].

3 Conclusion

Theorem 2 generalizes the result of Mihet [2] and Vijayaraju [3] because it does not required the completeness of spaces and containment amongst range sets of involved mappings.

References

1. Kramosil, I., Michalek, J.: Fuzzy metric and Statistical metric spaces. *Ky-bernetica* **11**, 326–334 (1975)
2. Mihet, D.: Fixed point theorems in fuzzy metric spaces using property (E.A.). *Nonlinear Anal.* **73**, 2184–2188 (2010)
3. Vijayaraju, P., Sajath, Z.M.I.: Some common fixed point theorems in fuzzy metric spaces. *Int. J. Math. Anal.* **3**, 701–710 (2009)
4. Zadeh, L.A.: Fuzzy sets. *Inform. Control* **8**, 338–353 (1965)
5. Gregori, V., Morillas, S., Sapena, A.: Examples of fuzzy metrics and applications. *Fuzzy Sets Syst.* **170**, 95–111 (2011)

6. George, A., Veeramani, P.: On some results in fuzzy metric spaces. *Fuzzy Sets Syst.* **64**, 395–399 (1994)
7. Schweizer, B., Sklar, A.: Statistical metric spaces. *Pacific J. Math.* **10**, 313–334 (1960)
8. Grabiec, M.: Fixed point in fuzzy metric spaces. *Fuzzy Sets Syst.* **27**, 385–389 (1988)
9. Banach, S.: Sur les operations dans les ensembles abstraits et leur application aux equations integrals. *Fundamenta Mathematicae.* **3**, 133–181 (1922)
10. Subrahmanyam, P.V.: A common fixed point theorem in fuzzy metric space. *Inform. Sci.* **83**, 109–112 (1995)
11. Vasuki, R.: A common fixed point theorem in fuzzy metric space. *Fuzzy Sets Syst.* **97**, 395–397 (1998)
12. Saini, R.K., Vishal, Gupta: Fuzzy version of some fixed points theorems on expansion type maps in fuzzy metric space. *J. Math. Assoc. Thai.* **5**, 245–252 (2007)
13. Saini, R.K., Vishal, Gupta: Common coincidence points of R-weakly commuting fuzzy maps. *J. Math. Assoc. Thai.* **6**, 109–115 (2008)
14. Vishal, Gupta, Ashima, Kanwar: Fixed point theorem in fuzzy metric spaces satisfying E.A property. *Indian J. Sci. Technol.* **5**, 69–71 (2012)
15. Liu, Y., Jun, Wu, Li, Z.: Common fixed points of single-valued and multi-valued maps. *Int. J. Math. Math. Sci.* **19**, 3045–3055 (2005)
16. Imdad, M., Ali, J.: Some common fixed point theorems in fuzzy metric spaces. *Math. Commun.* **11**, 153–163 (2006)
17. El Naschie, M.S.: A review of E-infinity theory and the mass spectrum of high energy particle physics. *Chaos, Solitons Fractals* **19**, 209–236 (2004)

Use of Evolutionary Algorithms to Play the Game of Checkers: Historical Developments, Challenges and Future Prospects

Amarjeet Singh and Kusum Deep

Abstract The objective of this paper is to study the historical development of computer programmers for playing the game of checkers. Since the game-playing is a NP-hard problem, it would be interesting to use evolutionary algorithms to solve them. The question is can a programme be developed which can beat humans with complete success, it may appears that some challenges may also be formed which may substantiate the argument of the paper. Further, these challenges also form a part of this study.

Keywords Checker game · Evolutionary algorithms · Chinook

1 Introduction

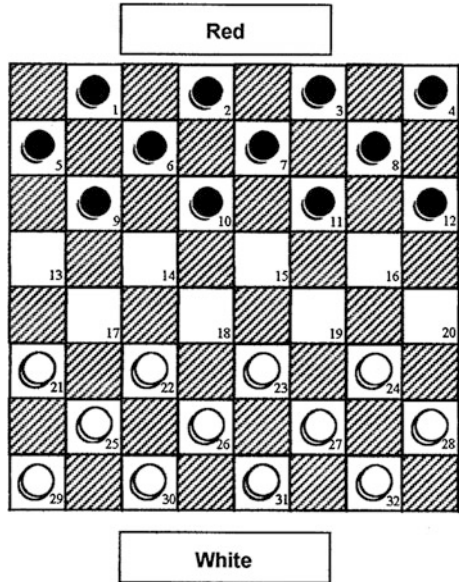
Checkers is one of the world's oldest, two players board game, played on square checker board which consists 64 light and dark squares. Checkers is widely played in the United States and the British Commonwealth. Each player has 12 playing pieces which are disk-shaped. Normally, the colours of pieces are red and white.

In the beginning of the game, each contestant has 12 pieces arranged on the board. The red pieces occupy square 1–12 and the white pieces occupy 21–32. Red moves first, players take their turn by advancing a piece diagonally forward to an adjoining vacant square. If an opponent's piece is in an adjoining diagonally vacant square, with a vacant space beyond, it must be captured and removed by

A. Singh (✉) · K. Deep
Department of Mathematics, Indian Institute of Technology Roorkee,
Roorkee 247667 Uttarakhand, India
e-mail: amarjeetitr@gmail.com

K. Deep
e-mail: kusumdeep@gmail.com

Fig. 1 Checker board



diagonally jumping over it to the empty square. If this square presents the same situation, successive jumps forward in a straight or zigzag direction must be completed. If there is more than one way to jump, then player has a choice. If a piece reaches the king row (last row) then it can move backwards also. A player will win when an opponent’s all pieces are captured or blocked. A game is declared draw, when neither side can force a victory nor the trend of play becomes repetitive (Fig. 1).

2 Historical Developments

The development of board game programs started during and after Second World War, the original work began on chess. The first work on chess was done Shannon [1]. In 1950, a paper on computer chess entitled “Programming a computer for playing chess” by Shannon was published, which describes how a machine or computer could be made to play a game of chess. Minimax procedure, based on evaluation function of a given chess position, was used. In evaluation function, the value of the black position was subtracted from that of the white position. Material was counted as 1, 3, 5 and 9 points for pawn, knight or bishop, rook and queen respectively. Some positional factor, subtracting 0.5 point for each doubled pawn, backward pawn and isolated pawn and mobility was added with 0.1 point for each legal move available. He gave the king artificial value of 200 points and used only checkmate to capture of the king.

Dietrich G. Prinz, colleague of Alan Turing, developed the first limited chess program in 1951. The computer, Ferranti Mark 1, could not play full game powerfully but was able to find best moves if it was only two moves away from checkmate.

Alan Turing, the father of modern computing, experimented machine routing for playing chess. Turing had no computer to run his heuristic chess program. Turing simulated the operation of the program by hand, using paper and pencil. The output of this program was not so good but his work was appreciated as this was the first time somebody defined rules to play a game like chess.

Samuel [2] wrote the first checkers program for IBM 701 and learning program in 1955 and discussed two procedures: (1) a rote learning procedure in which a record was kept of the board situation encountered in actual play together with information as to the results of the machine analysis of the situation; this record could be referenced at terminating board situations of each newly initiated tree search and, in effect, allow the machine to look ahead further than time would otherwise permit, (2) a generalization learning procedure in which the programme continuously re-evaluated the coefficients for the linear polynomial used to evaluate the board positions at the terminating board situations of a look-ahead tree search. Minimax algorithm is used to backup scores assigned to the terminating situations and to select the best move and assumed that opponent would also apply the same selection rules on his turn. The rote learning procedure, learn continuously but was very slow and effective in the opening and end game phases of the play. Two copies of his program played against each other and the weaker program was eliminated, and produced a new second copy. After repetition of this process, a stronger program was produced. In 1962, Samuel's program beat Robert Nealey, a blind checkers master, but lost 8 games in a row in 1966 against Walter Hellman and Derek Oldbury.

The shortcomings of this program may be highlighted as follows:

1. The incorrectness of the assumption of linearity which underlie the use of a linear polynomial.
2. The inadequacies of the heuristic procedures used to prune and to terminate the tree search.
3. Slowness of the learning procedure.
4. The absence of any strategy considerations for altering the machine mode of play in the light of the tactical situations as they develop during play and
5. The absence of an effective machine procedure for generating new parameters for the evaluation procedure.

The signature table technique was also introduced by Samuel [3] to express the observed relationship between parameters in subsets; values read from the tables for a number of subsets are then combined for the final evaluation.

In 1970, at Duke University [4], a checkers program was written by Eric Jensen, Tom Truscott and Alan Bierman, which bet Samuel's program in a 2-game match, and went on to lose against grandmaster Elbert Lowder in a 5-game match with 2 losses, 2 draws and a win.

In 1989 and 1990, three strong programs emerged during the first and second computer Olympiads that was held in London (1) Colossus (Martin Bryant) (2) Checkers (Gil Dodgen) (3) Chinook (Jonathan Schaeffer et al.). The programs Colossus and Checkers, commercially available, were aimed at PC market and both were ranked among the top twenty players in the world that time, while Chinook was a research project. Jonathan Schaeffer made a team of people to work with him on Chinook. Norman Treolar was, checkers team expert, responsible for the evaluation function and the opening book; Robert Lake was in charge of the endgame databases. Later on, Martin Bryant supplied a very big and very good opening book for Chinook. Endgame databases, distance to win database, introduced in Chinook, which gave the computer perfect knowledge for all positions with 8 pieces or less on the board of the form win/loss/draw. Chinook uses some fixed predefined strategies for game start and game end section. In the middle of the game, Chinook uses evaluation function in which expert knowledge was captured. Chinook was based on game tree and alpha-beta search (traditional game theory mechanics) and expert knowledge. The win/loss/draw scheme is interesting, because it has to store less information and therefore you can compress these databases much more. The 8-piece database turned out to be about 6 GB in compressed form.

Initially, the American Checker Federation and English Draughts Association, the governing bodies for checkers, did not allow the Chinook-Tinsley (world champion) match and said that the world championship was for humans, not for machines but at the end, they agreed to allow the match and gave the title “Man-Machine World Championship” to that match.

In 1992, Chinook-Tinsley match was played but Chinook lost with 2 wins and 4 losses [5–7]. In 1994, after introducing the 8-piece database, a rematch was started. Due to bad health, Tinsley confiscated the match after 6 drawn games, he died shortly afterwards. Chinook went on to beat Don Lafferty, the second best player in the world after Tinsley in 1995 in a very close match (1 win and 31 draws), and finished clear first far ahead of the field in the 1996 national tournament in the US.

Researchers, working in the field of artificial intelligence, want to create intelligent game program which is capable of defeating human experts. Several approaches have been proposed for different games including neural networks for Backgammon, special purpose hardware called Deep Blue for chess and the application of expert knowledge with relatively small computational power for checkers. Recently introduced methods to play the checker game, mostly based on a learning theory to generate an intelligent agent, and learn how to play without prior knowledge. Mostly, these techniques depend on expert knowledge and training evaluation function, relevance factor for the evaluation, the weight of the evaluation factors, opening knowledge, and an end game database.

Chisholm and Bradbeer [8] used genetic algorithm to optimize the board evaluation function of a game playing program and a pool of checkers programs was played against each other in a round-robin tournament to evaluate the fitness of each player and genetic algorithm was used to preserve and improve the best

performance. One of the greatest achievements in the realm of checkers game is Blondie24.

Chellapilla and Fogel [9], in which after injecting little expert knowledge into the algorithms, Blondie24, was able to play a game of checkers at the human expert level. To find the potentially good moves, Fogel used a minimax search tree as a look ahead and evolution strategies with neural networks. In Blondie24, the strategies do not all play the same number of game because some would be selected as opponents more often than others.

Franken and Engelbrecht [10] investigated the effectiveness of various particle swarm optimizer structure to learn and how to play the game of checkers and used co-evolutionary techniques to train the game playing agents, performance was compared against the player making moves at random. In co-evolutionary process, population of game playing agents consist a neural network and genetic makeup, translate to the neural network weights and updates to produce stronger game playing agent based on the evolutionary strategy. During each play's move, a single ply depth game tree was constructed and NN was used as an evaluation function for the leaf node of the game trees. After taking different swarm sizes, varying hidden layers sizes, comparison in performance was made and the highest scores of the agent were 82.4.

Hughes [11] used an on-line evolutionary algorithm to co-evolve move sets for both players in the game, playing the entire length of the game tree for each evaluation and chromosome was defined such that each chromosome has 100 genes, each per move. Each gene, a real value in the range [0,1), was decoded into a move by multiplying the real value by the total number of currently available moves, rounding the nearest integer and using this as an index for a single move from the list of available moves.

In "Evolutionary approach to the game of checkers" by Kusiak et al. [12] presented a new method of genetic evolution of linear and non linear evaluation functions in the game of checkers, in which evolutionary heuristic generators in two player games domain was evaluated using the game of US checkers. Twenty five components, based on basic board features, were used in the evaluation functions. Linear and non linear heuristic evaluations were considered. Each linear heuristic consisting of linear combination of the parameter set and non linear heuristics were composed of a small number of IF-conditions which divided the entire game into disjoint stages. At each stage, linear combination of parameters was considered and the coefficients of linear combination were optimized by evolutionary process.

Dura and de Oliveira [13] used board features such as material strength, piece mobility, pawn structure, king safety and control of the centre in parameterized board evaluation function whose weight are optimized using PSO and through simulation result shows that PSO can provide faster learning results than simulated annealing under similar experimental conditions, especially in the presence of bounded computing time.

Al-khateeb [14], in his Ph.D. thesis, introduced a round robin tournament into the evolutionary phase of evolutionary checkers program to eliminate the

randomness and to enhance its learning ability. Individual and social learning were also introduced and n-tuple systems into evolutionary checkers produced a good player which learn faster and take less computational time in comparison to other approaches. Piece difference has a significant effect on learning abilities, shown through experiments. Thirty feed forward neural network players was played against each others for 140 generations and obtained best player was tested against an evolutionary program based on Blondie24.

“Immune based fuzzy agent plays checkers game” by Cheheltani and Ebadzadeh [15] discussed immune based approach, in which permanent memory cells cause to omit the process of learning for any played strategy and consequently increase the speed of decision making process. In this method, memory cells represent actions that have the best local payoff, for that current state of the game and are generated simultaneously by learning process. Because of these cells decision making system decide better, considering the previous and future state of the game.

Based on Fogel’s work to evolve Blondie24, Al-Khateeb and Kendall [16] evolved different checkers players. Through experiments, many checkers players were produced of various depths of ply during learning. Through three different experiments, it was shown that players with higher ply perform better than those with a lower ply, when playing at fixed ply of six, players trained with higher plies also performed better than the players trained with a lower ply. When playing at the ply that they were trained at and players trained on higher ply perform significantly worse when playing at a lower ply. Also results show that increasing the depth level by one will give a different performance depending upon the level.

Generally, the games draw; when the world’s best players play the game of checkers. To avoid this and make the game more competitive, two-move ballot is used. In which first move of both players are discussed. There are 49 possibilities to play first two moves. After a lot of research on first two moves, it is found that 6 out of 49 are unbalanced, as it gives an advantage to one side over the other. So, only 43 moves are considered. A card is randomly chosen, showing which of 43 moves is to be played. The original game, with no forced opening moves, is called Go-As-You-Please (GAYP).

Rating of checkers players is calculated by standard system formula, in which initial rating for a player R_0 is 1,600.

Rating of opponent:

$$R_{new} = R_{old} + C(outcome - W).$$

Where outcome: outcome of match 1 for a win, 0.5 for a draw and 0 for a loose.

R_{new} , R_{old} are new rating for the player and current player’s rating and is calculated by the formula:

$$W = 1/(1 + 10^{((R_{opp}-R_{old})/400)}),$$

R_{opp} : Opponent’s rating and C is 32 for ratings less than 2,100, 24 for rating between 2,100 and 2,399, and 16 for ratings at or above 2,400.

A player is called expert, master and senior master if its rating are in 2,000–2,199, 2,200–2,399 and 2,400+ respectively; and class A, B, C, ..., J if the player has rating in 1,800–1,999, 1,600–1,799, 1,400–1,599, ..., below 200.

3 Conclusions and Future Challenges

The discussion in the context reveals the fact that people often enjoy playing board games because it does not only place some intelligent challenge before the player, it also evinces a sense of satisfaction after playing well. People use knowledge and search to make their decisions during this process. Without perfect knowledge, mistakes are made and even world champions occasionally loose the match.

Many matches were played between checkers program and checker masters and in which checker master have been defeated however the checker game is not perfectly solved i.e. there is no checker program that may always win whenever be the match being played. It is a challenge to develop a tireless, robust, unbeatable Checker player program, which is available to the user at any time to play.

With the advent of some new nature inspired optimization techniques being invented, there is a wide scope of research in this area where the newly developed nature inspired optimization techniques can be used to design program for playing board game like checkers.

References

1. Shannon, C.E.: Programming a computer for playing chess. *Phil. Mag.* **41**(314), 256–275 (1950)
2. Samuel, A.L.: Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 210–229 (1959)
3. Samuel, A.L.: Some studies in machine learning using the game of checkers. II—recent progress. *IBM J. Res. Dev.* **11**(6), 601–617 (1967)
4. Fierz, M.: A brief history of computer checkers. <http://www.fierz.ch/history.htm>
5. Schaeffer, J.: Man versus machine: The silicon graphics world checkers championship, technical report 92-120, Department of Computing Science, University of Alberta (1992)
6. Schaeffer, J., Treloar, N., Lu, P., Lake, R.: Man versus machine for the world checkers championship. *AI Mag.* **14**(2), 28–35 (1993)
7. Schaeffer, J., Lake, R., Lu, P.: Chinook the world man-machine checkers champion. *AI Mag.* **17**, 21–30 (1996)
8. Chisholm, K.J., Bradbeer, P.V.: Machine learning using a genetic algorithm to optimise a draughts program board evaluation function. In: *IEEE International Conference on Evolutionary Computation*, pp. 715–720 (1997)
9. Chellapilla, K., Fogel, D.B.: Evolving an expert checkers playing program without using human expertise. *IEEE Trans. Evol. Comput.* **5**(4), 422–428 (2001)
10. Franken, N., Engelbrecht, A.P.: Comparing PSO structures to learn the game of checkers from zero knowledge. In: *The 2003 IEEE Congress on Evolutionary Computation, CEC'03*, vol. 1, pp. 234–241 (2003)

11. Hughes, E.J.: Checkers using a co-evolutionary on-line evolutionary algorithm. In: The 2005 IEEE Congress on Evolutionary Computations, vol. 2, pp. 1899–1905 (2005)
12. Kusiak, M., Wałędzik, K., Mańdziuk, J.: Evolutionary approach to the game of checkers. In: Adaptive and Natural Computing Algorithms, pp. 432–440. Springer Berlin Heidelberg (2007)
13. Duro, J.A., De Oliveira, J.V.: Particle swarm optimization applied to the chess game. In: IEEE World Congress on Computational Intelligence, IEEE Congress on Evolutionary Computations, pp. 3702–3709 (2008)
14. Al-Khateeb, B.: Investigating evolutionary checkers by incorporating individual and social learning, N-tuple systems and a round robin tournament. PhD Dissertation, University of Nottingham, UK (2011)
15. Cheheltani, S.H., Ebadzadeh, M.M.: Immune based fuzzy agent plays checkers game. *Appl. Soft Comput.* **12**(8), 2227–2236 (2012)
16. Al-Khateeb, B., Kendall, G.: Effect of look-ahead depth in evolutionary checkers. *J. Comput. Sci. Technol.* **27**(5), 996–1006 (2012)
17. Schaeffer, J.: *One Jump Ahead: Computer Perfection at Checkers*. Springer, New York (2009)

Development of Computer Aided Process Planning System for Rotational Components Having Form Features

D. Sreeramulu, D. Lokanadham and C. S. P. Rao

Abstract This paper presents a feature based Computer Aided Process Planning (CAPP) system for rotational components. While developing a feature based CAPP system, a set of flexible process plans were encountered and a population based heuristics namely Genetic Algorithm (GA) is used to obtain a most optimal process plan for a defined objective function. The objective function for optimization of process plan is the minimization of manufacturing score. Proposed methodology has been implemented for an example part having different rotational features.

Keywords Rotational parts · CAPP · Genetic algorithm

1 Introduction

The main objective of any manufacturing industry is to produce high quality products at the lowest possible cost. To achieve this goal, manufacturers had to adapt new production techniques and were forced to look forward for alternatives to the traditional approaches to design and manufacturing because of increased cost and competition. The use of computers in various fields such as design, control and manufacturing has been of particular importance [1] Experience in the

D. Sreeramulu (✉) · D. Lokanadham
Department of Mechanical Engineering, Aditya Institute of Technology and Management,
Tekkali, India
e-mail: dowlurusreeram@gmail.com

D. Lokanadham
e-mail: dloknadham@gmail.com

C. S. P. Rao
Department of Mechanical Engineering, National Institute of Technology, Warangal, India
e-mail: csp_rao63@yohoo.com

use of various computer based systems in industry has shown that while each can individually benefit the production process. CAPP can help to reduce the planning time and increase the consistency and efficiency for producing new products by any manufacturing industry. CAPP integrates automation of product design with that of manufacturing by linking CAD and CAM.

Literature on CAPP is huge and a considerable amount of work has been carried out on the CAPP over the last few decades. In the past Altung and Zhang [2], Steudel [3], Weill et al. [4] and Cay and Chassapis [5] have made an extensive study on CAPP. Zhang et al. [6] presented an approach that deals with process planning problems using Genetic Algorithms in a concurrent manner in generating the entire solution space by considering the multiple decision making activities, i.e., operation selection, machine selection, setup selection, cutting tool selection, and operations sequencing, simultaneously. In this paper the authors proposed a CAPP system for the rotational parts having form features and the sequence of operations are optimized using GA. Since we have chosen a rotational part with form features, the process plan evaluation criterion chosen for optimization is minimization of manufacturing scores, which includes the change of datum features, feature adjacency and preference in processing the features.

When the various decision-making tasks are carried out simultaneously, process planning becomes a combinatorial problem. During the last decade, GA's have been applied to many combinatorial problems. These include job shop scheduling [7, 8], the traveling salesman problem (TSP) [9], and other NP-complete problems [10]. A process planning problem (PPP) is similar to a TSP in that every operation has to be traversed once and once only, although a PPP is more complicated due to precedence constraints among operations and nonfixed "distance" between operations (time required for machine, setup, and tool change). It is expected that GA's can provide a valid option for solving the PPP's so long as a suitable string representation and a corresponding search operator can be devised. Recently, there have been reports on applying GA's to process planning [11–13]. These developed GA's, however, still suffer from the above-mentioned problems. The system presented in this paper is developed to address these problems comprehensively and effectively.

1.1 Classification of Turning Features

The form features of rotational parts are mainly classified into primary, secondary and C-axis features [14]. The primary features given the overall shape to part. Both secondary and C-axis features reside on the primary features and add details to the part. The hierarchical structure of typical form features of a rotational part is shown in Fig. 1.

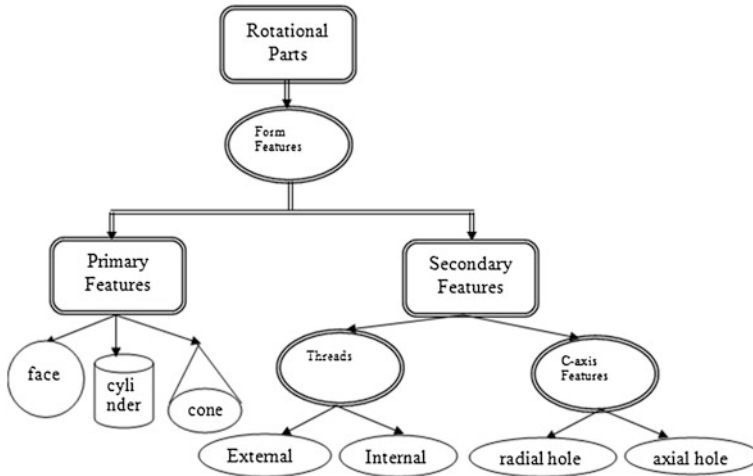


Fig. 1 Hierarchical structure of form features of a rotational part

1.2 Classification of Turning Operation Based on Feature

Turning operations can be classified as longitudinal turning, facing, angle cutting, and profiling. Owing to the fact that the rotational parts are symmetrical along their axis, designing their processes can be done according to the symmetry axis so that feature recognition process is performed on the axis of symmetry [15]. Features including rotational parts could be classified as outside features and inside features according to machining attributes. These features are also classified into sub features by the system such as long turning and grooving according to machining attributes. Some of the typical features like cylindrical surface, conical surface, toroidal surface, threads and axial holes on rotational components can be generated by the following turning operations and can be shown in Fig. 2.

Straight Turning: Reducing the diameter of a part to a desired dimension can generate cylindrical surfaces, e.g. cylindrical surface

Threading: Threads are cut using lathes by advancing the cutting tool at a feed exactly equal to the thread pitch, e.g. Threads on cylindrical surface

Taper Turning: Cutting tool has a simple shape, but the feed motion is complex; cutting tool is fed along a contour thus creating a contoured shape on the workpiece. Producing tapers on a lathe is a specific task and contour turning is just one of the possible solutions, e.g. Conical surface

Form Turning: In form turning cutting tool has a shape that is imparted to the workpiece by plunging the tool into the workpiece. Feed of the cutting in the form turning is linear, e.g. Toroidal Surface

Axial Drilling: Making a hole/holes along axial direction with the use of a drill bit and tailstock of the lathe, e.g. axial holes

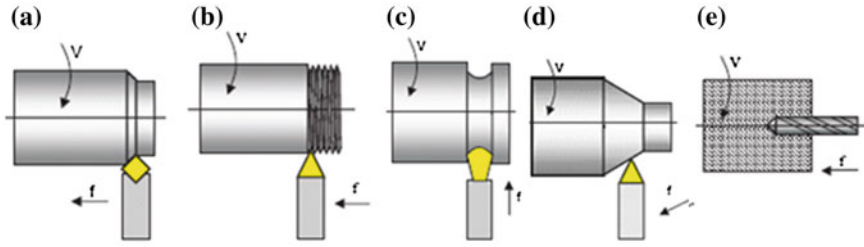


Fig. 2 Types of turning operations. **a** Cylindrical. **b** Threading. **c** Toroidal. **d** Conical. **e** Axial holes

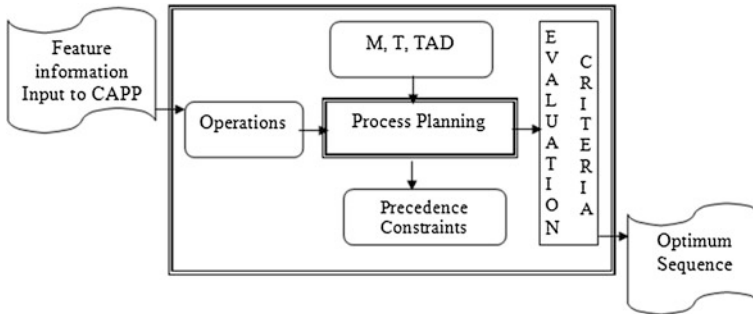


Fig. 3 Framework of CAPP

1.3 Computer Aided Process Planning

Process Planning is an engineering task that determines the detailed manufacturing requirements for transforming a raw material into a completed part, with the available machining resources. The output of process planning generally includes operations, machine tools, cutting tools, fixtures, machining parameters, etc. With the advent of computer technologies, there is a general demand for CAPP systems to assist human planners and achieve the integration of CAD and CAM. Moreover, the introduction of new manufacturing technologies, such as CIM system and design for manufacturing (DFM), causes an even greater demand for automated CAPP systems. To support production scheduling, process plans with alternative routes and sequences must be generated to suit the changes in the workshop.

Figure 3 shows the framework of CAPP system. Process planning activities includes interpretation of features, selection of process to manufacture the component, selection of machine (M), tool (T) and Tool Approach Direction (TAD), determination of optimum sequence of operations without violating precedence constraints based on evaluation criteria chosen. Features are the input to the system and it will give the optimum sequence as output.

2 Development of CAPP System for Rotational Parts

Process planning problem for a given part can be defined as selection of operations and operation sequences utilizing the set of resources available in the shop-floor so as to convert raw material into a finished part [6]. This section describes the development of CAPP system for the rotational parts having primary and secondary features.

2.1 Process Planning Model to Generate Initial Feasible Sequences

For a given rotational part, the process planning model is developed by the following steps.

Step-I: Operation selection: Map all the features into operation(s) and list out all such machining operations for each feature, e.g. feature-1 is turning, feature-2 is threading etc.

Step-II: Precedence constraints: Generation of operation sequence is a complex task that is influenced by the constraints imposed by the features and possible interactions among the features. These constraints are given below:

- (i) **Locating constraint:** It is concern with the examination of defined part features to determine which reference face is to be used to locate a particular feature. It refers to plane surfaces such as end or a face. The locating surface is to be machined prior to the reference surface. For example, for a rotational part, end face could be the reference feature.
- (ii) **Accessibility constraint:** It establishes precedence constraint among the associated features, since a secondary feature is defined as the one, which resides on primary feature. It is not possible to machine the secondary feature until the primary feature is formed.
- (iii) **Non-destructive constraint:** This is taken into consideration so that the current machined feature do not destroy the properties of features machined previously. For example, when thread and chamfer are two secondary features on a cylinder, chamfering must proceed threading.
- (iv) **Geometric tolerance constraint:** This refers to the datum requirements on features according to the geometric tolerance scheme. It results in the identification of features which must be machined in same set-up, otherwise it would increase set-up time or cost. A feasible sequence is the one which does not violates any of the above feasibility constraints.

Step-III: Algorithm for Initial Sequence Generation:

1. Select an operation randomly, which has no constraint.
2. From the remaining operations select an operation at random such that it obeys the precedence relationship.
3. Continue step 2 till all the operations are finished.

2.2 Optimization Using GA

GA is a population based search method requiring domain specific knowledge to solve an optimization problem. In GA, a candidate solution is represented by sequence of numbers known as chromosome or string. In this work each element in string represents a operation to be done. The order of elements in string represents the sequence to be followed. A prudently selected set of feasible strings from feature precedence makes the initial population. Optimization these sequences are done by GA using appropriate defined fitness evaluation.

Fitness Evaluation: Since the holding and adjacent features plays a major role while machining the rotational components, minimization of manufacturing scores is taken as evaluation criteria for optimization. The optimality criteria may include change of datum/reference features, feature adjacency. The change of datum/reference features represents abstract setup changes. It is preferable to machine as many features as possible in one setup. The feature adjacency criterion refers to abstract tool changes. If the features are adjacent, it is possible to reduce tool changes and idle tool motions [14].

- (i) **Holding score:** This score is based on change of datum/reference features for holding the part for a given sequence:

$$U_h = \text{number of changes in each sequence} \quad (1)$$

- (ii) **Adjacency score:** This score is based on the adjacency of features and it is evaluated for each sequence based on the number of elements not matching with the sequence given in each template:

$$U_a = \sum_i [\text{no. of out of sequence features in adjacency} - \text{template (i)}] \quad (2)$$

where i represents the template under consideration.

- (iii) **Total score:** For each sequence total score is calculated as the weighted sum of individual scores

$$\text{Total score, } U = w_1(U_h) + w_2(U_a) \quad (3)$$

The assignment of weights w_1 and w_2 depends on the user preference. In this work w_1 and w_2 are taken to be 3 and 2 respectively.

The objective function is to minimize the total score.

The Objective function = $\text{Min}(U)$

The fitness value for each string in population is calculated as follows

$$\text{Fitness value, } \mathbf{f} = U_{\max} - U_i. \quad (4)$$

where

U_{\max} maximum score of a string in a population and

U_i score of a string i in a population

After evaluation of each string based on its fitness value, these strings are further optimized by applying three major GA operators namely reproduction, cross over and mutation.

Reproduction: The selection of better strings is based on the actual count arrived in initial sequence. The generic operator is used to generate new population that has better strings than old population is called parent-1 and is used for the next generic operation i.e. crossover.

Crossover: The strings obtained from reproduction are then mated at a given probability (crossover rate). To ensure that the crossover will not result in any violation of precedence constraints's and each operation in the offspring is executed once and only once, the cyclic crossover operator proposed by **Dagli et al. [16]**, is adopted. The algorithm is described as follows.

Algorithm for Crossover of String 1 and String 2:

1. Determine a cut point randomly from the all the positions of a string. Each string is then divided into two parts, the left side and the right side, according to the cut point.
2. Copy the left side of string-1 to form the left side of offspring-1. The operator constructs the right side of offspring-1 according to the order of operations in string-2.
3. Copy the left side of string-2 to form the left side of offspring-2. The operator constructs the right side of offspring-2 according to the order of operations in string-1.

Cyclic Crossover selects a string pair with Probability P_c select a cut point randomly. This process is illustrated with an example shown in Fig. 4. A pair of strings, parent-1 and parent 2, is under the crossover operation in which the cut point is chosen between positions 3 and 4. The left side of parent-1, op4-op1-op2, is used to form the left side of offspring-1. The order of the right side of parent-1, op5-op7-op3-op6, is adjusted according to the order of parent-2 to form the right side of offspring-1. By doing so, the sequences among the operations in both parent-1 and parent-2 are maintained in offspring-1. A similar operation is applied to parent-2 and parent-1 to form offspring-2.

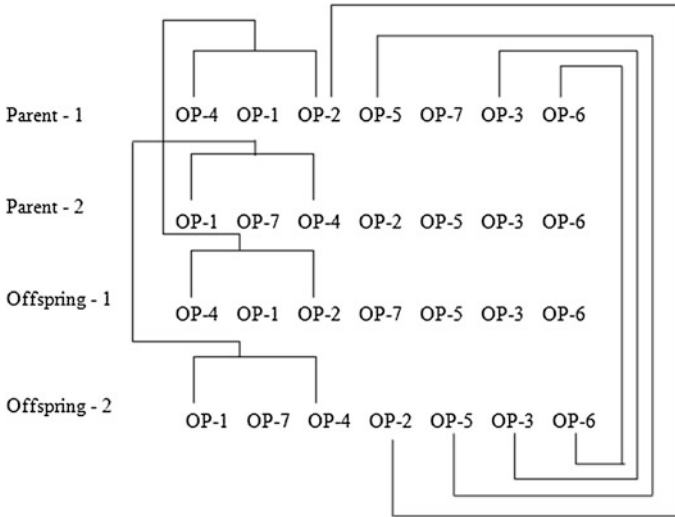


Fig. 4 An example of applying the cyclic crossover

Mutation: The mutation operator makes random changes in one or more elements of the string. Mutation is done with a small probability (mutation probability). This is done to protect loss of some potentially useful strings. The mutation operator proposed for rotational parts randomly modifies two elements to obtain the resulting population.

GA Parameters chosen

The settings of GA parameters were as follows.

- Population size = 100
- Crossover probability $P_c = 0.7$
- Mutation probability $P_m = 0.35$
- Stopping criterion: 6,000 generation was selected.

3 Illustrative Example

This section describes the implementation of the proposed methodology of feature based CAPP system for a rotational part.

An example of rotational part [17] having form features is shown Fig. 5 (Two views were shown to view all the features).

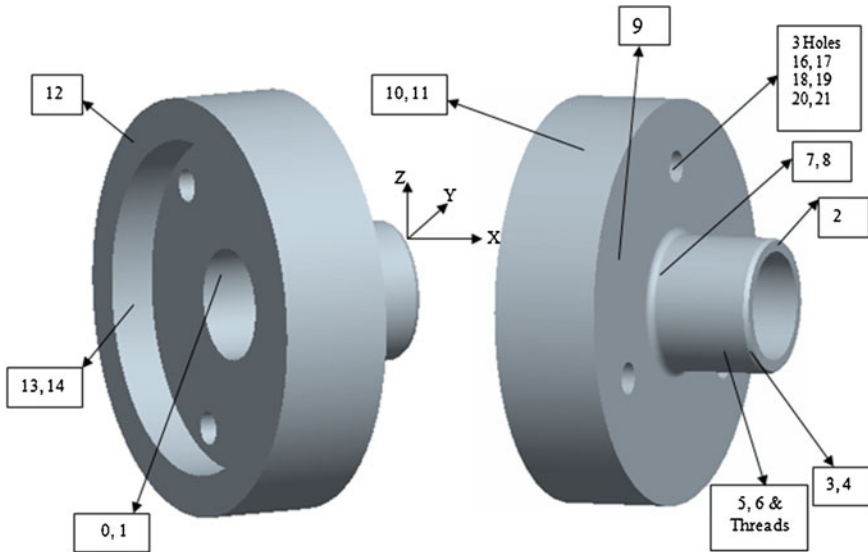


Fig. 5 Example part

3.1 Selection of Operations from Feature List

An operation candidate(s) has to be assigned to all the features. The cylindrical feature having maximum radius is mapped into external cylinder and rest of the cylindrical features are mapped into step turning.

Conical feature is mapped into chamfering, since it is located at the extreme end and having a smaller inclined length i.e. 1 mm. Grooving operation is assigned to toroidal feature and drilling operation is assigned to all the holes. The list of operations are given in Table 1, where columns 1 and 2 show the features and the operations respectively.

3.2 Precedence Constraints

The precedence relation among all the operations is shown in Table 2. It is developed based on the locating, accessibility, non destruction and geometric tolerance constraints given by **Gopala Krishna** [17].

The details of holding the feature and adjacency templates are shown in Table 3.

Table 1 Operation selection for the example part

Features	Operation candidates
F-1, Face (Face-12)	Facing (op1)
F-2, Face (Face-2)	Facing (op2)
F3-Cylindrical surface (Face No. 10&11)	External cylinder (op3)
F4-Cylindrical surface (Face No. 5&6)	Step turning (op4)
F5-Hole (Face No. 13&14)	Drilling (op5)
F6-Hole (Face No. 0&1)	Drilling (op6)
F7-Conical surface (Face No. 3&4)	Chamfering (op7)
F8-Thread (Face No. 5&6)	Threading (op8)
F9-Toroidal Surface (Face No. 7&8)	Groove (op9)
F10-(3-Holes)	Drilling (op10)
F-11, Face(Face-9)	Facing-3 (op11)

Table 2 Precedence relationship between operations

	Op1	Op2	Op3	Op4	Op5	Op6	Op7	Op8	Op9	Op10	Op11
Op1			1		1	1					1
Op2				1			1	1	1	1	1
Op3											1
Op4							1	1	1		1
Op5											
Op6					1						
Op7								1			
Op8											
Op9								1			1
Op10											
Op11											

Table 3 Details of holding and adjacency templates

Features								
Set-1:	1	3	6	5				
Set-2:	4	9	7	8	11	10	2	
Adjacency								
Templet-1:	1	3						
Templet-2:	2	4	11					

3.3 Generation of Optimum Sequence Against Evaluation Criterion

The proposed methodology for optimizing sequence of operations using GA is coded in C++ and executed in P-IV computer. The most optimum sequence has been generated based on the defined objective function (discussed in Sect. 2.2) for the following two cases.

Table 4 Optimum sequence in case-1

Sequence	1	3	6	5	2	4	10	7	9	11	8
----------	----------	----------	---	---	----------	----------	----	---	---	-----------	---

Holding score = 1, adjacency score = 1, total score = 5

Table 5 Optimum sequence in case-2

Sequence	1	3	6	5	4	2	9	10	11	7	8
----------	----------	----------	---	---	----------	----------	----------	----	----	---	---

Holding score = 1, adjacency score = 0, total score = 3

Case 1: Minimization of manufacturing score for holding and adjacency templates shown in Table 3.

The string having the least manufacturing score is shown in Table 4.

The optimum value obtained here is 5 units and this value can further be optimized by making a small change in template-2 of adjacency.

Case 2: Minimization of manufacturing score by changing the Template-2 of adjacency with out violating precedence.

In this second case, OP-11 in Template-2 of adjacency is replaced with OP-9 (*Modified Template-2: 2 4 9*). The optimum sequence having the score of 3 units is shown in Table 5

The manufacturing score obtained in case-2 is less than the case-1, hence it would be better to chose the template of adjacency proposed in case-2 and we have also evaluated the results with Differential Evaluation method (This work is not reported in this paper) and found a improvement in the result and hence GA is used as a toll in all the CAPP systems.

4 Conclusions

In this paper an attempt is made successfully to develop feature based CAPP systems for rotational parts. The features were mapped into operation candidate(s) and feature precedence is established to generate initial sequence of operations. Among these alternate sequences, an optimum sequence was generated based on defined objective function using GA. Genetic algorithm (GA) is used as a meta heuristic to perform search in the state of space of alternate process plans and obtain an optimized process plan based on minimization of manufacturing score for rotational parts. A number of cases were studied to test the proposed method of CAPP and successful runs were recorded in all our trials.

References

1. Ssemakula, M.E.: Role of Process Planning in Integration of CAD/CAM Systems. SRC-TR, pp. 87–108 (1986)
2. Alting, L., Zhang, H.: Computer-aided process planning: the state-of-the-art survey. *Int. J. Prod. Res.* **27**(4), 553–585 (1989)
3. Steudel, H.J.: Computer-aided process planning: past, present and future. *Int. J. Prod. Res.* **22**(2), 253–266 (1984)
4. Weill, R., Spur, G., Eversheim, W.: Survey of computer aided process planning systems. *Ann. CIRP* **31**(2), 539–551 (1982)
5. Cay, F., Chassapis, C.: An IT view on perspectives of computer aided process planning. *Comput. Ind.* **34**, 307–337 (1997)
6. Zhang, F., Zhang, Y.F., Nee, A.Y.C.: Using genetic algorithms in process planning for job shop machining. *IEEE Trans. Evol. Comput.* **1**(4), 278–289 (1997)
7. Davis, L.: Job shop scheduling with genetic algorithms. In: Grefenstette, J.J. (ed.) *Proceedings of 1st International Conference on Genetic Algorithms and Their Applications*, pp. 136–140. Lawrence Erlbaum, Hillsdale, NJ (1985)
8. Biegel, J.E., Daver, J.: Genetic algorithms and job scheduling. *Comput. Ind. Eng.* **19**, 81–91 (1990)
9. Suh, J.Y., Van Gucht, D.: Incorporating heuristic information into genetic search. In: Grefenstette, J.J. (ed.) *Proceedings of 2nd International Conference on Genetic Algorithms*, pp. 100–107. Lawrence Erlbaum, Hillsdale, NJ (1987)
10. De Jong, K.A., Spears, W.M.: Using genetic algorithms to solve NP complete problems. In: Schaffer, J.D. (ed.) *Proceedings of Third International Conference on Genetic Algorithms*, pp. 124–132. Morgan Kaufmann, San Mateo, CA (1989)
11. Vancza, J., Markus, A.: Genetic algorithms in process planning. *Comput. Ind. Eng.* **17**, 181–194 (1991)
12. Awadh, B., Sepehri, N., Hawaleshka, O.: A computer-aided process planning model based on genetic algorithms. *Comput. Oper. Res.* **22**(8), 841–856 (1995)
13. Yip-Hoi, D., Dutta, D.: A genetic algorithm application for sequencing operations in process planning for parallel machining. *IIE Trans.* **28**, 55–68 (1996)
14. Shunmugam, M.S., Mahesh, P., Bhaskara Reddy, S.V.: A method of preliminary planning for rotational components with C-axis features using genetic algorithm. *Comput. Ind.* **48**, 199–217 (2002)
15. Kalpakjian, S., Schmid, S.R.: *Manufacturing Engineering and Technology*. Addison-Wisely Longman, Singapore (2000)
16. Dagli, C., Sittisathanchai, S.: Genetic neuro-scheduler for job shop scheduling. *Comput. Ind. Eng.* **25**(1/4), 267–270 (1993)
17. Gopala Krishna, A.: Selection of optimal sequence of machining operation sequence in CAPP. *J. Manuf. Eng.* **2**, 4 (2007)

Comparative Analysis of Energy Efficient Protocols for Prolonged Life of Wireless Sensor Network

Gagandeep Singh, H. P. Singh and Anurag Sharma

Abstract The efficiency of Wireless Sensor Networks (WSNs) depends on the routing protocols used, since routing protocol provide the best possible data transmission route from sensor nodes to sink to save energy of nodes in the network. The clustering schemes enhance the network lifetime, raise the scalability and reduce the energy consumption of the sensor network. The work in this paper presents the comparative analysis of the energy efficient routing protocols for WSN such as SEP, TSEP and DSEP. The optimized routing protocol has been proposed on the basis of the network life time, stability and cluster head selection for efficient working of the sensor networks.

Keywords Wireless sensor networks · Clustering · SEP · TSEP · DSEP

1 Introduction

Wireless sensor networks are composed of many homogeneous or heterogeneous sensor nodes with restricted resources. A single sensor node is the combination of three components: processor, sensor and wireless communication device (transceiver) as presented in Fig. 1. These sensor nodes spread throughout it to observe, collect, and transmit data. The sensors are economical, simple and their power supply is irreplaceable. Early study on wireless sensor networks generally focused on technologies based on the homogeneous wireless sensor network in

G. Singh (✉)

Department of Electronics and Communication Engineering, Quantum School of Technology, Roorkee, India
e-mail: er.gagan89@gmail.com

H. P. Singh · A. Sharma

Department of Electronics and Communication Engineering, CT Institute of Engineering Management and Technology, Jalandhar, India

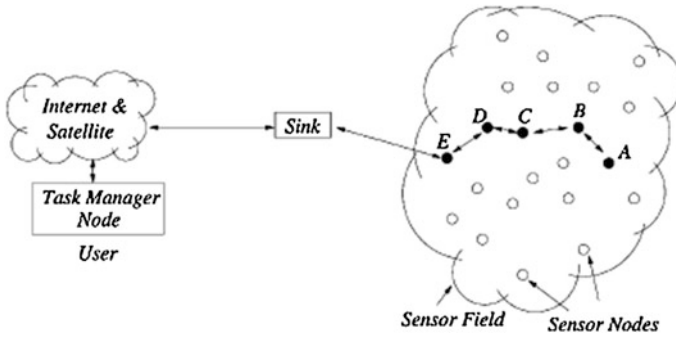
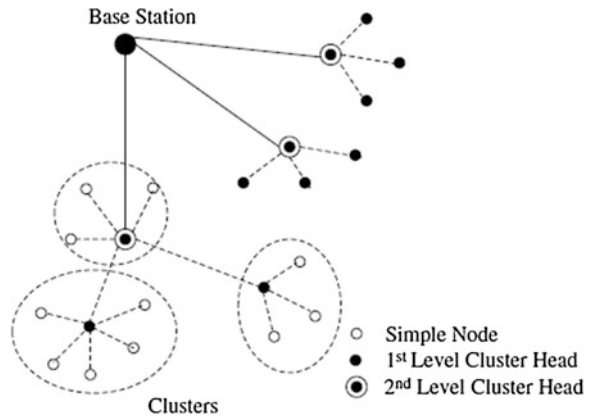


Fig. 1 Wireless sensor network [12]

Fig. 2 Nodes communicate to base station through cluster heads



which each node have same energy. Nevertheless, now a days heterogeneous wireless sensor networks are becoming more and more popular. The heterogeneous nodes can extend network lifetime and improve network reliability without extensively increasing the cost [1, 2]. A classic heterogeneous wireless sensor networks consists of a large number of normal nodes and a few heterogeneous nodes. The normal node is inexpensive and source-constrained whose major tasks are to sense and issue data report. The heterogeneous node, which provides data fusion, filtering and transport, is more costly and more proficient [3]. In a hierarchical topology, nodes are organized into number of clusters according to specific requirements or metrics and perform different tasks in WSNs. Clustering in WSNs contains grouping nodes into clusters and electing a cluster head. The member nodes of a cluster can communicate with their cluster head directly. The cluster head is able to forward the combined data to the central base station with the help of other cluster heads [4].

The nodes having higher energy can act as cluster head which carry out the task of data processing and information transmission, whereas nodes having low energy are normal nodes which carry out the task of information sensing as shown in Fig. 2.

In the previous few years, a comparatively large number of hierarchical clustering routing protocols have been developed for WSNs. In this paper, we broadly review and significantly discuss the most vital hierarchical clustering routing algorithms that have been used for WSNs. The related work has been presented in the next section and the brief description the energy efficient routing protocols has been presented in the [Sect. 3](#). The [Sect. 4](#) presents the comparative results and discussion. The last section concludes the work and presents the future work.

2 Related Work

LEACH is the first clustering based routing protocol for WSNs. This protocol relates arbitrary rotation of cluster head to equally distribute energy load between sensor nodes to improve stability period and network lifetime [5]. LEACH is basically designed for homogeneous networks in which the nodes having equal initial energy.

PEGASIS protocol [6] uses greedy algorithm to organize the sensor nodes into a sequence chain, thus every node transmits and receives the data from the nearby neighbor. But if the node fails, discovering of new route becomes complicated as each node has a set route before the transmission towards the base station.

Stable Election Protocol (SEP) [7] uses weighted election probabilities for every node to become cluster-head according to their respective energy. The cluster head selection is randomly selected and distributed based on the fraction of energy of every node. SEP considered two types of nodes, normal nodes and advanced nodes respectively and also uses two levels of heterogeneity. This protocol is basically designed to raise time interval of first dead node.

TSEP [8] is a reactive routing protocol in which nodes have three different levels of energies. Cluster heads selection is threshold based which causes increase in stability period and network life.

A Deterministic-SEP (D-SEP) [9] is proposed, for selecting cluster heads in a circulated fashion in two-, three-, and multi-level hierarchical wireless sensor networks. This protocol uses improved SEP algorithm for wireless sensor networks in the presence of energy heterogeneity. By using a heterogeneous three-level node setting in a clustering algorithmic approach, nodes select themselves as cluster heads based on their energy levels, retaining uniformly distributed energy among all sensor nodes.

3 Routing Protocols for WSN

In this section, we analyze three standard WSN clustering routing algorithms in detail and present a broad and vital survey of important clustering routing protocols for WSNs. For the purpose of this study, we use similar radio communication and consumption model as reported in [7].

3.1 SEP

SEP protocol [7] is based on two levels of heterogeneity. A fraction ‘ m ’ of total ‘ n ’ nodes is given with an extra energy factor ‘ A ’, which is known as, advanced nodes. Therefore, probability of normal node and advanced node to become cluster head is $p_{nrm} = p/1 + mA$ and $p_{adv} = p * (1 + A)/1 + mA$ respectively, where ‘ p ’ is the best possible probability of every node to become cluster head. Cluster heads selection in SEP is completed arbitrarily on the basis of probability of every type of node as in LEACH. Nodes sense data and transmit it to associated cluster head which transmit it to base station. By increasing ‘ p_{adv} ’ or ‘ m ’, we can further improve our system. Thus, SEP results in better stability period and network life due to advance nodes however two-level heterogeneity also caused enlarged throughput [8].

3.2 TSEP

TSEP (Threshold sensitive Stable Election Protocol) is based on three levels of heterogeneity and has a reactive routing protocol [8]. Advance nodes having energy greater than all other nodes; intermediate nodes have energy in between normal and advance nodes whereas the remaining nodes are normal nodes. Intermediate nodes can be elected with ‘ x ’, a part of nodes which are intermediate nodes and by using the relation that energy of normal nodes is ‘ μ ’ times additional than that used for the normal nodes. In the case of SEP energy considered for normal nodes is E_o , for advance nodes it is $E_{adv} = E_o(1 + A)$ and energy for intermediate nodes can be calculated as $E_{in} = E_o(1 + \mu)$, where $\mu = A/2$. As a result, the total Energy of all the nodes will become $n * E_o(1 + m * A + x * n)$ where, ‘ n ’ is number of nodes, ‘ m ’ is fraction of advanced nodes to entire number of nodes ‘ n ’ having energy greater than remaining of nodes and ‘ x ’ is fraction of intermediate nodes. The best possible probability of nodes, which are separated on the basis of energy, to be selected as a cluster head can be computed by using following formulas:

$$P_{nrm} = p / (1 + m * A + x * n) \quad (1)$$

$$P_{int} = p * (1 + \mu) / (1 + m * A + x * n) \quad (2)$$

$$P_{adv} = p * (1 + A) / (1 + m * A + x * n) \quad (3)$$

Thus, to ensure that cluster head selection is made in the similar method as we have assumed, we have taken an additional factor into consideration, which is threshold level. Every node produces randomly a number inclusive of 0 and 1, when produced value is less than threshold then this node turns into cluster head [5].

For every this types of node, there are different formulas for the computation of threshold depending upon their probabilities, which are shown below:

$$T_{nrm} = p_{nrm}/[1 - p_{nrm}(r \bmod 1/p_{nrm})], \quad \text{if } n_{nrm} \in G' \quad (4)$$

$$T_{int} = p_{int}/[1 - p_{int}(r \bmod 1/p_{int})], \quad \text{if } n_{int} \in G'' \quad (5)$$

$$T_{adv} = p_{adv}/[1 - p_{adv}(r \bmod 1/p_{adv})], \quad \text{if } n_{adv} \in G''' \quad (6)$$

Here G' , G'' and G''' are the set of normal nodes, intermediate nodes and advanced nodes that has not turn out to be cluster heads in the past respectively.

3.3 DSEP

The threshold value is modified in DSEP by using the residual energy and set as [9]:

$$T_{(S_i)} = [p_i/1 - (p_i * (r \bmod 1/p_i))] * [E_{residual} + (r_s \bmod 1/p_i) * (1 - E_{residual})] \quad (7)$$

Now ' r_s ' is the number of successive rounds in which a node has not been cluster-head. If ' r_s ' reaches the value of $1/p_i$, the threshold $T(S_i)$ is reset to the value. Therefore, the possibility of node ' n ' to become cluster head improves because of a high threshold. Moreover, ' r_s ' is reset to '0' if a node becomes cluster head. So, we ensure that data is reached to the base station as long as nodes are alive.

In weighed election probability of three- level heterogeneity nodes known as normal, advanced and intermediate nodes are measured based on partial difference in their initial energy level. At this time the reference value of ' p_i ' is different for nodes. The probabilities of normal, advanced and intermediate nodes are [9–11]:

$$P_i = \{ P_{nrm}, P_{int}, P_{adv} \} \quad (8)$$

where

$$P_{nrm} = p * E_{residual} / (1 + m * A + x * n) * E_{average} \quad (9)$$

$$P_{int} = p * (1 + \mu) * E_{residual} / (1 + m * A + x * n) * E_{average} \quad (10)$$

$$P_{adv} = p * (1 + A) * E_{residual} / (1 + m * A + x * n) * E_{average} \quad (11)$$

Threshold value for cluster head selection is considered for normal, advanced, intermediate nodes by putting above values in Eq. (7) or else it is zero. G' , G'' and G''' is the set of normal and advanced nodes.

4 Results and Discussion

In our present work, a comparative analysis of DSEP with SEP and TSEP protocols on the basis of network lifetime, stability period and throughput is achieved after creating a $100\text{ m} \times 100\text{ m}$ region of 100 sensor nodes spread randomly. The sink or base station is located at the center point (50×50). The packet size that the nodes send to their cluster heads as well as the combined packet size that a cluster head sends to the sink is set to 4,000 bits. The parameters used in the simulation are mentioned below in Table 1.

4.1 Stability Period

Stability period is defined as the time interval from the start of network operation until the death of the first sensor node. Figure 3 illustrates the number of dead nodes for $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$ over 5,000 rounds. It is observed that for SEP with two types of nodes having different initial energy, the first sensor node dies at the round of around 863 whereas due to availability of more nodes with extra energy in TSEP, the first sensor node dies at the round of around 1,184 which is more than SEP. It is observed that in DSEP, stability period is greater than SEP and TSEP protocols as the first node dies at around 1,417.

Further, for $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$ over 6,000 rounds, it is observed from the Fig. 4, that for the SEP the first sensor node dies at the round of around 1,132 whereas the first sensor node dies for TSEP is at around 1,420 which is again more than SEP. And the first sensor node for DSEP is around 1,483 which is greater than SEP and TSEP. So, it shows that in DSEP, stability period is greater than TSEP and SEP.

4.2 Network Lifetime

Network lifetime is defined as the time interval from the start of operation (of the sensor network) until the death of the last alive node. Figure 5 illustrates the lifetime of the sensor network for $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$ over 5,000 rounds. We can observe that for the SEP protocol the last sensor node dies at 2,565 and for TSEP protocol, the last sensor node dies 3,905 rounds whereas for DSEP the last sensor node still alive over 5,000 rounds. It shows that the lifetime of network for DSEP is more than TSEP and SEP.

Further, for $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$ over 6,000 rounds, it is observed from the Fig. 6, that for the SEP the last sensor node dies at the round of around 3,120 and the last sensor node dies for TSEP is at around 3,732 which is

Table 1 Simulation parameters

Parameters	Value
Network field	(100, 100)
Number of nodes	100
E_o (Initial energy of nodes)	0.5 J
Message size	4,000 bits
E_{elec}	50 nJ/bit
E_{amp}	0.0013 pJ/bit/m ⁴
E_{fs}	10 nJ/bit/m ²
E_{DA}	5 nJ/bit/signal
D_o (Threshold distance)	70 m

Fig. 3 Number of dead nodes for $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$

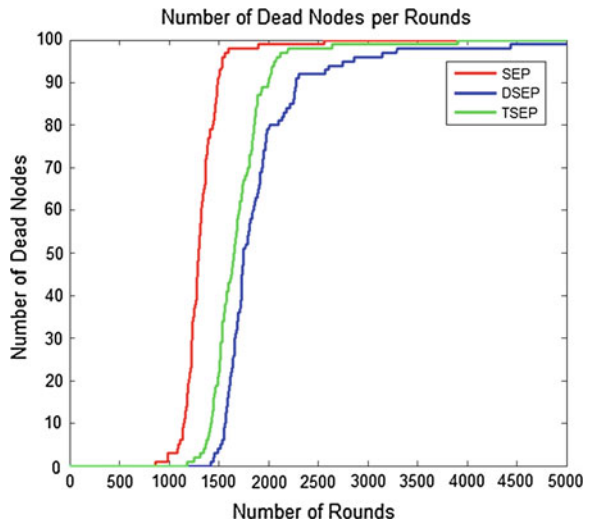


Fig. 4 Number of dead nodes for $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$

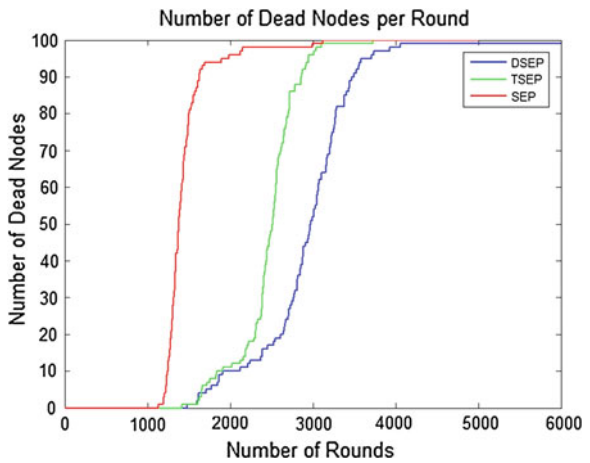


Fig. 5 Lifetime of the sensor network for $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$

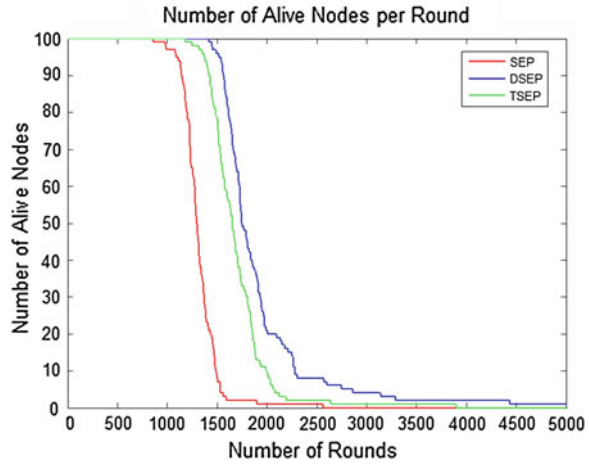
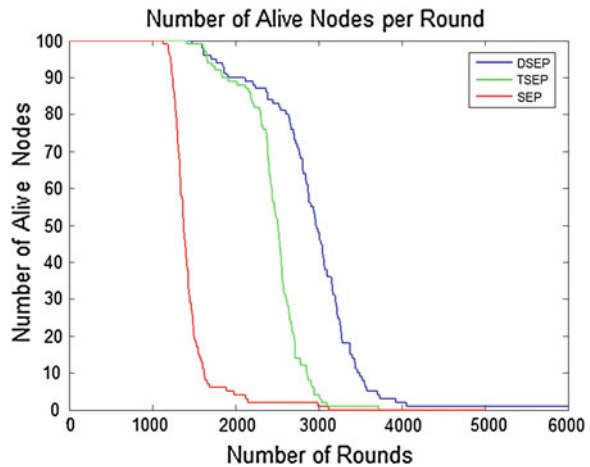


Fig. 6 Lifetime of the sensor network for $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$



again more than SEP whereas for DSEP the last sensor node still alive over 5,000 rounds. It is again observed that the lifetime of network for DSEP is more than TSEP and SEP.

4.3 Throughput of the Network

Figures 7 and 8 illustrate the throughput of the sensor network in terms of the cluster heads alive to send the received packets to the base station or sink. We can observe that for the existing SEP protocol having two types of sensor nodes, the throughput is stable enough up to around 1,104 rounds for the values $m = 0.4$,

Fig. 7 Throughput of the sensor network for $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$

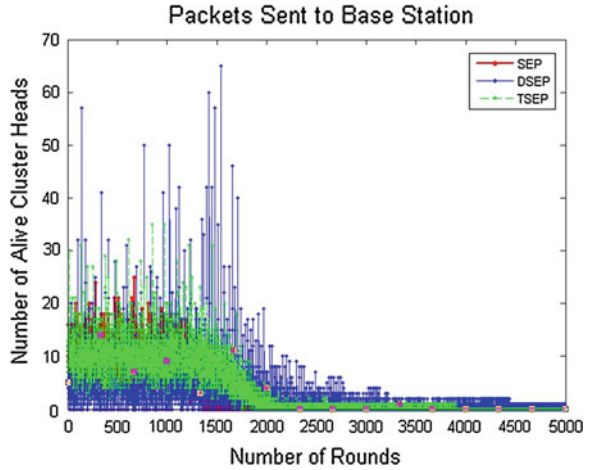
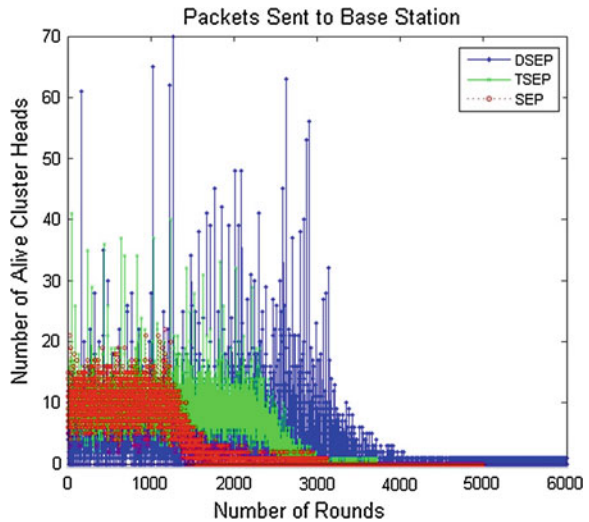


Fig. 8 Throughput of the sensor network for $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$



$A = 1$ and up to 1,456 rounds for the values $m = 0.6$, $A = 1.5$. Moreover, for the TSEP protocol having three types of sensor nodes, the throughput is stable enough up to 1,783 rounds for the values $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$ and up to 2,550 for the values $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$. But with the increasing number of rounds the network throughput decreases whereas for the DSEP protocol, the throughput is stable enough up to 2,100 rounds for the values $m = 0.4$, $A = 1$, $\mu = 0.5$, $x = 0.2$ and up to 3,300 for the values $m = 0.6$, $A = 1.5$, $\mu = 0.75$, $x = 0.3$.

5 Conclusion

SEP is based on weighted election probabilities of every node to become cluster head according to the remaining energy and TSEP is reactive routing protocol in which nodes have three different levels of energies. Cluster heads choice is threshold based, due to three levels of heterogeneity. A comparative analysis of DSEP, which is based on the weighted probabilities to get the threshold for normal, super and advanced nodes points us to elect the cluster head in every round, provides a longer network lifetime, stability period and higher throughput as compared to the existing SEP and TSEP protocols. For the future scope the present work can be extended as an advanced DSEP to deal with clustered wireless sensor networks with more than three levels of hierarchy and more than three types of nodes.

References

1. Kumar, R., Tsiatsis, V., Srivastava, M.B.: Computation hierarchy for in-network processing. In: Proceedings of the 2nd International Workshop on Wireless Networks and Application, San Diego, CA (2003)
2. Rhee, S., Seetharam, D., Liu, S.: Techniques for minimizing power consumption in low data-rate wireless sensor networks. In: Proceedings of IEEE Wireless Communications and Networking Conference, Atlanta, GA (2004)
3. Yu, L., Wang, N., Zhang, W.: Deploying a heterogeneous wireless sensor network. In: Proceedings of International Conference on WiCom, pp. 2588–2591. Singapore (2007)
4. Younis, O., Krunz, M., Ramasubramanian, S.: Node clustering in wireless sensor networks: recent developments and deployment challenges. In: IEEE Network, pp. 20–25. (2006)
5. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Hawaii International Conference on System Science (HICSS-33), pp. 1–10. (2000)
6. Lindsey, S., Raghavenda, C.S.: PEGASIS: power efficient gathering in sensor information system. In: Proceedings of the IEEE Aerospace Conference on Big Sky, pp. 1125–1130. Montana, March 2002
7. 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 (SANPA 2004), 2004
8. Kashaf, A., Javaid, N., Khan, Z.A., Khan, I.A.: TSEP: threshold-sensitive stable election protocol for WSNs. In: Proceedings of 10th International Conference on Frontiers of Information Technology, pp. 164–168. (2012)
9. Bala, M., Awasthi, L.: Proficient D-SEP protocol with heterogeneity for maximizing the lifetime of wireless sensor networks. *Int. J. Intell. Syst. Appl.* **7**, 1–15 (2012)
10. Saini, P., Sharma, A.K.: Energy efficient scheme for clustering protocol prolonging the lifetime of heterogeneous wireless sensor networks. *Int. J. Comput. Appl.* (0975–8887) **6**, 31–36 (2010)
11. Mao, Y., Liu, Z., Zhang, L., Li, X.: An effective data gathering scheme in heterogeneous energy wireless sensor networks. In: Proceedings of International Conference on Computational Science and Engineering, pp. 338–343. (2009)

12. Islam, M.M., Matin, M.A., Mondol, T.K.: Extended stable election protocol (SEP) for three level hierarchical clustered heterogeneous WSN. In: Proceedings of IET Conference on WSS, London, 1–4 (2012)
13. Liu, X.: A survey on clustering routing protocols in wireless sensor networks. *Sensors* **12**, 11113–11115
14. Manjeshwar, E., Agrawal, D.P.: TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: Proceedings of the 15th International Parallel and Distributed Processing Symposium (IPDPS), San Francisco, CA, USA, 2009–2015 (2001)
15. Li, Q., Qingxinand, Z., Mingwen, W.: Design of a distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput. Commun.* **29**, 2230–2237 (2006)
16. Khan, A.A., Javaid, N., Qasim, U., Lu, Z., Khan, Z.A.: HSEP: heterogeneity-aware hierarchical stable election protocol for WSNs. In: Proceedings of Seventh International Conference on Broadband, Wireless Computing, Communication and Applications, pp. 373–378. (2012)
17. Jaiswal, V.P.N., Garg, A.K.: An efficient protocol for reducing energy consumption in wireless sensor networks. *Int. J. Eng. Res. Appl.* **2**, 530–533 (2012) (ISSN: 2248-9622)

Condition Monitoring in Induction Motor by Parameter Estimation Technique

P. Kripakaran, A. Naraina and S. N. Deepa

Abstract This paper addresses a new approach for rotor parameter estimation of induction motors. Condition monitoring of electric motors avoids unexpected motor failures and greatly improves system reliability and maintainability. These are very important issues in motor-driven and power-electronics systems since they can greatly improve the reliability, availability, and maintainability of the system. Induction motors are critical components in many industrial processes early fault diagnosis and Condition monitoring can increase machinery availability and performance, reduces consequential damage, prolong machine life (Mirahl et al., Condition monitoring of squirrel-cage induction motors fed PWM -based drives using a parameter estimation approach, 2004) [1], reduce spare parts and breakdown maintenance (Siddique et al., IEEE Trans. Energy Convers., 20:106–114, 2005) [2]. A reliable parameter estimation technique for induction motors is critical for the development of high-performance drive systems, and it can also be utilized for condition monitoring applications as well. An accurate parameter estimation technique can also be used for motor condition monitoring purposes. In this paper, a simple and reliable technique, based on parameter estimation methods, is introduced for rotor broken bar fault detection (Watson, The use of line current as a condition monitoring tool for three phase induction motors) [3].

Keywords Condition monitoring · Recursive least square · Rotor fault

P. Kripakaran (✉)

Department of EEE, Rajiv Gandhi College of Engineering and Technology,
Kirumambakkam, India
e-mail: kripa_pt@yahoo.co.in

A. Naraina · S. N. Deepa

Department of EEE, Anna University Regional Centre, Coimbatore, India
e-mail: Narainagoa@gmail.com

S. N. Deepa

e-mail: deepapsg@gmail.com

1 Introduction

Condition Monitoring (CM) is an important issue in many fields, including railways, power delivery, and electrical machines and motors. An Unexpected fault or shutdown can result in a serious accident and financial loss for the company. Energy companies must find ways to avoid failures, minimize downtime, reduce maintenance costs, and lengthen the lifetime of their equipment. With reliable condition monitoring, machines can be utilized in a more optimal fashion. It can be defined as a technique or process of monitoring the operating characteristics of a machine so that changes and trends of the monitored signal can be used to predict the need for maintenance before a breakdown or serious deterioration occurs, or to estimate the current condition of the machine. (Or) Condition monitoring is defined as the continuous evaluation of the health of the plant and equipment throughout its service life [4].

1.1 Advantages of CM

The following are the major advantages of performing condition monitoring in induction machines.

- Increases Reliability of the machines.
- Increases machinery availability.
- Decreases loss of production due to faulty motors.
- Increases machinery performance.
- Reduces consequential damage due to faults in machines.
- It prolongs machine life.
- Reduces spare parts.
- Reduces breakdown maintenance.

Through Fig. 1 shown below we might think in spite of all protective equipments in modern days what will be the reason for destruction of the motor. The end result in this picture is that the motor is completely destroyed, but what started it? Was it a bearing fault, excessive starts, or a poor ventilation system? In this example, the rotor was locked when it started and could not reach running speed. The resultant high currents overheated the rotor, stator, shaft and other components in the motor. Inspection later revealed that the overloads in the power circuit it had failed and did not trip the motor, resulting in complete destruction of the motor. There were multiple problems in this situation, which led to the destruction of the motor. If condition monitoring has been carried out in this case complete destruction of the motor could have been avoided and hence a huge amount of cost might have been saved. Thus CM is desirable for such cases. This motivates the need for developing new reliable and efficient CM methods to avoid such damages.

Fig. 1 Typical fault in induction motor



1.2 Classification of Faults in Induction Motor

The most prevalent faults in Induction Motor are briefly categorized as

- Rotor faults
- Bearing faults
- Eccentricity faults
- Stator faults.

The surveys indicate that in general, failures in electrical machines are dominated by bearing and stator faults with rotor winding problems being less frequent. Figure 2 shows the statistical spread in the various dominant mechanisms.

2 Condition Monitoring Using Parameter Estimation Method

2.1 Introduction

There are many papers regarding rotor condition monitoring, and rotor fault detection. Some of these works have mainly used the frequency spectrum of the stator current for rotor condition monitoring. The rotor magnetic field orientation pendulous oscillation, due to broken bars, was recently presented as an index for rotor fault diagnostic purposes. There are other techniques which are based on the artificial intelligence and data mining methods as well as some investigations based on parameter estimation or parameter identification techniques.

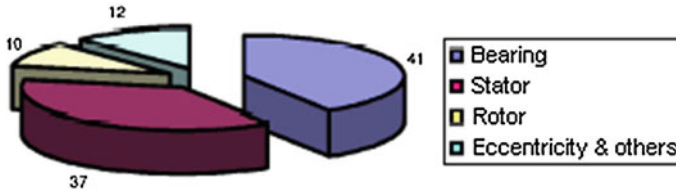


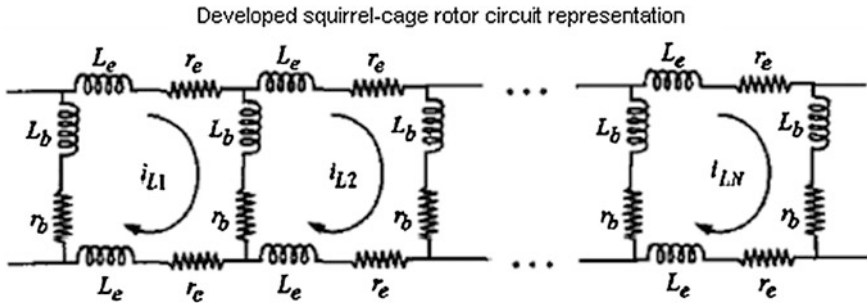
Fig. 2 Classification of faults

A reliable parameter estimation technique for induction motors is critical for the development of high-performance drive systems, and it can also be utilized for condition monitoring applications as well. However, in the context of existing literature the main thrust of parameter estimation techniques in motor-drive control applications is its use to achieve fast and controlled torque response of an induction motor utilizing the principle of vector (field oriented) control. The widely used squirrel-cage rotor aids in the robustness and economy of the drive, but rotor quantities are not accessible. However, the full advantage of vector control is available only if the instantaneous position of the rotor flux vector relative to a stationary reference frame can be indirectly obtained. Hence, knowledge of motor parameters is needed.

An accurate parameter estimation technique can also be used for motor condition monitoring purposes. Here in this paper, a simple and reliable technique, based on parameter estimation methods, is introduced for rotor broken bar fault detection. The main idea is that the apparent rotor resistance and leakage inductance of a squirrel-cage induction motor will increase when a rotor bar breaks. Meanwhile, the stator resistance, inductance and the magnetizing inductance will not be directly impacted by such rotor bar breakages. It is a well established fact that a broken bar fault generates modulation envelopes superposed on the amplitudes of stator currents over a slip cycle. This type of envelopes cannot be observed under no-load condition because the squirrel-cage bars virtually conduct no current under such a condition. Accordingly, the rotor circuit's effects are hardly reflected on the stator side. This means that the stator parameters are not affected by a bar breakage. This is due to the fact that in a no-load situation only the electrical circuits of the stator windings carry the currents and the rotor circuit behaves like an open circuit secondary winding.

This paper addresses two issues, first it presents a new approach for rotor parameter estimation of induction motors, and second it presents the utilization of this parameter estimation approach for purposes of rotor condition monitoring of induction motors.

2.2 Basic Principle



- Rotor Resistances and Inductance of a IM will increase when a rotor bar breakage occurs.
- If Rotor bar breaks the bar resistance r_b , will assume an infinite value.
- Consequently increase in the overall equivalent resistance of the rotor cage is affected.

3 Methodology

The present rotor parameter estimation approach constitutes a combination of signal processing and least squares techniques. In this approach, the motor terminal currents, voltages and motor speed are sampled over a specific period of time. The measured voltages and currents are transformed from an ABC frame of reference to a stationary dq0 reference frame. Then, the obtained voltages and currents are further broken down to their frequency components. Furthermore, those waveforms are mathematically expressed in a form of summation of sinusoidal waveforms with their associated frequencies. Consequently, the rotor currents, which are not physically accessible, can be calculated. Based on the measured stator quantities and obtained rotor currents, all represented in the stationary dq0 reference frame, a least squares method was implemented to calculate the rotor's inductance and resistance estimations.

The matrix equation of IM in dq0 reference frame is given by

$$\begin{bmatrix} v_{qs} \\ v_{ds} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} r_s + L_s D & \omega L_s & L_m D & \omega L_m \\ -\omega L_s & r_s + L_s D & -\omega L_m & L_m D \\ L_m D & (\omega - \omega_r) L_m & r_r + L_r D & (\omega - \omega_r) L_r \\ -(\omega - \omega_r) L_m & L_m D & -(\omega - \omega_r) L_r & r_r + L_r D \end{bmatrix} \begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{qr} \\ i_{dr} \end{bmatrix} \tag{1}$$

where,

- R_s is Stator resistance
- R_r is Rotor resistance
- L_s is Stator inductance
- L_r is Rotor inductance
- L_m is Magnetizing inductance
- ω_r is Rotor speed
- ω is Speed of reference frame
- D is time differential operator.

This matrix equation can be represented in a dq0 reference frame fixed to the stator, by substituting the reference frame speed, $\omega = 0$ in Eq. 3.43, which leads to the following

$$\begin{bmatrix} v_{qs} \\ v_{ds} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} r_s + L_s D & 0 & L_m D & 0 \\ 0 & r_s + L_s D & 0 & L_m D \\ L_m D & -\omega_r L_m & r_r + L_r D & -\omega_r L_r \\ \omega_r L_m & L_m D & \omega_r L_r & r_r + L_r D \end{bmatrix} \begin{bmatrix} i_{qs} \\ i_{ds} \\ i_{qr} \\ i_{dr} \end{bmatrix} \quad (2)$$

The first two rows of (2), which express the stator differential equations, can be rewritten as follows

$$Di_{qr} = 1/L_m(v_{qs} - r_s i_{qs} - L_s Di_{qs}) \quad (3)$$

$$Di_{dr} = 1/L_m(v_{ds} - r_s i_{ds} - L_s Di_{ds}) \quad (4)$$

Here, v_{qs} , v_{dr} , i_{qs} , i_{ds} , i_{qr} , and i_{dr} are represented in the stationary frame of reference with the rotor currents referred to the stator side. Using the Fast Fourier Transformation (FFT), one can obtain all frequency components of any stator or rotor voltages and currents. From the obtained frequency components, the waveforms can be reconstructed in a time domain as the summation of the sinusoidal waveforms.

From Eq. 4, the rotor differential equations can be rewritten as follows

$$i_{qr} r_r + (Di_{qr} - \omega_r i_{dr}) L_r = \omega_r L_m i_{ds} - L_m Di_{qs} \quad (5)$$

$$i_{dr} r_r + (Di_{dr} + \omega_r i_{qr}) L_r = -\omega_r L_m i_{qs} - L_m Di_{ds} \quad (6)$$

where, the unknown parameters are the rotor resistance, r_r , and the rotor inductance, L_r . In order to estimate these parameters, either (5) or (6) can be used for implementing a Least Squares (LS) method. Here, equation (5) is considered for the remainder of the discussion on estimating the unknown parameters.

4 Recursive Least Square Technique

Let vectors θ and X , and scalar Y be defined as follows

$$\theta = [\theta_1 \theta_2]^T = [r_r L_r]^T \quad (7)$$

$$X = [x_1 x_2]^T = [i_{qr} D i_{qr} - \omega_r i_{dr}]^T \quad (8)$$

$$Y = L_m (\omega_r i_{ds} - D i_{qs}) \quad (9)$$

Hence, (8) can be expressed based on the vectors of the unknown parameter, θ , X , and Y as follows

$$X^T \theta = Y \quad (10)$$

If we make N measurements, $N > 2$, we could minimize the squared error between the actual output, $y(t_k)$, and the predicted output, $\hat{y}(t_k) = \hat{\theta}^T X(t_k)$, by minimizing

$$J(\hat{\theta}) = \frac{1}{N} \sum_{k=1}^N \alpha [y(t_k) - \hat{\theta}^T x(t_k)]^2 \quad (11)$$

and typically $(1/\alpha)$ is set to a value just less than one. After further mathematical manipulation the following formulations are obtained (see **Appendix A** for more details)

$$\hat{\theta}(t_k) = \hat{\theta}(t_{k-1}) + P(t_k) x(t_k) \alpha [y(t_k) - (\hat{\theta}(t_{k-1}))^T x(t_k)] \quad (12)$$

$$P(t_k) = P(t_{k-1}) - \frac{P(t_{k-1}) x(t_k) (x(t_k))^T P(t_{k-1})}{(x(t_k))^T P(t_{k-1}) x(t_k) + (1/\alpha)} \quad (13)$$

where,

It has to be pointed out that the denominator in Eq. (13) is a scalar, although x is a vector, and P is a matrix. Typically, $P(t_0) = C I_{2 \times 2}$, where, C is a large constant number [5, 6].

5 Simulation Results and Analysis

5.1 Modelling

Parameter estimation approach based on recursive least square algorithm discussed in the previous chapter is validated for the following parameters using **MATLAB/SIMULINK** Software as shown in Fig. 3 [7].

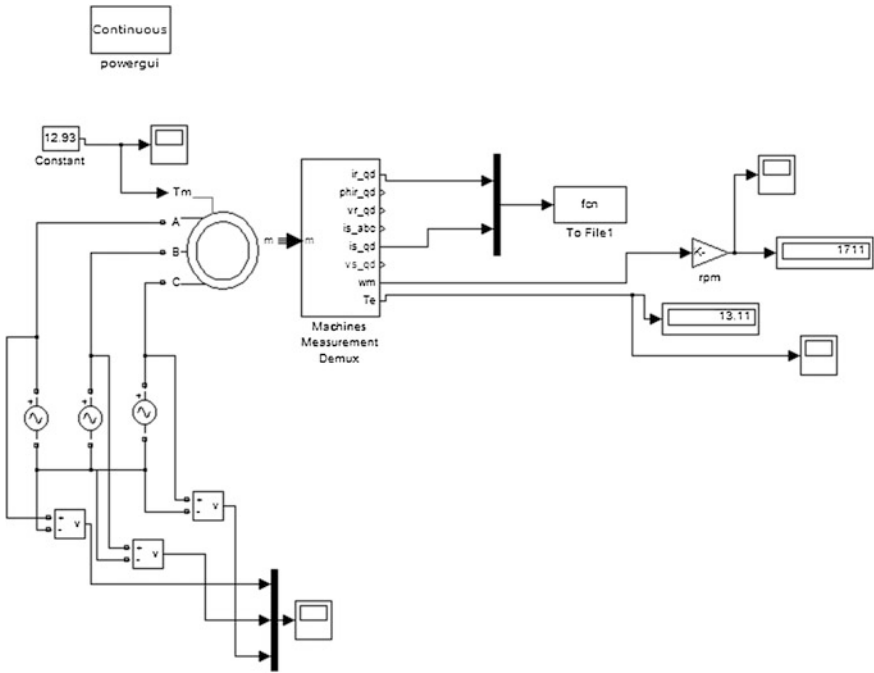


Fig. 3 Simulation circuit

5.2 Specifications of Motor Used

HP rating	3 hp
Voltage	220 (L–L)
Frequency	60 Hz
Stator Resistance	0.435 ohms
Stator Inductance	4 mH
Rotor Resistance	0.816 ohms
Rotor Inductance	2 mH
Mutual Inductance	69.31 mH
Moment of Inertia	0.089 Kgm ²
Rated Speed	1725 rpm
No. of Poles	4
Rated Torque	12.93 Nm
No. of Rotor bars	36.

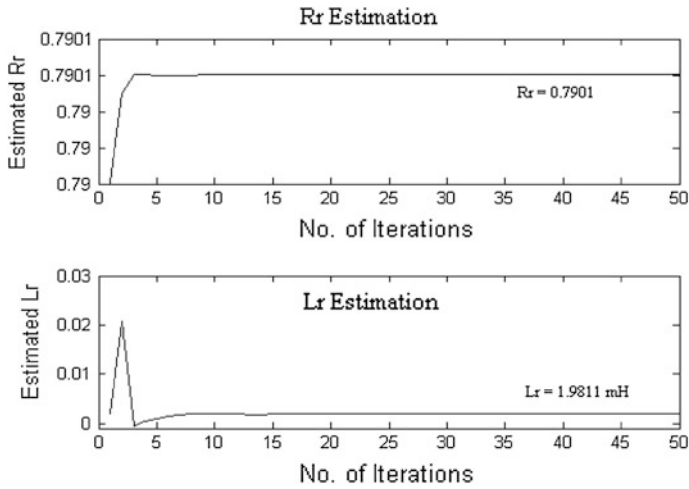


Fig. 4 Estimated values of rotor resistance and inductance (healthy case)

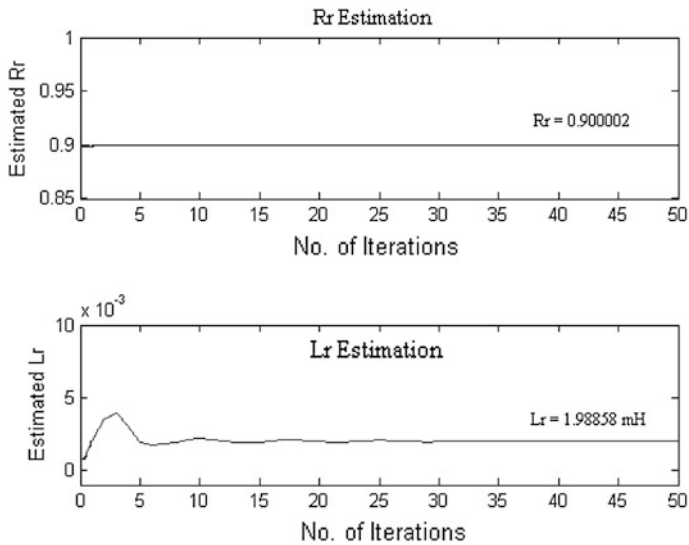


Fig. 5 Estimated values of rotor resistance and inductance (one bar broken case)

The induction motor model of MATLAB is simulated for various cases namely healthy, one bar broken, three bars broken and five bars broken conditions. Corresponding values of speed, current and voltages in dq0 frame is sampled and stored for further processing in least square algorithm. Then the M-file for the algorithm is run to estimate the values of rotor resistance and inductances.

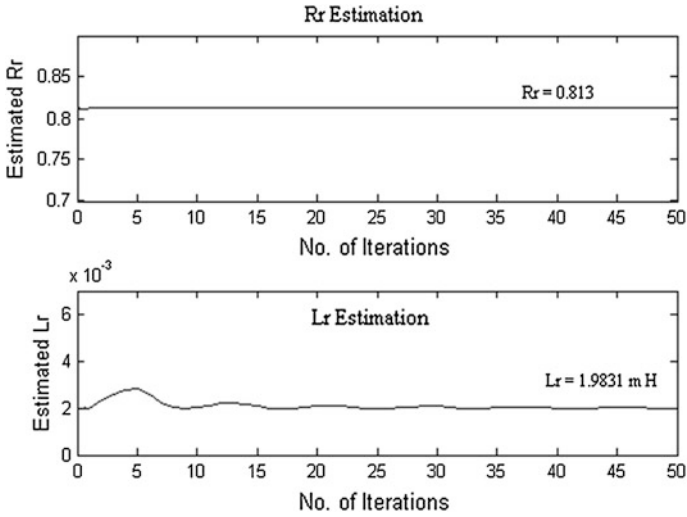


Fig. 6 Estimated values of rotor resistance and inductance (three bar broken case)

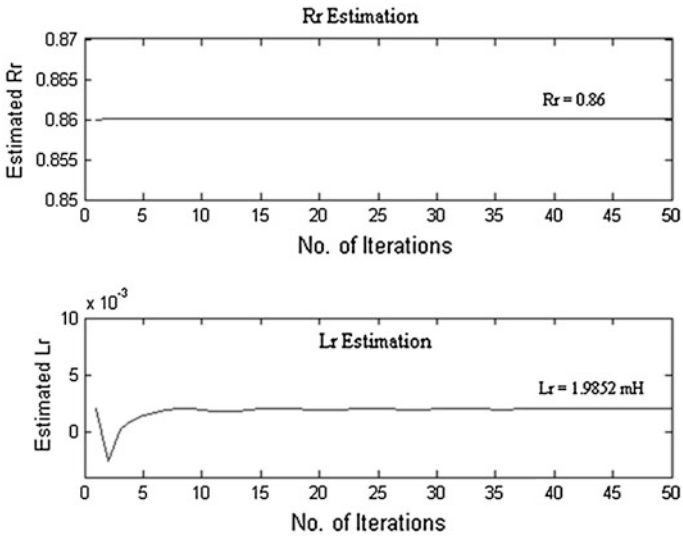


Fig. 7 Estimated values of rotor resistance and inductance (five bar broken case)

5.3 Simulation Results

The simulation results for the various cases are listed in Figs. 4, 5, 6, 7.

Table 1 Actual and estimated rotor resistance

Case	Actual value	Estimated value	% Error
Healthy condition	0.816	0.7901	3.17
One bar broken	0.8387	0.813	3.06
Three bar broken	0.884	0.86	2.71
Five bar broken	0.929	0.9	3.15

Table 2 Actual and estimated rotor inductance

Case	Actual value (mH)	Estimated value (mH)	% Error
Healthy condition	2	1.9811	0.945
One bar broken	2.0012	1.9831	0.904
Three bar broken	2.0036	1.9852	0.918
Five bar broken	2.00722	1.9885	0.929

5.4 Analysis

Simulation results are shown in the previous section. Table 1 gives the estimated values of rotor resistance for all the four cases. We can see that that resistance value increases for every case which is according the basic principle of the methodology. Also it can be noted that error also remains near 3 %. Table 2 gives the estimated values of rotor inductance for all the four cases. Similarly hear also there is slight increase in the value of the inductance for each case. But it can be noted that the change is not as much as the change in the resistance value.

6 Conclusion

It has been shown that a simple and reliable rotor parameter estimation approach presented here used for the rotor condition monitoring purposes. The proposed methodology has been validated through simulation for one, three and five bars broken conditions. The estimated rotor resistance and inductance indicates that these parameters increase in value with an increase in the number of broken bars. Moreover, from the simulation results, it has been found that estimation of the rotor inductance is more reliable than the rotor resistance, because the error is less when compared to the resistance.

References

1. Mirahl, B., Fateh, F., Yeh, C.C., Povinelli, R., Demerdash, N.A.O.: Condition monitoring of squirrel-cage induction motors fed PWM -based drives using a parameter estimation approach. In: International Conference on Power System Technology (POWERCON 2004), Singapore, 21–24 Nov (2004)
2. Siddique, A., Yadav, G.S., Singh, B.: A review of stator fault monitoring techniques of induction motors. *IEEE Trans. Energy Convers.* **20**(1), 106–114 (2005)
3. Watson, J.F.: The use of line current as a condition monitoring tool for three phase induction motors
4. Rodriguez-Cortes, H., Hadjicostis, C.N., Stankovic, A.M.: Model-based broken rotor bar detection on an IFOC driven squirrel cage induction motor. In: Proceeding of the American Control Conference, Boston, June 2004
5. Fault Zone Analysis—Rotor. Phillip cole, Motor Reliability Technical Conference (2004)
6. Trutt, F.C., Sottile, J., Kohler, J.L.: On-line condition monitoring of induction motors, IEEE
7. Yeh, C.C., Mirafzal, B., Povinelli, R.J., Demerdash, N.A.O.: A condition monitoring vector database approach for broken bar fault diagnostics of induction machines, IEEE
8. Ilonen, J., Kamarainen, J.K., Lindh, T., Ahola, J., Kalviainen, H.: Diagnosis tool for motor condition monitoring. *IEEE Trans. Ind. Appl.* **41**(4), 963–971 (2005)
9. Nandi, S., Toliyat, H.A.: Condition monitoring and fault diagnosis of electrical machines—a review

Word Recognition Using Barthannwin Wave Filter and Neural Network

Abhilasha Singh Rathor and Pawan Kumar Mishra

Abstract Speech Recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Recognition technique makes it possible to the speaker's voice to be used in verifying their identity and control access to services such as voice dialing, telephone shopping, banking by telephone, database access services, information service, voice mail, remote access to computers and security control for the confidential information areas. Digital filter plays an important role in digital signal processing applications. Digital filter can also be applied in speech processing applications, such as speech enhancement, speech filtering, noise reduction and automatic speech recognition among others. Filtering is a widely researched topic in the present era of communications. As the received signal is continuously corrupted by noise where both the received signal and noise change continuously, and then arises the need for filtering. This paper provides introduction to Barthannwin Wave Filter based on Barthannwin window for speech signal modeling suited to high performance and robust isolated word recognition. It provides efficient performances with less computational complexity. A Barthannwin Wave filter is designed based on the estimated noise statistics, and it is useful for noise reduction of the speech. The proposed filtering scheme outperforms other existing speech enhancement methods in terms of accuracy in a word recognition system.

Keywords Speech recognition · Speech processing · Word recognition · Filter design

A. S. Rathor (✉) · P. K. Mishra
Faculty of Technology, UTU, Dehradun, India
e-mail: abhilasha.rathor@gmail.com

P. K. Mishra
e-mail: pawantechno@gmail.com

1 Introduction

Speech is considered to be a primary mode of communication among human beings. It is also a most efficient and natural form of exchanging information among human [1]. Speech comes so naturally to everybody that we don't realize that speech is very difficult phenomenon to understand. We as human being understand speech very easily but computers have trouble to understand it. Computers are very good at recognizing and understanding instructions given to them but they have trouble in understanding speech. There are several problems in recognizing speech such as: (1) None of the words sound the same. (2) All speakers say the word in different way. (3) Context of language is meaningless.

Speech recognition systems can be categorized in different classes on the basis of what type of word they can recognize [1]. Isolated word can be defined as single word at a time. These recognizers generally involve every utterance to include quiet on both sides of sample windows [1]. Connected word can be said as group of words spoken together with minimum pause in between [1]. Continuous speech is similar to speaking a line or a paragraph [1]. Spontaneous speech is when human being speaks in natural way with or without pause between the words [1]. Digital representation is required to process a Signal, thus speech processing is regarded as a special area of digital signal processing (DSP), which can be applied to speech signal. Speech processing contains the acquirement, manipulating, storing, transferring and obtaining output of speech signals. Since the 1960s computer scientists have been researching ways and means to make computers able to record, interpret and understand human speech [1]. Speech Recognition, SR is conversion of spoken word into text. Some speech recognition systems use training" and some do not. In trained SR systems every individual speaker reads a part text into the speech recognition system. After that these systems evaluate the specific voice of person after that they use this voice to fine tune the person's speech recognition, which results in more perfect transcript. Speaker Dependent systems are those systems which are required to be trained while Speaker Independent systems are those systems which are not required to be trained. Speech Recognition can be said as special case of pattern recognition. Training and Testing are two phase in supervised pattern recognition. In the training phase, the parameters of the arrangement model are anticipated using a large number of class examples (Training Data) while the testing or recognition phase, the attributes of test pattern is matched with the trained model of each and every class. The test pattern is said to belong to that whose model matches the test pattern best [1]. Speech Recognition systems performance is usually calculated in terms of speed and accuracy. Speed is measured along with real time factor while Accuracy is considered as Word error rate (WER). Additional measurements of accuracy are Command Success Rate (CSR) and Single Word Error Rate (SWER) [2]. Idea of speech to text can be hard to implement for intellectually disabled persons because of the fact that there are rare chances that anyone will try to learn these technologies to teach the persons with the disabilities [3].

1.1 Problem Definition

When a word is spoken, it is assumed that speech segments can be unfailingly separated from the non-speech segments. The process of separating the speech signals of an utterance from the background noise, i.e., the non-speech segments obtain while recording the signal, is known as endpoint detection. Isolated word recognition systems are used to accurately detect the endpoints of spoken words. It is important for two reasons: (1) Reliable word recognition is importantly dependent on the accurate endpoint detection. (2) When the endpoints are accurately located the computation for processing the speech becomes minimum. There are several problems in accurately locating the endpoints of isolated words while recordings. The problems in endpoint detection occur from transients connected with the speaker or the transmission system. Background noises also complicate the endpoint detection problem significantly. Thus accurate endpoint detection method is very essential component of word recognition. The necessary components of a speech recognition system are feature extraction, pattern comparison, and decision rule. While Endpoint detection is performed in the processing. Noise-robust speech endpoint detection is significant to get practical speech recognition in a noisy real-world environment [4]. The endpoints can be founded in explicitly, implicitly, or hybrid manner.

1.2 Literature Survey

In past years a number of dissimilar methodologies have been proposed for continuous speech and isolated word recognition. These are usually grouped in two classes called as speaker-dependent and speaker-independent. Speaker dependent methods involve training a system to recognize each vocabulary words uttered single or multiple times by specific set of speakers while for speaker independent systems such training methods are not applicable and words are recognized by analyzing their intrinsic acoustical properties. Recent research is focusing on three main features [5]: Large vocabulary size, Continuous Speech Capabilities, Speaker Independent System. Many systems use Hidden Markov models (HMMs) widely. While other uses Neural Networks, some also use Dynamic Time Wrapping and many more techniques [6, 7].

Al-Alaoui et al. [8] analyzed and discussed the applicability of artificial neural networks to speech recognition. A total number of 200 vowel signals from individuals with different gender and races were recorded. The filtering process was performed using the wavelet approach to de-noise and to compress the speech signals. Kotnik et al. [9] proposed a multiconditional robust mel frequency Cepstral coefficients feature extraction algorithm as an alternative to commonly used symmetrical Hamming window. He also proposed Cosine window (hHCw) in the preprocessing stage. Betkowska et al. [10] discussed the problem of speech

recognition in the presence of nonstationary sudden noise, like in home environments. The proposed FHMMs achieved better recognition accuracy than clean-speech HMMs for different SNRs. The overall relative error reduction given by phoneme FHMMs was 12.8 % compared to that given by clean-speech HMMs. Lim et al. [11] implemented a new pattern classification method, by Neural Networks trained using the Al-Alaoui Algorithm. The new method gave comparable results to already implement HMM method for recognition of words, and it has overcome HMM in the recognition of sentences. He compared two different methods for automatic Arabic speech recognition for isolated words and sentences. The KNN classifier gave better results than the NN in the prediction of sentences. Muda et al. [12] discussed two voice recognition algorithms Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW), which are important in improving voice recognition performance. Pandey et al. [13] discussed how binary recurring neural network can be used to solve the problem of recognizing sentences having similar meaning but different lexico-grammatical structures in English. Abushariah et al. [14] designed and implemented English digits speech recognition system using Matlab (GUI). It was based on Hidden Markov Model (HMM), which provided extremely trustworthy technique for speech recognition. Rafiee and Khazaei [15] proposed a new model for a noise robust Automatic Speech Recognition (ASR) based on parallel branch Hidden Markov Model (HMM) structure with novel approach for robust speech recognition. The characteristics of a novel model are presented by exploring vibrocervigraphic and ectromyographic ASR methods and some other successful approaches to achieve the best results. Paul et al. [16] proposed a methodology for automated recognition of isolated words independent of speakers. She computed ZCR by partitioning audio signal into segments and calculates the number of times the signal crosses zero amplitude level within each segment. Wijoyo [17] implemented speech recognition system on mobile robot for controlling movement of the robot. They used Linear Predictive Coding (LPC) and Artificial Neural Network (ANN) for speech recognition system. LPC method is used for extracting feature of a voice signal and ANN is used as the recognition method. Back propagation method is used to train ANN. Experimental results show that highest recognition rate that can be achieved by this system is 91.4 %. Ittichaichareon et al. [18] discussed an approach of speech recognition by using the Mel-Scale Frequency Cepstral Coefficients (MFCC) extracted from speech signal of spoken words. Based on experimental database of total 40 times of speaking words collected under acoustically controlled room, MFCC extracts have shown the improvement in recognition rates significantly when training the SVM with more MFCC samples by randomly selected from database, compared with ML.

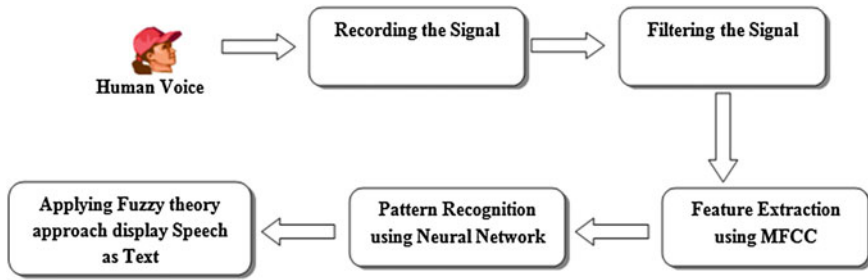


Fig. 1 Word recognition system design

2 Word Recognition System Design

The basic idea is to design a filter so that we can remove the noises and unwanted signal from the recorded voice and obtain high efficiency in Word Recognition. The scheme used for Word Recognition along with the algorithm of proposed Barthannwin Wave Filter is discussed. Word Recognition is an area of Speech Recognition, where we recognize the words spoken by several speakers with the words stored in our database. Word recognition requires the extraction of features from the recorded utterances followed by a training phase [19, 20]. There are several steps in word recognition system, our main focus in on filtering the signal and designing an efficient filter for word recognition system. A word recognition system design is shown in Fig. 1.

For recording the signal we have used MATLAB frontend. After recording the signal for filtering process Barthannwin Wave Filter is implemented. Filtering is a process that removes some unwanted distortion or noise from a signal. It aims at removing some frequencies in order to suppress interfering signals and reduce background noise. Filters are widely used in signal processing and communication systems in applications such as channel equalization, noise reduction, radar, audio processing, video processing, biomedical signal processing, and analysis of economic and financial data.

The primary functions of filters are: (1) To confine a signal into a prescribed frequency band as in low-pass, high-pass, and band-pass filters. (2) To decompose signal into two or more sub-bands as in filter-banks, graphic equalizers, sub-band coders, frequency multiplexers. (3) To modify frequency spectrum of a signal as in telephone channel equalization and audio graphic equalizers. (4) To model input-output relationship of a system such as telecommunication channels, human vocal tract, and music synthesizers.

Filter designing means to select the filter coefficients in such a way so that the system has specific characteristics. These characteristics are stated as filter specifications. Mostly the time filter specifications are known as frequency response of the filter. We also need to choose the transfer function and filter structure. Mapping the transfer function to the filter structure gives the element values of analog filters

elements and the coefficients of digital filter. The methods to find the coefficients of a filter from its frequency specifications are: (1) Window design method (2) Frequency Sampling method (3) Weighted least squares design (4) Parks-McClellan method (5) Equiripple FIR filters design using the FFT algorithms. We have used Window Design Method since in filter design, windows are typically used to reduce unwanted ripples of the filter in the frequency response. FFT windows reduce the effects of leakage but cannot eliminate leakage entirely. In effect, they only change the shape of the leakage. In addition, each type of window affects the spectrum in a slightly different way. Many different windows have been proposed over time, each with its own advantage and disadvantage relative to the others. Some are more effective for specific types of signal types such as random or sinusoidal. Some improve the frequency resolution, that is, they make it easier to detect the exact frequency of a peak in the spectrum. Some improve the amplitude accuracy, that is, they most accurately indicate the level of the peak. The best type of window should be chosen for each specific application. We have chosen Barthannwin window which is a modified form of Bartlett-Hann window.

2.1 Proposed Filter-Barthannwin Wave Filter

Barthannwin Wave Filter is a high pass FIR filter which is designed by using Barthannwin window. We have designed a FIR filter because of its following basic characteristics:

1. Linear phase characteristic
2. High filter order (more complex circuits); and
3. Stability

Barthannwin window is a combination of two windows Bartlett Window and Hann (also known as Hann) Window.

Bartlett Window: The Bartlett window is same as a triangular window. Bartlett window ends with 0's at samples 1 and n, while at those points the triangular window is nonzero. The center L-2 points of Bartlett (L) are equivalent to triang (L-2) for L odd. If we specify L = 1 for a one-point window, the value 1 is returned.

w = bartlett(L) returns an L-point Bartlett window in the column vector w, where L should only be a positive integer. The Bartlett window coefficients can be computed as follows:

$$w(n) = \begin{cases} \frac{2n}{N}, & 0 \leq n \leq \frac{N}{2} \\ 2 - \frac{2n}{N}, & \frac{N}{2} \leq n \leq N \end{cases}$$

The length of window $L = N+1$.

Hann Window: The Hann window has the shape of one cycle of a cosine wave with 1 added to it so it is always positive. After that the sampled signal values are

multiplied by the Hann function. The ends of the time record are forced to zero regardless of what the input signal is doing. The Hann window should always be used with continuous signals, but must never be used with transients. The reason is that the window shape will distort the shape of the transient, and the frequency and phase content of a transient is intimately connected with its shape.

w = hann(L) returns an L -point symmetric Hann window in the column vector w . L should be a positive integer. Hann window coefficients are computed from the equation.

$$w(n) = 0.5 \left(1 - \cos \left(2\pi \frac{n}{N} \right) \right), 0 \leq n \leq N$$

The window length is $L = N + 1$.

Barthannwin Window: This window consist main lobe at its origin and decaying side lobes on both sides asymptotically. It is a result of linear combination of Hann and weighted Bartlett windows with near side lobes lower than both Bartlett and Hann and with far side lobes lower than both Bartlett and Hamming windows. The main lobe width of the modified Bartlett-Hann window is not increased relative to either Bartlett or Hann window mainlobes.

w = barthannwin(L) returns an L -point modified Bartlett-Hann window in the column vector w . The following equation is used to compute the coefficients of a Modified Bartlett-Hann window is

$$w(n) = 0.62 - 0.48 \left| \left(\frac{n}{N} - 0.5 \right) \right| + 0.38 \left(2\pi \left(\frac{n}{N} - 0.5 \right) \right)$$

where $0 \leq n \leq N$ and the window length is $L = N + 1$.

After deciding which window to be used a high pass FIR filter was designed using Barthannwin Window known as Barthannwin Wave Filter.

For the feature extraction of the speech signal the Mel-frequency cepstrum coefficient is used in this model. In speech processing, mel-frequency cepstrum (MFC) represent the short-term power spectrum of the sound signal, which is based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. The feature measurements of speech signals are typically extracted using one of the following spectral analysis techniques: MFCC Mel frequency filter bank analyzer, LPC analysis or discrete Fourier transform analysis. Currently the most popular features are Mel frequency Cepstral coefficients MFCC. There are two ways to analyze the MFCCs: (a) as a filter-bank processing adapted to speech specificities and (b) as a modification of the conventional cepstrum, a well-known deconvolution technique based on homomorphic processing [33]. The block diagram for calculating MFCCs is shown in Fig. 2.

Neural networks are composed of simple elements operating in parallel, inspired by biological nervous systems. The network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Neural networks are adjusted, or trained, so that a particular

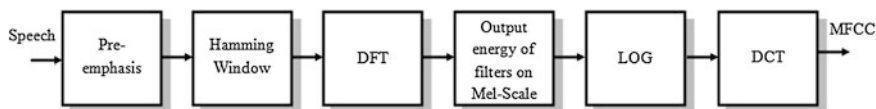


Fig. 2 MFCC calculation

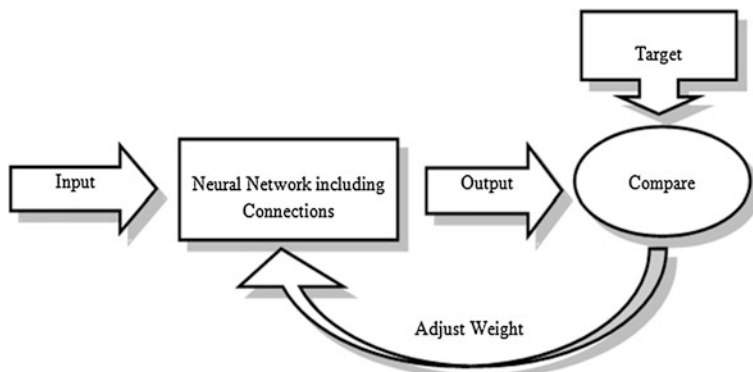


Fig. 3 Neural network training

input leads to a specific target output. There, the network is adjusted, based on a comparison of the output and the target, until the network output matches the target.

Neural networks have been trained to perform complex functions in various fields of application including pattern recognition, identification, classification, speech, vision and control systems. The supervised training methods are commonly used, but other networks can be obtained from unsupervised training techniques or from direct design methods. Radial basis functions are dominant techniques used for interpolation in the multidimensional space. A RBF is a function which has constructed into a distance criterion with respect to a center. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the ‘hidden’ layer. The RBF chosen is usually a Gaussian.

Radial basis function (RBF) networks typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer [39]. Architecture of a radial basis function network is shown in Fig 3. An input vector x is used as input to all radial basis functions, each with different parameters. The network output is a linear combination of outputs from radial basis functions.

A common method of training a neural net in which the initial system output is compared to the desired output and the system is adjusted until the difference between the two is minimized. It is a controlled learning method, and simplification of the delta rule. For making the training set it requires a dataset of desired

output for different inputs. Back propagation needs that the activation function which is used by artificial neurons be differentiable. Back propagation networks are essentially multilayer Perceptrons (typically with one input, hidden, and output layer). The back propagation learning algorithm can be divided into two stages: propagation and weight update.

2.2 Steps for Word Recognition

Steps for Word Recognition process are explained below.

Step 1: Begin

Step 2: Get The Input Data **/ Record the Samples*

**/ Check if Samples are correct or not, if samples are not correct, record again*

Step 3: Check The Input Data

**/ Apply designed filter so that important information is retained and rest is discarded*

Step 4: Apply Barthannwin Wave Filter

**/ Extract the features of the signal using MFCC*

Step 5: Extract Features

**/ Create the network where P denotes the Input and T denotes the Target*

Step 6: Create The Network

P ← Input

T ← Target

**/ Train the Neural Network*

Step 7: Training The Neural Network

**/ Apply the Fuzzy Set-Theoric Approach to match Input to Target*

Step 8: Apply Fuzzy Set-Theoric Approach

**/ If match found display the word*

Step 9: If Output Is Equal To Target

Display Word

Play Voice

End if

**/ If word not found Display that word is not found*

Step 10: If Word Not Found Display “Word Not Found”

Step 11: End

2.3 Algorithm for Barthannwin Window

1. Begin

2. if number of input arguments < 1

Use barthannwin function

- End if
- 3. if length of Signal < 0
 - Display "Error"
- End if
- 4. Compute $L \leftarrow \text{round}(L)$ **/L denotes the length of signal*
- 5. Compute $N \leftarrow L - 1$ **/N denotes maximum Order*
- 6. For $n \leftarrow 0$ to N
 - Compute $w = 0.62 - 0.48|(n/N - 0.5)| + 0.38 \cos[2\Pi(n/N - 0.5)]$
- End
- 7. Compute $w \leftarrow w'$
- 8. End

2.4 Algorithm for Barthannwin Wave Filter

- 1. Begin
 - */All frequency values are in Hz.*
- 2. Set $F_s \leftarrow 32000$ **/Fs denotes sampling frequency*
- 3. Set $N \leftarrow 32000$ **/N denotes Order*
- 4. Set $F_c \leftarrow 10800$ **/Fc denotes Cutoff frequency*
 - */Sampling Flag, 'scale' to normalize the filter so that the magnitude response of the filter at the center frequency of the passband is 0 dB. */*
- 5. Set Flag \leftarrow 'Scale'
 - */Create the window vector for the design algorithm.*
- 6. Set win \leftarrow barthannwin ($N + 1$)
 - */calculate the coefficients using the FIR1 function.*
- 7. Compute $b = \text{fir1}(N, F_c/(F_s/2), \text{'high'}, \text{win}, \text{flag})$;
 - */'high' is for a highpass filter with cutoff frequency $F_c/(F_s/2)$.*
 - */returns a discrete-time, direct-form finite impulse response (FIR) filter, Hd, with numerator coefficients, b. */*
- 8. Compute $H_d = \text{dfilt.dffir}(b)$;
- 9. End

3 Experimental Results

The experiments are executed in order to assess performance of the newly developed Barthannwin Wave filter applied on fuzzy word recognition. In interpreting the results, consideration must be taken since the behaviour of the algorithm is highly dependent on the parameter values. The results are only meaningful for the specific test signal set. For other test signals, the performance may be different.

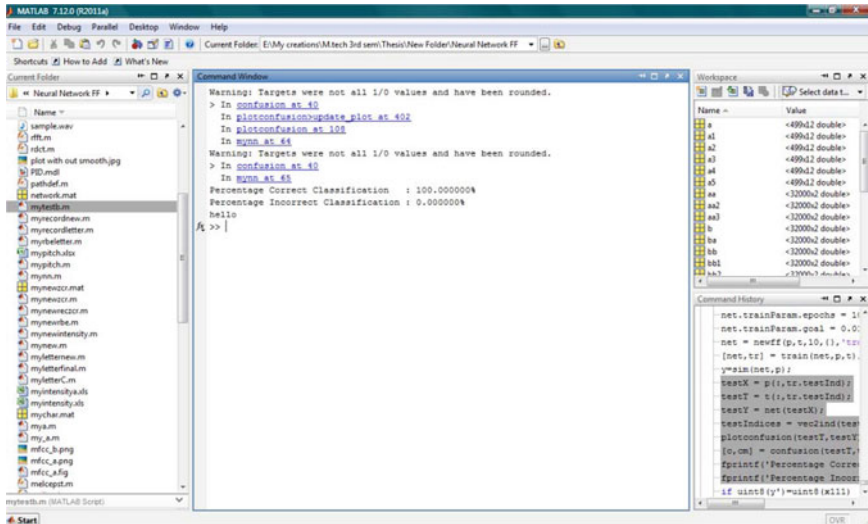


Fig. 4 Output window for word “Hello” with 100 % correct classification

The new filter was implemented in a word recognition system and its accuracy is shown with the help of different words spoken by different age group.

The results obtained are shown in Fig. 4.

The 5 words were recorded that are “Hello”, “How”, “Are”, “You” and “Fine” using microphones. 2 s time was allotted to record a word. System was trained for these 5 words with 1 male and 2 female voice. After that different samples were collected from people belonging to different age groups and accuracy of system was calculated.

All the results of word “Hello” in both Male and Female voices is as shown. Which shows more than 97 % of accuracy is achieved in all samples of Male voice and 90 % of accuracy is achieved in the samples of Female voice (Tables 1, 2).

4 Conclusion

The basic of Barthannwin Window, system identification and Barthannwin Wave Filter using Barthannwin Window was designed. A new word recognition algorithm for speech corrupted by slowly varying additive background noise based on Barthannwin Wave filtering was developed. It was demonstrated that this new algorithm can enhance the accuracy of word recognition system. The algorithm was implemented and evaluated. The parallel version proves that speech enhancement algorithms can profit from parallel computing techniques. Nevertheless, the rather modest quality improvement of the resulting speech over alternative methods would not by itself justify the significant increase in

Table 1 Accuracy of word “hello” in male voice

Number of samples Age group	10–15 years	15–25 years
10	99 %	98.6 %
50	98 %	98.2 %
80	97.6 %	97.2 %

Table 2 Accuracy of word “hello” in female voice

Number of samples Age group	10–15 years	15–25 years
10	95 %	95.6 %
50	93 %	93.4 %
80	92.6 %	92.2 %

computational complexity. This work also demonstrates the use of a model based approach for word recognition. In the case studied, to identify the model parameters correctly and reliably was found to be the critical part of the algorithm. Finally, it is important to keep in mind that for speech enhancement systems designed for human listener, it is the human listener who is the ultimate judge. The conversion to an optimization problem is a design and engineering decision and must be always verified by listening tests. As far as enhancement systems for speech recognition are concerned, their design criterion should closely match the recognizer structure and therefore, it seems unlikely that a single enhancement algorithm would perform well for both tasks. Constant progress is being done in the domain of speech enhancement but it is still a challenging and rewarding field where full of problems are waiting to be solved.

References

1. Gaikwad, S.K., Gawali, B.W., Yannawar, P.: A review on speech recognition technique. *Int. J. Comput. Appl.* (0975–8887) **10**(3), 16–24 (2010)
2. Singh, G.N.: Literature review on automatic speech recognition. *Int. J. Comput. Appl.* **41**(8), 42–50 (2012)
3. Hawley, M.S., Green, P., Enderby, P., Cunningham, S., Moore, R.K.: Speech technology for e-inclusion of people with physical disabilities and disordered speech. In: *INTERSPEECH 2005*, pp 445–448, Sept 2005
4. Kitayama, K., Goto, M., Itou, K., Kobayashi, T.: Speech starter: noise-robust endpoint detection by using filled pauses. In: *Proceedings of the 8th European Conference on Speech Communication and Technology (Eurospeech 2003)*, pp. 1237–1240, Sept 2003
5. Vimala, C., Radha, V.: A review on speech recognition challenges and approaches. *World Comput. Sci. Inf. Technol. J. (WCSIT)* **2**(1), 1–7 (2012)
6. Das, B.P., Parekh, R.: Recognition of isolated words using features based on LPC, MFCC, ZCR and STE, with neural network classifiers. **2**(3), 854–858 (2012), ISSN: 2249–6645
7. Chunsheng, F.: From Dynamic Time Warping (DTW) to Hidden Markov Model (HMM). Final project report for ECE742 stochastic decision, March 2009

8. Al-Alaoui, M.A., Al-Kanj, L., Azar, J., Yaacoub, E.: Speech recognition using artificial neural networks and hidden markov models. *IEEE Multi. Eng. Educ. Mag.* **3**(3), 419–423 (2008)
9. Kotnik, B., Vlaj, D., Kacic, Z., Horvat, B.: Robust MFCC feature extraction algorithm using efficient additive and convolutional noise reduction procedures. In: *ICSLP'02 Proceedings*, pp. 445–448 (2002)
10. Betkowska, A., Shinoda, K., Furui, S.: Speech recognition using FHMMs robust against nonstationary noise. In: *ICASSP 2007*, pp. 1029–1032
11. Lim, C.P., Woo, S.C., Loh, A.S., Osman, R.: Speech recognition using artificial neural networks. In: *Web information systems engineering*, IEEE, Inc. 2000, pp. 77–86
12. Muda, L., Begam, M., Elamvazuthi, I.: Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. *J. Comput.* **2**(3), 138–143 (2010)
13. Pandey, B., Ranjan, S., Shukla, A., Tiwari, R.: Sentence recognition using hopfield neural network. *IJCSI Int. J. Comput. Sci. Issues* **7**(4), 12–17 (2010)
14. Abushariah, A.A.M., Gunawan, T.S., Khalifa, O.O., Abushariah, M.A.M.: English digits speech recognition system based on hidden markov models. In: *International Conference on Computer and Communication Engineering (ICCCE 2010)*, 11–13 May 2010, Kuala Lumpur, Malaysia
15. Rafiee, M.S., Khazaei, A.A.: A novel model characteristics for noise-robust automatic speech recognition based on HMM. In: *Proceedings of IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS)*, 2010, pp. 215–218
16. Paul, D., Parekh, R.: Automated speech recognition of isolated words using neural networks. *Int. J. Eng. Sci. Technol. (IJEST)* **3**(6), 4993–5000 (2011)
17. Wijoyo, T.S.: Speech recognition using linear predictive coding and artificial neural network for controlling movement of mobile robot. *International Conference on Information and Electronics Engineering IPCSIT vol. 6*, pp. 179–183 (2011)
18. Ittichaichareon, C., Suksri, S., Yingthawornsuk, T.: Speech recognition using MFCC. In: *International Conference on Computer Graphics, Simulation and Modeling (ICGSM'2012)*. Pattaya (Thailand), 28–29 July 2012
19. Lipeika, A., Lipeikien, J., Telksnys, L.: Development of isolated word speech recognition system. *Informatika* **13**(1), 37–46 (2002)
20. Ashouri, M.R.: Isolated word recognition using highorder statistics and time -delay neural networks. In: *IEEE Signal Processing Workshop on Higher-Order Statistics*, IEEE, Inc., 1997

Graph Coloring Problem Solution Using Modified Flocking Algorithm

Subarna Sinha and Suman Deb

Abstract Graph coloring is a widely studied method of assigning labels or colors to elements of a graph. This can also be mapped with bio-inspired bird flocking algorithms to solve the NP complete graph coloring problem in optimum time complexity. This paper proposes an application of the Bird flocking algorithm that uses the concepts of a flock of agents, e.g. birds moving together in a complex manner with simple local rules namely cohesion, alignment, separation and avoidance. Each bird representing one data, move with the aim of creating homogeneous groups of data in a two dimensional environment producing a spatial distribution that can be used to solve a particular computational problem. The combination of these characteristics can be used to design and solve the task of 3 coloring graphs. This graph labeling can hierarchically or linearly be applied on a domain specific network or set of items.

Keywords Graph coloring · Swarm · Bird flocking · NP complete · Spatial distribution

1 Introduction

The research area dealing with autonomous agents have gained a huge popularity in the past few years due to its application in a wide range of fields. This paper provides information about how autonomous agents work and follow a certain set

S. Sinha (✉) · S. Deb

Department of Computer Science and Engineering, National Institute of Technology,
Jirania, Agartala 799055, Tripura, India
e-mail: subarna.sinha9@gmail.com

S. Deb

e-mail: sumandeb.cse@nita.ac.in

of rules to solve a particular problem. They work based on collection of individual units working in unison each of which follow a certain defined protocol. These include Complex Adaptive Systems, Swarm Intelligence [1], Multi-Agent Systems and Self Organizing Systems etc. The way of a single ant, bee, termite and wasp behaves is often too simple, but their collective and social behavior is of great significance. Analysis of any wild life documentary reveals that advanced mammals also enjoy social lives. The collective and social behavior of living creatures motivated researchers to undertake the study of today what is known as Swarm Intelligence (SI) [2, 3]. SI systems are typically made up of a population of simple agents (an entity capable of performing/executing certain operations) interacting locally. Local interactions between such agents often lead to the emergence of global behavior. The problem addressed in this paper is to identify the computational capabilities that allow the emergent behavior of Swarms to solve NP-complete problems. In this work, the simple distinction between two different kinds in a population of boids [4] is enough to produce an emergent behavior that can be interpreted as solving the problem of graph coloring [5]. The swarm intelligence concept is used to model the graph coloring problem as follows: agents correspond one-to-one with the nodes of the proposed graph. The graph topology defines the agent affinities as follows: the agents whose nodes are directly connected are “different kind”, agents whose nodes are not directly connected are “same kind”. Agents are attracted to “same kind” while try to fly from or to avoid “different kind”. The colors for the graph coloring correspond to specific colored flock of birds distributed initially in spatial region. When all the agents are placed in their respective flocks to which they belong, the system configuration can be interpreted as defining a complete coloration of the graph. The combination of Cohesion, alignment, separation and avoidance to obstacles, is being proposed as a solution to the graph coloring problem.

2 Related Work

Craig Reynolds in 1986 developed an idea which uses the flocking behavior of birds and simulates it. His paper was published in 1987 in the proceedings of the ACM SIGGRAPH conference [4].

Emergent behavior was artificially simulated through his proposed. Each of the birds in the flock adheres to a set of simple rules but complexity exists due to the interaction between boids/agents. The rules proposed in Boids paper are mentioned below:

- (A) *Separation (steer to avoid crowding local flock mates)*
- (B) *Alignment (steer towards the average heading of local flock mates)*
- (C) *Cohesion: steer to move toward the average position of local flock mates.*
- (D) *Avoidance: steer to avoid obstacles in d way.*

In graph theory, graph coloring [5] is a special case of graph labeling; it is an assignment of labels traditionally called “colors” to elements of a graph subject to certain constraints. In its simplest form, it is a way of coloring the vertices of a graph such that no two adjacent vertices share the same color; this is called a vertex coloring. Similarly, an edge coloring assigns a color to each edge so that no two adjacent edges share the same color, and a face coloring of a planar graph assigns a color to each face or region so that no two faces that share a boundary have the same color.

In 1912, Birkhoff [6] introduced the chromatic polynomial to study the coloring problems. Kempe [7] had already drawn attention to the general, non-planar case in 1879 and many results on generalizations of planar graph coloring to surfaces of higher order followed in the early 20th century. Graph coloring has been studied as an algorithmic problem since the early 1970s: the chromatic number problem is one of Karp’s 21 NP-complete problems from 1972 [8].

3 Behavior Rules of Flocking of Birds in Detail

Each rule returns a vector of how much the bird’s destination should be changed by that particular rule, so simply adding all of these to the birds’ current destination very effectively points the bird to its new corrected destination.

3.1 Algorithm

Update_Position_Bird(Data: List of birds)

```
{
  If (bird! = different community bird)
    Flock (bird)
  End if
  Checkbound()//[check the boundary]
  Checkspeed()//[check the speed]
  dx = dx + value//value = value returned from cohesion, separation, alignment
  or avoidance rule.
  dy = dy + value//value = value returned from cohesion, separation, alignment
  or avoidance rule.
  New_value_x = current_value_x + dx;
  New_value_y = current_value_y + dy;
  New_position = (New_value_x, New_value_y);
}
```


3.2 Algorithm

Cohesion rule(Data: List of birds)

```

{
  For each bird in the flock
  if (distance < sight)//sight = sphere around each bird that shows its perception
  update bird.position.x
  update bird.position.y
  End if
  dx = dx + (bird.position.x- position.x)*weight
  dy = dy + (bird.position.y- position.y)*weight
  End for
}

```

3.3 Algorithm

Separation rule(Data: List of birds)

```

{
  For each bird in the flock
  If (distance < space)//space = maximum desired space between two birds in a flock
  dx = dx + (position.x- bird.position.x)
  dy = dy + (position.y - bird.position.y)
  End if
  End for
}

```

3.4 Algorithm

Alignment rule(Data: List of birds)

```

{
  For each bird in the flock do
  If (distance < sight)//sight = sphere around each bird that shows its perception
  distance
  dx = dx + bird.dx*weight
  dy = dy + bird.dy*weight
  End if
  End for
}

```

Birds move towards the average destination of their neighbors keeping the flock in alignment and moving together towards the same general heading. The alignment rule calculates the average destination of all birds within a set neighborhood distance from the original bird (including the destination of the original bird) and returns the average of those destinations.

3.5 Algorithm

Obstacle Avoidance(Data: List of birds)

```
{
  For each bird in the flock
  if(bird of _different_community && distance < sight)//sight = sphere around each
  bird that shows its perception distance
  dx = dx + (position.x - bird.position.x)
  dy = dy + (position.y - bird.position.y)
  End if
  End for
}
```

Birds must avoid hitting other type of birds as they fly while still obeying all the other rules of cohesion, separation and alignment. When an obstacle is found to be within a minimum set distance to a bird, the bird moves away from the object in order to avoid it. This rule has a high weighting as it would be important for real birds not to fly into obstacles and this is reflected in the simulation.

3.6 Supporting Algorithm: Euclidean Distance

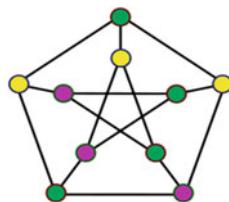
For all the above steering rules the distance between birds or birds and obstacles needs to be calculated. This is accomplished by calculating the Euclidean distance between them [9]. To calculate Euclidean distance between two points (x_1, y_1) and (x_2, y_2) in two dimensional space (used in the 2D simulation) the following formula is used:

$$\text{Distance} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

4 Coloring of Graphs

The coloring of a graph is almost always a proper vertex coloring, when used without any qualification. The vertex coloring is a labelling of the graph's vertices with colors such that no two vertices sharing the same edge have the same color.

Fig. 1 General form of graph coloring



Since a vertex with a loop could never be properly colored, it is understood that graphs in this context are loopless (Fig. 1).

The terminology of using colors for vertex labels goes back to map coloring. Generally labels like red and blue are only used. They are used when the number of colors is small, and normally it is understood that the labels are drawn from the integers $\{1, 2, 3, \dots\}$.

A coloring [10] using at most k colors is called a (proper) k -coloring. The smallest number of colors needed to color a graph G is called its chromatic number, and is often denoted $\chi(G)$. Sometimes $\gamma(G)$ is used, since $\chi(G)$ is also used to denote the Euler characteristic of a graph. A graph that can be assigned a (proper) k -coloring is k -colorable, and it is k -chromatic if its chromatic number is exactly k . A subset of vertices assigned to the same color is called a color class; every such class forms an independent set. Thus, a k -coloring is the same as a partition of the vertex set into k independent sets.

4.1 Applications

A. Scheduling—Vertex coloring models are used in a number of scheduling problems [11]. A given set of jobs need to be assigned to time slots, each job requires one such slot. Jobs can be scheduled in any order, but pairs of jobs may be in conflict in the sense that they may not be assigned to the same time slot, for example because they both rely on a shared resource. The corresponding graph contains a vertex for every job and an edge for every conflicting pair of jobs. Details of the scheduling problem define the structure of the graph.

B. Register allocation—To improve the execution time of the resulting code, one of the techniques of compiler optimization is register allocation, where the most frequently used values of the compiled program are kept in the fast processor registers. Ideally, values are assigned to registers so that they can all reside in the registers when they are used. The textbook approach to this problem is to model it as a graph coloring problem [12]. The compiler constructs an interference graph, where vertices are symbolic registers and an edge connects two nodes if they are needed at the same time.

C. Other Applications—The problem of coloring a graph has found a number of applications, including pattern matching. The recreational puzzle Sudoku can be seen as completing a 9-coloring on given specific graph with 81 vertices.

5 Proposed Method

In this paper it is shown that the simple distinction between “we” and enemies “them” in a population of boids is enough to produce an emergent behavior that can be interpreted as solving the problem of graph coloring. We use the swarm [13] metaphor to model the graph coloring problem as follows: agents correspond one-to-one with the nodes of the proposed graph. The graph topology defines the agent affinities as follows: the agents whose nodes are directly connected are “different kind”, agents whose nodes are not directly connected are “same kind”. Agents are attracted to “same kind” while try to fly from or to avoid “different kind”. The colors for the graph coloring correspond to specific colored flock of birds distributed initially in spatial region. All agents are attracted to stay in this region. Each of the agents tries to move towards and along the flock type it belongs to. When all the agents are placed in their respective flocks to which they belong, the system configuration can be interpreted as defining a complete coloration of the graph. When some agents are outside their respective flocks, the system configuration corresponds to a partial solution to the coloration problem. This emergent behavior of the birds which means combination of Cohesion, alignment, separation and avoidance to obstacles, is being proposed as a solution to the graph coloring problem.

Let $G = (V, E)$ be a graph, [14] composed of a set of vertices V and a set $E \subseteq V^2$. A k -coloring (where $k = \text{number of colors}$) of G is a function that maps each vertex in $V = \{b_1, \dots, b_n\}$ to a color in the set $C = \{1, 2, \dots, k\}$ in a way such that two connected vertex have different colors. The minimum number of colors that are used to color a graph is called its chromatic number. When $k \geq 3$, the problem of graph coloring becomes an NP complete one. Initial position and velocity of each of the bird or agent are decided at random. The network formed by the birds is represented with the help of graph concept. So G represents the bird network or the population of flying birds of different types. Let each agent be denoted by b_i . Then two nodes $i, j \in V$ are enemies if they are connected in the graph G , that is $(i, j) \in E$. Hence, from the point of view of agents b_i , the set of agents or birds actually inside its neighborhood is partitioned into two subsets as the set of enemies A_i and the set of friends B_i with an amity relationship (Fig. 2).

$$S_i = A_i \cup B_i \quad (2)$$

where $A_i = \{b_j : (b_i, b_j) \in E\}$.

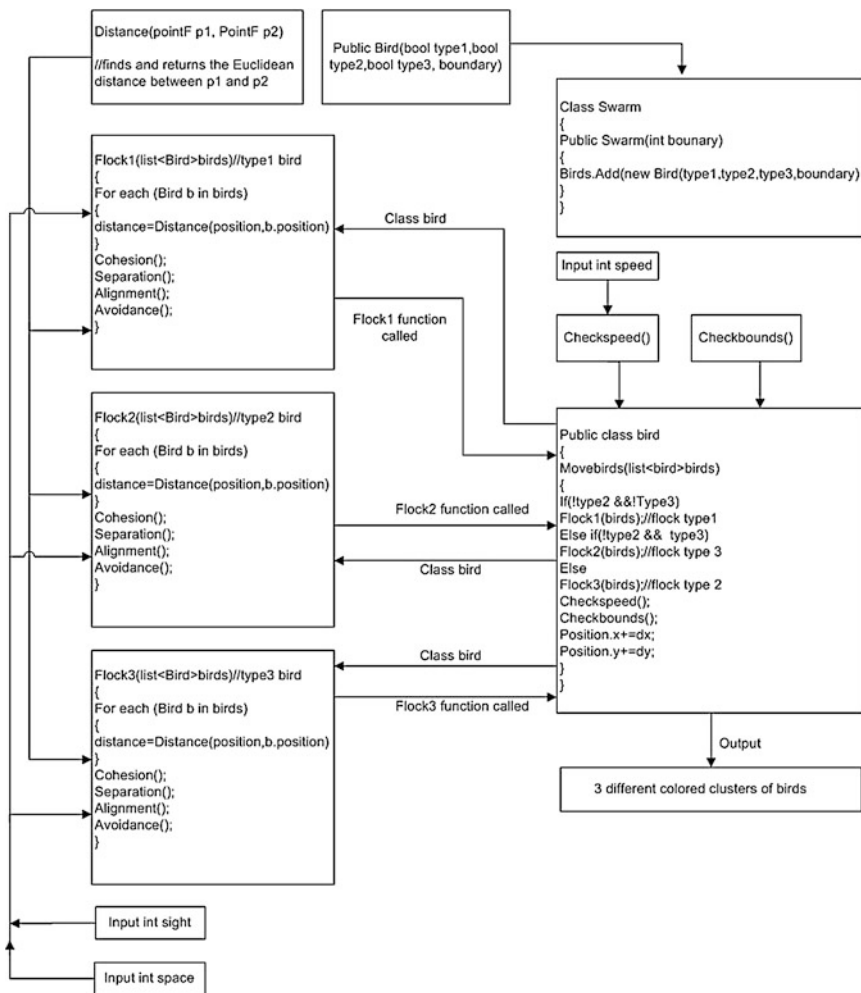


Fig. 2 Inter-functional relationship diagram

6 Results

6.1 Attraction or Repulsion Behavior

The application of the Boids swarms algorithm to colorable small graphs is verified by proving the algorithms namely Cohesion, Alignment, Separation, Avoidance against obstacles. All the above mentioned rules function together to produce the flocking behavior which is our main concern to solve the graph coloring problem. Figure 3a shows the behavior of the bird with all the four

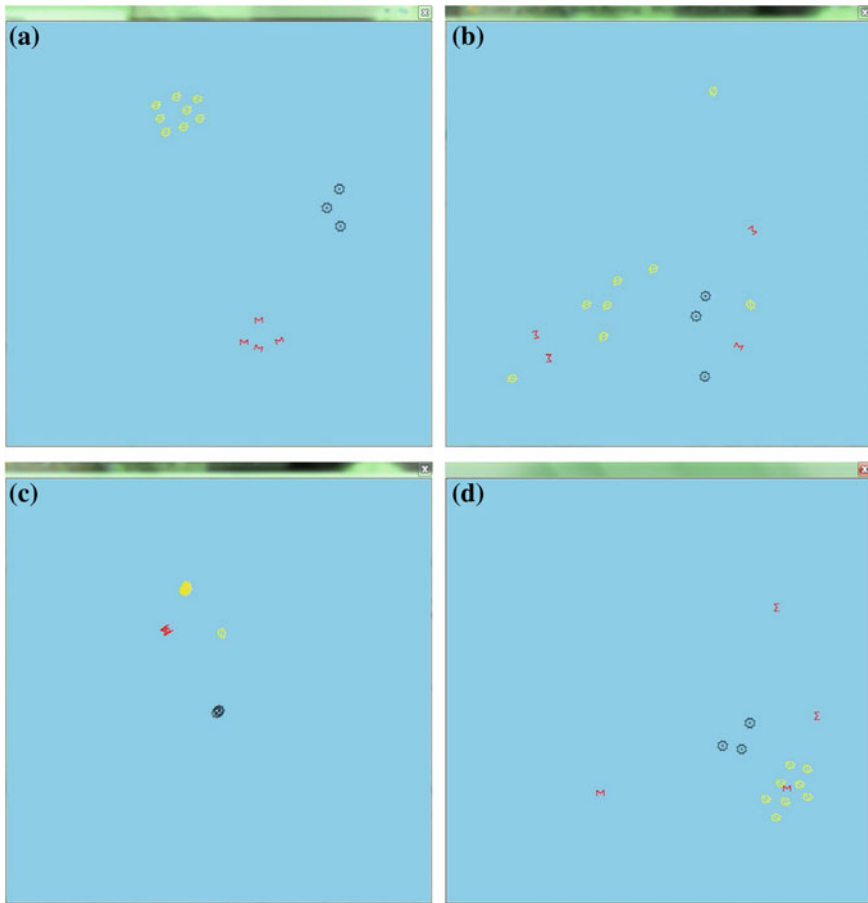


Fig. 3 a Bird behavior with all the four rules; b Bird behavior without cohesion rule; c Bird behavior without separation rule; d Bird behavior without avoidance rule

flocking rules, which is the ideal behavior for solving the 3 graph coloring problem. Figure 3b shows the behavior without the cohesion rule. We can observe that all the three kinds of birds got mixed forgetting about heading towards the average position of their respective flocks. Similarly if we want to differentiate the 3 differently colored vertices of a graph vertex from each other using the above mentioned Boids algorithm, then we will require the use of Cohesion algorithm along with the remaining two algorithms. Figure 3d shows the behavior without avoidance rule. Here we can notice that the birds forget their property to avoid birds belonging to a different flock. The importance of the remaining two algorithms of the flocking method is proved by the snapshots of the implementation of the algorithm in several scenarios (without separation, without alignment).

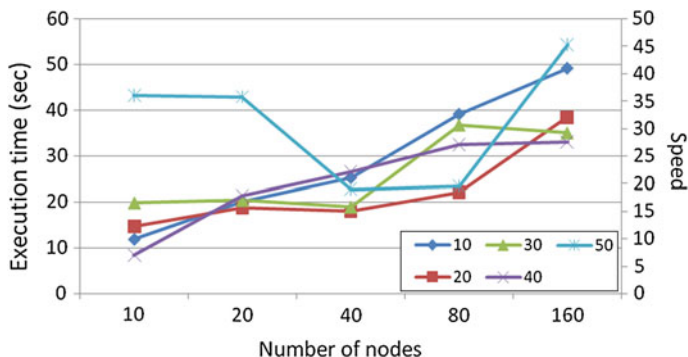


Fig. 4 Graph plotted between execution time, number of nodes and speed

6.2 Implementation

The performance of the algorithm is evaluated in the following section under different evaluation criteria. The algorithm is implemented in Microsoft Visual Studio and the .Net framework was chosen as a development environment with C# as the chosen language and executed on a core i3 processor, 2.27 GHZ, 3 GB RAM computer. The evaluation criteria considered for the proposed system is (1) no of nodes/agents (2) Execution time and (3) Speed.

The snapshots show the behavior of the swarm agents in various conditions (with or without separation etc.). All the behavioral rules produce the concept of graph coloring. Each of the yellow, black, red shapes as shown in Fig. 3b belongs to their respective community. Each of yellow colored shapes belongs to a single flock or in other words they are non adjacent vertices in the graph. But the black and red ones can be considered as adjacent vertices in the graph so yellow shapes simply avoid them using the avoidance algorithm. The same thing happens for black and red shapes (Fig. 4).

7 Conclusion

The coloration of graphs in this paper work is mainly approached geometrically, such that the nodes of a graph are represented as a flocking bird situated geographically. The solution to the graph coloring emerges from the whole population configuration. The aim of this paper is not to create a model of coloring graphs from an existing model. Rather our aim was the research of a swarm individual's special ability of perception of attraction or repulsion to another individual of the swarm. This ability of a swarm individual or agents can be used to solve the coloring problem in graphs which has been used extensively in scheduling, register allocation etc. This paper proposes solution to 3 coloring problem as shown in the

implementation where agents of only three colors can be seen. Each agent or bird belonging to a swarm perceives another agent or bird as belonging to its own type or to the other type. Each type has its own properties. This amity–enmity protocol allows easy modeling of the solving process for coloring graphs. Another contribution of this paper has to do with the complexity of swarms, understood as the complexity of the behavior of the emergent organism with respect to the computational capabilities of individuals.

References

1. Huffman, D.A.: A method for the construction of minimum-redundancy codes. In: Proceedings of the I.R.E., pp. 1098–1110, Sept 1952
2. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm intelligence: from natural to artificial systems. Oxford University Press, New York (1999)
3. Rodríguez, A., Reggia, J.A.: Extending self organizing particle systems to problem solving. *Artif. Life* **10**(4), 379–395 (2004)
4. Reynolds, C.W.: Flocks, herds and schools: A distributed behavioral model. SIGGRAPH'87. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, pp. 25–34. New York, NY, USA, ACM Press (1987)
5. Dániel, M.: Graph colouring problems and their applications in scheduling. *Periodica Polytech., Electr. Eng.* **48**(1–2), 11–16 (2004). (CiteSeerX: 10.1.1.95.4268)
6. Birkhoff, G.D.: A determinant formula for the number of ways of coloring a map. *Ann. Math.* **14**(1/4), 42–46 (1912)
7. Kempe, A.B.: On the geographical problem of the four colors. *Am. J. Math.* **2**, 193–200 (1879)
8. Karp, R.M.: Reducibility among combinatorial problems. In: Miller, R.E., Thatcher, J.W. (eds.) *Complexity of Computer Computations*, pp. 85–103. Pelnum Press, New York (1972)
9. David, M.F., de Castro, L.N.: A New Clustering Boids Algorithm for Data Mining. Mackenzie University, Brazil (2008)
10. Galinier, P., Hertz, A.: A survey of local search methods for graph coloring. *Comput. Oper. Res.* **33**, 2547–2562 (2006)
11. Leighton, F.T.: A graph colouring algorithm for large scheduling problems. *J. Res. Natl Bur. Stan.* **84**(6), 489–503 (1979)
12. Chaitin, G.J.: Register allocation and spilling via graph coloring. In: Proceedings of SIGPLAN'82 Symposium on Compiler Construction, SIGPLAN Notices, vol. 17(6), June 1982
13. Lodding, K.N.: The Hitchhiker's guide to biomorphic software. *ACM Queue* **2**(4), 66–75 (2004). (Topic: Bioscience)
14. Borodin, O.V., Glebov, A.N., Raspaud, A., Salavatipour, M.R.: Planar graphs without cycles of length from 4 to 7 are 3-colorable. *J. Comb. Theory Ser. B* **93**, 303–311 (2005)

Augmented Human Interaction with Remote Devices Using Low Cost DTMF Technology

R. Singathiya, Neeraj Jangid, Prateek Gupta and Suman Deb

Abstract In present world, we must use assorted high tech devices and equipments to get our jobs done and make the life simpler. These devices can be controlled by home care taker from any place since the home care taker might not present at home. Thus we need a remote interaction system with our every day essential devices for a technology blended state of the art life of present time. Smart home is equipped with such system so that we can control our home appliances from any location. In order to examine a true remote and enough secure solution to be really favorable and practicable, mobile technology is better than any other solution. In this paper we introduce new criteria so that the unremarkable services of the mobile phones can enlarged to communicate with and operate the home appliances and make our home a really well groomed one with obvious cost effective technology.

Keywords DTMF · Appliance control system · Smart home · Augmented control

R. Singathiya (✉) · P. Gupta

Department of Electronics and Communication Engineering,
National Institute of Technology Agartala, Jirania, Agartala 799055, Tripura, India
e-mail: ram.singathiya@gmail.com

P. Gupta

e-mail: 2291prateekgupta@gmail.com

N. Jangid

Department of Electronics and Instrument Engineering, National Institute of Technology
Agartala, Jirania, Agartala 799055, Tripura, India
e-mail: neerajjangid3@gmail.com

S. Deb

Department of Computer Science and Engineering, National Institute of Technology
Agartala, Jirania, Agartala 799055, Tripura, India
e-mail: sumandebcs@gmail.com

1 Introduction

As we know today's world is connected with the modern electronic equipments, so it is necessary that every human being should get such facilities which can be easily sustained. Although the DTMF Technology was discovered earlier but we have tried to make its use in day to day life as it is cost effective and it does not require any external remote system. There are many corporations which are trying to research in this sector; it enables them to control system easily in home or outdoor. The system is provided with mobile phone normally registered in communication service and another phone which receives the call automatically. In this research we have investigated the different ways we could use the cell phones to go beyond making calls and sending SMS and device some ways to implement the remote control system. This paper proposes methods of control that use simple voice call. The method proposed uses the Dual-Tone-Multi-Frequency (DTMF) generated when a keypad button of the mobile phone is pressed by the user. In this manner user controls the system.

2 Traditional DTMF Technology

It is widely used in remote control system and is mostly available in every cellular phone. It comprises of 16 keys as mentioned in the Table 1.

2.1 DTMF Tone Generation

When we press the key of mobile phone, an authentication code (basically DTMF tone) generates, consisting of two sinusoidal waves which can be represented as follows.

$$f(t) = a \sin(2\pi f_a t) + b \sin(2\pi f_b t) + \dots \dots \dots$$

where f_a and f_b are frequencies of two sinusoidal waves with a and b as their peak amplitudes and f as the resultant DTMF signal. Here,

$$(0.7) < (a/b) < (0.9) \text{ V}$$

The frequencies are adjusted in such manner that they are not the harmonics of each other (Fig. 1).

The frequencies associated with assorted keys on the keypad are shown in Table 2.

Table 1 The frequencies corresponding to a particular key on the mobile phone used

	1209 Hz	1336 Hz	1477 Hz
697 Hz	1	2	3
770 Hz	4	5	6
852 Hz	7	8	9
941 Hz	*	0	#

It can be noticed that for every key there are two frequencies, one specified for the row while the other for the column. There are also 4 special keys i.e. A, B, C and D which are used for special purposes

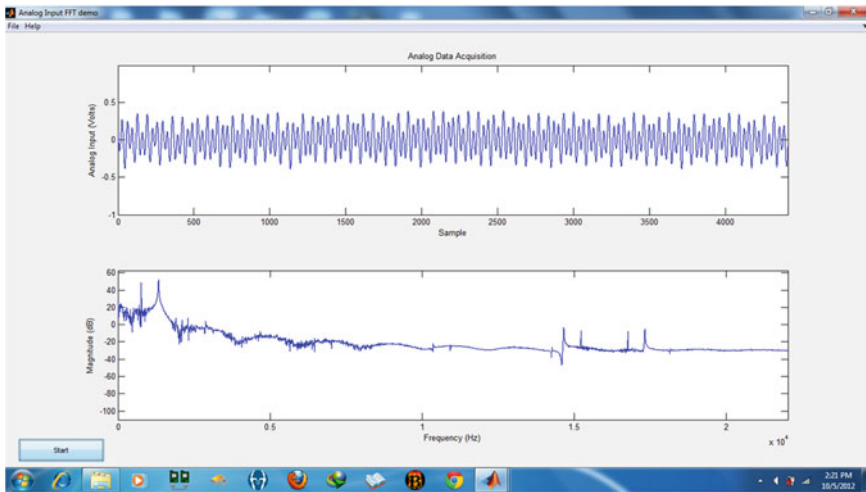


Fig. 1 Frequency spectrum analysis of DTMF tone of key '5' on mobile phone

2.2 DTMF Signal Through Cellular Network

The DTMF signals work in similar manner as the voice would have as you usually talk mobile passing through many base towers and even satellites in case of long

Table 2 Individual key press at transmitter end reflects as a BCD value A3, A2, A1 and A0 at the outputs of M-8870 DTMF decoder

Button	Low frequency (Hz)	High frequency (Hz)	A3	A2	A1	A0
1	697	1,209	0	0	0	1
2	697	1,336	0	0	1	0
3	697	1,477	0	0	1	1
4	770	1,209	0	1	0	0
5	770	1,336	0	1	0	1
6	770	1,477	0	1	1	0
7	852	1,209	0	1	1	1
8	852	1,336	1	0	0	0
9	852	1,477	1	0	0	1
0	941	1,209	1	0	1	0
*	941	1,336	1	0	1	1
#	941	1,477	1	1	0	0

distances. When a call is connected the number buttons send an authentic code which is carried as a digitized and compressed audio stream.

3 Description of Equipment

3.1 Input Device

The telephony device sends the DTMF signal to the receiver telephony device connected with the DTMF Decoder circuit through the earpiece.

3.2 Decoder Circuit

The Decoder M-8870 consists of a filter which is basically a band split filter and its work is to separate the high and low frequency of specific received tone. Also there is a digital decoder which verifies both the frequencies that we receive as the filter output before passing it to output bus. When we get the valid tone, then the early steering flag bit gets high, and for any signal loss it falls down (Fig. 2).

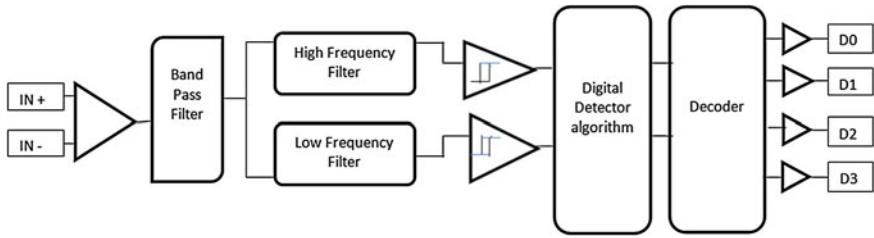


Fig. 2 This figure represents working of decoder circuit

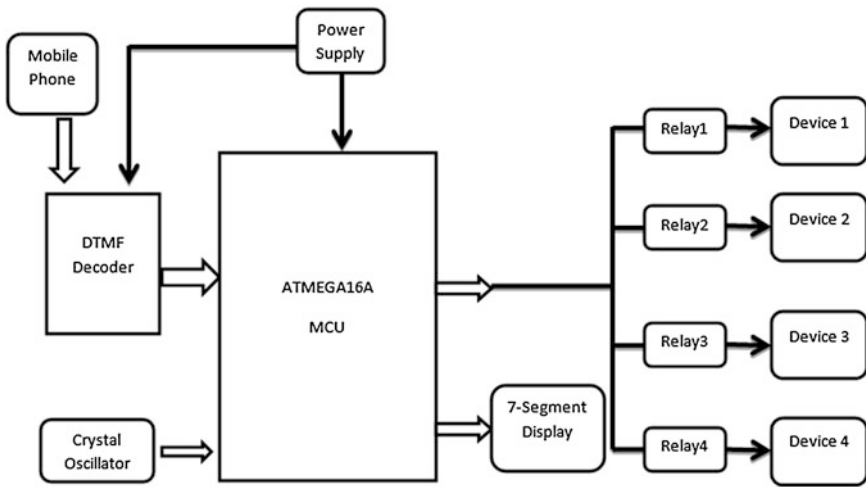


Fig. 3 Block diagram of circuit

3.3 Microcontroller ATmega16A

The role of microcontroller is to monitor the input signal to get the desired output. When the binary data comes from decoder IC it is transferred to microcontroller, and it gives the desired output as we program in the flash memory in the microcontroller.

3.4 Relay

It is used to operate the AC device from the microcontroller's output (Fig. 3).

4 Structure of Proposed Work

The signal is transmitted to the second mobile phone which is at the receiver end connected with the DTMF decoder circuit. The receiver end comprises of input device, decoder, microcontroller, seven-segment display and condenser microphone. For the decoder section, the underlying concept is DTMF signal and this signal is fed to display driver and microcontroller. The microcontroller is connected to relay circuit which performed the required action.

Once the connection is maintained between the two phones whatever phone key is pressed at the transmitting end, the corresponding DTMF tone is heard in the earpiece of the receiver phone. The earpiece is connected to a condenser microphone which picks up the DTMF tone. Its output is amplified by the DTMF decoder. The DTMF decoder will give the corresponding BCD value of the tone. Seven-segment display acts as a visual indicator when the valid signal is received by the system. This output, through a driver circuit is connected to PORTS of the microcontroller (ATmega16A). This microcontroller's output is fed to real circuit to trigger a voice feedback.

5 Working Structure

The circuit was built and tested for the given picture of tested circuit. The device controlled was a 230 V/20 W Power saving CFL (Fig. 4).

5.1 Schematic Diagram

See Fig. 5

5.2 Major Areas of Applications

Industrial Automation
Home Security system
Remote switches
Mobile Robot Control
Wireless Radio Control.



Fig. 4 Working model of remote interaction system

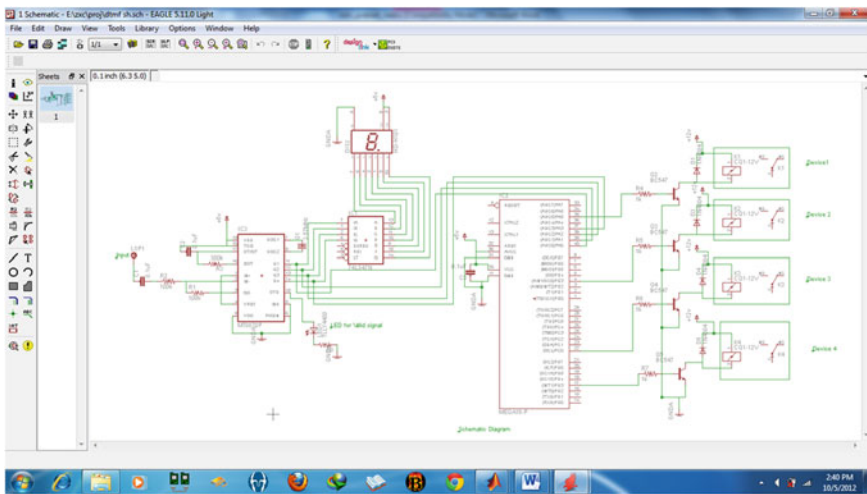


Fig. 5 This figure shows the schematic of complete structure designed in easily applicable graphical layout editor (*EAGLE*) software

6 Conclusions

It can be controlled by computer by using software which generates an artificial DTMF tone and can be further used in agriculture for sprinkling pesticides in the field using a control system which gives a message automatically to the farmer about the amount of chemical to be mixed to form pesticide according to different seasons. Moreover, DTMF technology is well applicable for irrigation of field by supplying water whenever it requires through mobile control system.

References

1. Smart Labs Technology: INSTEON—Wireless Home Control Solutions for Lighting, Security, HVAC, and A/V Systems. <http://www.insteon.net/>
2. Selman, S., Paramesran, R.: Comparative analysis of methods used in the design of DTMF tone detectors. In: Proceedings of the 2007 IEEE International Conference on Telecommunications and Malaysia International Conference on Communications, Penang, Malaysia, 14–17 May 2007
3. Koyuncu, B.: PC remote control of appliances by using telephone lines. *IEEE Trans. Consum. Electron.* **41**(1), 201–209 (1995). doi:[10.1109/30.370328](https://doi.org/10.1109/30.370328)
4. White, B., Jacobs, G., Landsburg, G.: A monolithic dual-tone multifrequency receiver. *IEEE international solid-state circuits conference. Digest of technical papers*, vol. 22, pp. 36–37 (1979). doi:[10.1109/ISSCC.1979.1155917](https://doi.org/10.1109/ISSCC.1979.1155917)
5. Wong, E.M.C.: A phone based remote controller for home and office automation. *IEEE Trans. Consum. Electron.* **40**(1), 28–33

Multi-Objective Ant Colony Optimization for Task Scheduling in Grid Computing

Nitu and Ritu Garg

Abstract Resource Management in Grid computing system is a fundamental issue in achieving high performance due to the distributed and heterogeneous nature of the resources. The efficiency and effectiveness of Grid resource management greatly depend on the scheduling algorithm. In this paper, the problem of scheduling is represented by a weighted directed acyclic graph (DAG). Ant Colony Optimization is used for scheduling tasks on resources in Grid which simultaneously pay attention to two objectives of makespan (schedule length) and the failure probability (reliability). These objectives are conflicting and it is not possible to minimize both objectives at the same time. With the help of concept of non-dominance, we are able to choose a trade-off between makespan minimization and reliability maximization. For evaluating the algorithm, ACO is compared with NSGA-II. The metrics for evaluating the convergence and diversity of the obtained non-dominated solutions by the two algorithms are reported. The results of simulation using JAVA programming language manifest that proposed approach can be used more efficiently for allocating the tasks as compared to NSGA-II.

Keywords Task scheduling · DAG · Makespan · Reliability

1 Introduction

Grid computing systems have emerged as a new environment for coordinated resource sharing and problem solving in multi-institutional virtual organizations while providing dependable, consistent, pervasive access to global resources. The

Nitu (✉) · R. Garg
Department of Computer Engineering, NIT, Kurukshetra, India
e-mail: me.nitumalhan@gmail.com

R. Garg
e-mail: ritu.59@gmail.com

sharing ranges from simple file transfer to direct access to computers, software, data, and other network accessible resources. Grid resources are heterogeneous, dynamic, complex and self-autonomic in nature which makes the resource management a significantly challenging job. Job Scheduling and Resource Management are the critical issues in Grid Computing [1]. The multiprocessor task schedule problem is known to be NP-complete. In order to address this problem, many heuristics algorithms have been proposed. These heuristics are classified into different categories such as list scheduling algorithms, clustering algorithms, duplication based algorithms [2].

Ant colony algorithm is one of the effective techniques used to solve NP-complete problems. ACO is inspired by the behavior of real ant colonies in nature to search for food and to connect to each other by pheromone trails laid on paths traveled. ACO was used to solve many NP-hard problems such as traveling salesman problem, vehicle routing problem, graph coloring problem [3–5] and so on. In this paper, we apply this technique for dependent task scheduling. The input to the scheduling algorithm is a directed acyclic graph (DAG), in which the node weights represent task processing times and the edge weights represent data dependencies as well as the communication times between tasks. The multi-objective ACO algorithm uses non-dominance approach for tackling the two objectives (makespan and reliability).

2 Related Work

Task scheduling is known to be NP-complete problem and lots of heuristics and meta-heuristics techniques have been examined to solve it. Most of them can be applied to the Grid environment with suitable modifications.

Min-min [6] set the tasks which can be completed earliest with the highest priority. The main idea of Min-min is that it assigns tasks to resources which can execute tasks the fastest. Max-min [6] set the tasks which has the maximum earliest completion time with the highest priority. The main idea of Max-min is that it overlaps the tasks with long running time with the tasks with short running time.

In [7] the qualities of the different obtained solutions were compared against the Particle Swarm Optimization (PSO), the Simulated Annealing (SA) and the Genetic Algorithms (GA) Meta-heuristic. The results show that PSO and GA are highly efficient and effective in the task scheduling problem. Later these Meta-heuristics were compared against a MOEA algorithm, optimizing the makespan and flowtime objective functions. In [8], the authors propose an algorithm called Multi-Objective Resource Scheduling Approach—MORSA, which is a combination between NPGA and NSGA Algorithms. They combine the sorting algorithm of non-dominated solutions with the process of Niche Sharing to ensure diversity. In [9], a deadline-based model is proposed which first generates all feasible scheduling solutions with makespan less than a predefined deadline, and then finds the best possible solution with the maximum reliability between feasible solutions.

Ant colony optimization (ACO) is a meta-heuristic alternative for solving the complicated optimization problems [10]. There are many different kinds of ACO algorithm, i.e., Ant Colony System (ACS), Max–Min Ant System (MMAS), Rank-based Ant System (RAS), Fast Ant System (FANT) and Elitist Ant System (EAS).

3 Scheduling Problem and Formulation

Generally grid applications in e-science and e-business falls in the category of workflow applications modeled by an ordered graph called Directed Acyclic Graph (DAG). The application can be represented by $G(V, E)$ where set V represents n number of the subtasks of the application and a set E of edges shows the dependencies among the subtasks. Each and every task is related by directed edge representing the communication directions between tasks and precedence constraints (i.e. data dependency). A directed edge $e_{ij} \in E$ indicates the data dependency constraint exists between the tasks v_i and v_j . In this model, a task can't start executing until all its parents have been completed. The value assigned to the edge represents the amount of data to be transferred between tasks if they are not executing on the same resource. If both parent and offspring tasks processed on a same resource, the value of communication time between them is considered to be zero.

In any DAG, there is always an input node, v_{entry} as a node with no parent and an output node, v_{exit} as a node with no offspring. When DAG has several entry or exit tasks, these tasks are connected to a pseudo entry-task or pseudo exit-task that has zero load weight and zero capacity edges.

A schedule is a function $S: V \rightarrow P$ assigns a task to the processor that executes it. Let $V(j, s) = \{i | s(i) = j\}$ be the set of tasks assigned to processor j .

The completion time of a processor j is calculated as

$$C_j(s) = \sum_{i \in V(j,s)} (st_{ij} + w_{ij})$$

where st_{ij} denotes the start time of the task i on processor j . The start time of the entry task is assumed to be zero. Other tasks' start time can be computed by considering the completion time of all immediate predecessors of the task. The communication time, st_{ij} is added to the start time, if the dependent tasks are allocated to different processors. The makespan of a schedule is the time where all tasks are completed

$$C_{\text{max}} = \max_j C_j(s)$$

Every processor has a constant failure rate, and let λ_j denote the failure rate of processor j . The probability that a processor j executes all its tasks successfully is given by

$$P_{\text{succ}}^j(S) = e^{-\lambda_j C_j(s)}$$

It is assumed that faults are independent, therefore, the probability that schedule S has finished correctly is

$$P_{succ} = e^{-rel(s)}$$

The reliability index (rel) is defined by

$$rel(s) = \sum_j \lambda_j c_j(s)$$

Minimizing the objective function rel is equivalent to maximizing the reliability of the schedule. For solving the task scheduling problem the objectives C_{max} and rel are to be minimized simultaneously.

4 Proposed Algorithm

In this section we are presenting the algorithm for the dependent task scheduling in Grid environment using Ant Colony Optimization technique, which aims at achieving high reliability and reducing the makespan. The algorithm consists of two mechanisms, a ranking mechanism [11], which is a modified version of the HEFT [2, 12] and a processor assignment mechanism.

Steps for Scheduling Algorithm:

1. Set the computation cost of tasks and communication cost among them.
2. Compute upward RRank value for all tasks by traversing graph, starting from the exit task using Ranking method ().
3. Sort the task in a scheduling list by non-increasing order of upward rank value.
4. While there are unscheduled tasks in the list do
 - Select the first task v_i from the list for scheduling
 - For each processor p_m in the processor set $p_m \in P$ do
 - Calculate the heuristic information (n_{ij})
 - Calculate current pheromone trail value $\Delta\tau_{ij}$
 - Update the pheromone trail matrix
 - Calculate the probability matrix
 - Select the task with highest probabilities of i and j as the next task v_j to be executed on the resource P_j .
 - Remove the task v_i from the unscheduled list
 - Modify the resource free time
5. For each local solution generated by all ants, find the makespan and reliability index.
6. Apply the concept of non-dominated sorting on local solutions to find the globally best solution.

4.1 Ranking Method ()

Our algorithm has used reliability rank (RRank) attribute to compute priorities of tasks. The RRank [11] is a rank of a task, from the exit task to itself, and equal to the sum of average communication costs of edges, average computation costs and reliability overhead of tasks over all processors. Communication costs between tasks scheduled on the same processor are assumed zero, and the execution constraints are preserved.

The RRank is recursively defined as:

$$RRank(v_i) = \overline{w(v_i)} + \max_{v_j \in succ(v_i)} \left\{ \overline{w(e_{i,j})} + RRank(v_j) \right\} + RC_{v_i}$$

where $succ(v_i)$, the set of immediate successors of task v_i , $\overline{w(v_i)}$ is the average computation cost of task v_i , and $\overline{w(e_{i,j})}$ is the average communication cost of edge $e_{i,j}$. The RC_{v_i} is the reliability overhead of task v_i and can be computed by

$$RC_{v_i} = \left(1 - \prod_{n=1}^m \exp\{-\lambda_{P_n} \times w(v_i)/w(p_n)\} \right) \times \overline{w(v_i)}$$

The rank is computed recursively by traversing the task graph upward, starting from the exit task. For the exit task v_{exit} , the rank value is equal to

$$RRank(v_{exit}) = \overline{w(v_{exit})} + RC_{v_{exit}}$$

4.2 Task Assignment Mechanism

In this mechanism, tasks are assigned to the processors in such a way that makespan is reduced and system reliability is improved. In order to achieve these goals, we find the best mapping of tasks and processors applying ACO. The ant is placed at first task in the generated order and that task is mapped on one of the available resources required by that task. When each of the tasks of the system is mapped to any specific resource then solution construction for that ant is completed and a complete and feasible solution is created. Each task is mapped to a specific resource and preemption is not allowed, that resource is selected based on probability rule that depends on both pheromone on edges between tasks and resources and heuristic information that depends on objective functions.

$$P_{ij} = \left[\tau_{ij}^\alpha \eta_{ij}^\beta \right] / \sum \tau_{il}^\alpha \eta_{il}^\beta$$

where τ is the value of pheromone on edge between tasks i and resource j and η is the value of heuristic information, α and β are two constants that represents the

relative importance of pheromone trail values and heuristic information values that depends on problem considered.

After each iteration of the algorithm i.e. after all the ants completed their solution construction, Global pheromone updating rule is performed to increase the value of pheromone on the edges of the solution that is found to be global best in case of single objective and Non-dominated solutions in case of multi-objective problem, so that the probability of edges of the best solution to be traversed by ants in next iteration of algorithm increases as probability depends on pheromone value also.

4.3 Non-Dominated Sorting

We have used concept of non-dominance. In [13], a sorting technique is proposed to sort a population into non-dominated fronts, the first non-dominated front is found and removed from the population, then the second is found from the remaining members and removed, then the third, and so on, until every member of the population has been assigned to its proper front.

For a population of individuals with $M = 2$ objectives, the individual with the lowest first-objective score must be part of the NDF, since it cannot be dominated by any other individual. If two or more individuals tie for the lowest first-objective score, then these solutions must also be compared in the second objective, and the individual(s) which score(s) best in the second objective are definitely in the population's first non-dominated front. The algorithm repeatedly applies this idea to efficiently find the population's NDF.

5 Experimental Results and Discussion

In order to assess the effectiveness of the proposed scheduling method, we have obtained the solutions for random task graph. Random task graph were generated using the method as proposed in [2]. There are few parameters involved in the random task graph generation were set with the following values:

- $SET_N = \{20, 40, 60\}$,
- $SET_{CCR} = \{0.1, 1\}$,
- $SET_\alpha = \{0.2, 0.4\}$,
- $SET_{out_degree} = \{1, 2, 3, 4, n\}$,
- $SET_\beta = \{0.1, 0.5, 1.0\}$

where N is number of nodes (tasks) in the graph, α is shape parameter of the graph. CCR is the ratio of the average communication cost to the average computation cost of the application DAG. We have also generated randomly a set of processors, where λ is chosen uniformly in the range $[10^{-3}, 10^{-4}]$.

Fig. 1 Obtained Pareto solutions for 20 numbers of tasks

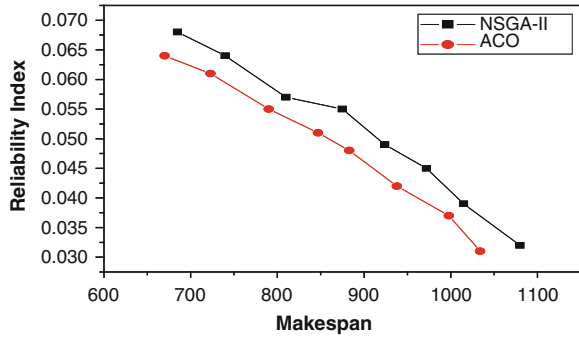


Fig. 2 Obtained Pareto solutions for 40 numbers of tasks

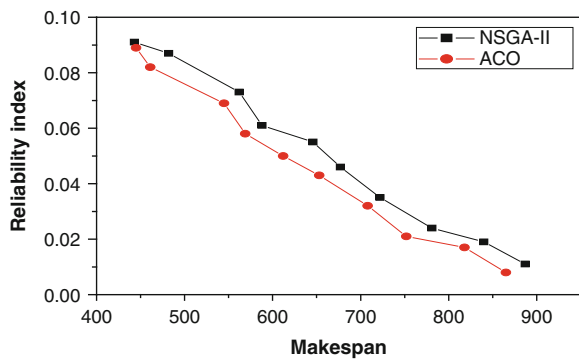
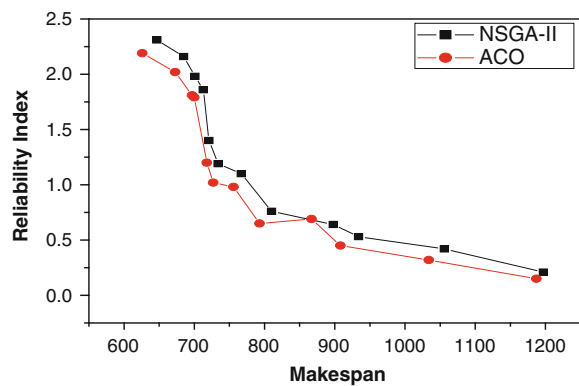


Fig. 3 Obtained Pareto solutions for 60 numbers of tasks



The obtained global Pareto solutions using multi-objective ACO and NSGA-II are as shown in Figs. 1, 2 and 3. The results show that the failure probability of the random task graph increases in proportion to the size of the application. This is due to the fact that when the size of an application increases, processors have to be failure-free for longer periods of time.

Table 1 Metrics for evaluating the diversity and convergence of ACO and NSGA-II

No. of tasks	Spacing		GD	
	ACO	NSGA-II	ACO	NSGA-II
20	1.15	1.19	0.25	0.28
40	0.86	0.94	0.43	0.49
60	0.46	0.49	0.54	0.57

There are many performance metrics proposed in the literature. One of the performance metrics namely, *spacing* [14, 15] is used to measure the diversity among obtained non-dominated solutions. The *Generational Distance Metric* (GD) [14, 15] used for measuring the convergence of the obtained non-dominated solutions. The spacing and GD values for the random task graphs are given in Table 1. The values confirm that ACO is better for the problem under study.

6 Conclusion

Scheduling is a critical issue for the execution of performance driven Grid applications. The work presented in the paper focuses on an efficient algorithm for multi-objective Grid Scheduling by assigning the submitted jobs to appropriate resources. In multi-objective optimization problem, multiple trade-off pareto solutions are produced for the maximum satisfaction of user. In this work, we proposed the use of Ant Colony Optimization algorithm using concept of non-dominance to solve bi-objective workflow scheduling problems. In our scheduling problem, we have considered the two major objectives: minimization of makespan (execution time) and maximization of reliability (to incorporate failure affect of resources in scheduling decision). But we have formulated the reliability in terms of minimization of reliability index.

The pareto solutions obtained by multi-objective ACO are compared with the solutions obtained by NSGA-II and a statistical analysis of their results has been presented to show the quality of each algorithm on different number of tasks. To measure the quality of these obtained solutions, we selected two metrics called GD (Convergence metric) and spacing (Diversity metric). A statistical analysis showed that multi-objective ACO outperforms NSGA-II in terms of both convergence towards pareto optimal front and maintaining good spread between solutions.

References

1. Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the grid. Int. J. Supercomput. Appl. **15**(3) (2001)
2. Topcuoglu, H., Hariri, S., Wu, M.Y.: Performance-effective and low complexity task scheduling for heterogeneous computing. IEEE Trans. Parallel Distrib. Syst. **13**(3), 260–274 (2002)

3. Dorigo, M., Gambardella, L.M.: Ant colony system: a cooperative learning approach to the traveling salesman problem. *IEEE Trans. Evol. Comput.* **1**(1), 53–66 (1997)
4. Salari, E., Eshghi, K.: An ACO algorithm for graph coloring problem. In: Conference on Computational Intelligence Methods and Applications, December 2005
5. Zhang, X., Tang, L.: CT-ACO-hybridizing ant colony optimization with cycle transfer search for the vehicle routing problem. In: Conference on Computational Intelligence Methods and Applications, December 2005
6. Maheswaran, M., Ali, S., Siegel, H.J., Hensgen, D., Freund, R.: Dynamic matching and scheduling of a class of independent tasks onto heterogeneous computing system. *J. Parallel Distrib. Comput.* **59**, 107–131 (1999)
7. Khafa, F., Abraham, A.: Computational models and heuristic methods for grid scheduling problems. *J. Future Gener. Comput. Syst.* **26**, 608–621 (2010)
8. Ye, G., Rao, R., Li, M.: A multiobjective resources scheduling approach based on genetic algorithms in grid environment. In: 5th International Conference on Grid and Cooperative Computing Workshops. pp. 504–509 (2006)
9. Dai, Y.S., Levitin, G.: Performance and reliability of tree structured grid services considering data dependence and failure correlation. *IEEE Trans. Comput.* **56**(7), 925–936 (2007)
10. Sallim, J., Shahrir, W.M., Hussin, W.: A background study on ant colony optimization metaheuristic and its application principles in resolving three combinatorial optimization problems. In: National Conference on Software Engineering and Computer Systems, Legend Resort Kuantan (2007)
11. Tang, X., Li, K., Li, R., Veeravalli, B.: Reliability-aware scheduling strategy for heterogeneous distributed computing systems. *J. Parallel Distrib. Comput.* **70**, 941–952 (2010)
12. Liu, G.Q., Poh, K.L., Xie, M.: Iterative list scheduling for heterogeneous computing. *J. Parallel Distrib. Comput.* **65**(5), 654–665 (2005)
13. Mazurek, M., Wesolkowski, S.: Non-dominated sorting on two objectives. Defence R&D Canada—CORA, Technical Note 027, pp. 1–13, July 2009
14. Deb, K.: *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley, New York (2001). ISBN: 0-471-87339-X
15. Deb, K., Jain, S.: Running performance metrics for evolutionary multi-objective optimization. In: *Simulated Evolution and Learning*, pp. 13–20 (2002)

An Evaluation of Reliability of a Two-Unit Degradable Computing System Using Parametric Non-linear Programming Approach with Fuzzy Parameters

Kalika Patrai and Indu Uprety

Abstract In this paper, we consider the problem of evaluating system reliability using Markov modeling approach, in which Times to failure and Times to repair of the operating units are assumed to follow exponential distribution. For this purpose, a method has been developed to construct a fuzzy set as an estimator for unknown parameters in the proposed statistical model. Using α -cut approach, the membership functions of MTTF and Availability are then constructed using Non-Linear Programming approach. With the proposed approach, explicit closed-form expressions of the system characteristics are obtained by inverting the interval limits of α -cuts of membership functions.

Keywords Markov modeling · MTTF · Availability · Fuzzy sets · Non-linear programming

1 Introduction

A gracefully degradable computing system is one which can demonstrate performance degradation [1]. The meaning of failure/success is hard to determine for gracefully degradable computing systems as the system can bear multiple faults in processing elements and/or links and still function with an acceptable performance degradation [2, 3]. Since the system degrades gracefully over long period, it is desirable to introduce coverage factors for successful detection of faulty units. If in

K. Patrai (✉)

Department of Computer Applications, Galgotia Institute of Management and Technology,
Greater Noida, India

e-mail: patrai.kalika@gmail.com

I. Uprety

School of Management, Gautam Buddha University, Greater Noida, India

e-mail: induuprety@yahoo.co.uk

a system, the failures are not successfully detected, located and recovered; then, this situation is called imperfect coverage [4]. Automatic recovery and reconfiguration mechanisms (detection, location, and isolation) play a crucial role in implementing fault tolerance because an uncovered fault may lead to a system or subsystem failure even when adequate redundancy exists. A lot of work has been done in this field by many researchers [5–7], Trivedi [8] considered the concept of detection and imperfect coverage and their effect on the repairable systems, whereas Ke et al. [9] used Bayesian approach to predict the performance measures of a repairable system with detection, imperfect coverage and reboot.

For the purpose of modeling and analysis, most of the classical reliability assessments are based on the probabilistic Markov models. In these models, the state space is completely defined, namely, either a failure or a success state and the failure state probability is calculated based on expected (crisp) values of mean time to failure/repair, wherein the uncertainties related to mean time to failure/repair affect the state probabilities [10–13]. Uncertainty is a condition where the possibility of error exists as a result of having less than total information about some existing environment. From pragmatic perspective, it is inadequate to represent the expected values of mean time to failure/repair by crisp numbers because it is difficult to obtain sufficient data to determine a fairly good distribution for representing these uncertainties [14, 15]. Similarly, the uncertainty in the failure occurrences and consequences and the uncertainties arising due to lack of up gradation of data also form another fundamental issue in reliability analysis. To deal with the problem of uncertainty and randomness in data, the fuzzy set theory is often utilized to quantify imprecision and uncertainty by making new assumptions, such as possibility assumption (possibility distribution), or the fuzzy-state assumption, in place of the probability assumption (probability distribution), or the binary-state assumption, respectively [16].

Fuzzy set theory was introduced by Lotfi Zadeh in 1965 to deal quantitatively with imprecision and uncertainty. It can be described as a body of concepts and techniques including fuzzy sets, fuzzy logic and fuzzy numbers that gives a form of mathematical precision to processes that are imprecise and ambiguous by the standards of classical methods [17]. A number of studies support the usefulness of fuzzy theory in reliability. In this regard, Wang et al. [18] presents an approach to predict the system reliability with fuzzy random variables to represent uncertainties. Wu [19] proposes a method to discuss system reliability in which the functioning probability of each component in the system is assumed as a non-negative fuzzy number. Some basic concepts of fuzzy set theory and their applications have been discussed in detail by Klir et al. [20]. Chen's [21] fuzzy arithmetic where a triangular fuzzy number represents component reliability simplified the fuzzy number operations of "extended algebraic operations" and interval arithmetic. Cai et al. [22] explicitly introduces fuzzy reliability theory and studies it for k-out-of-n, parallel, parallel-series, series and series-parallel systems.

Further, in the literature described above, the system parameters like, times to failure, and times to repair are required to follow certain possibility distributions, or probability distributions with fixed parameters. Parametric probability

distributions are used mostly in stochastic analyses of reliability of systems, where they are assumed to be fully known and corresponding properties of the system are analyzed. However, defining the exact parameter pattern is somewhat difficult and more subjective/possibilistic in nature, thus revealing uncertainty concerning these parameter patterns. In fuzzy theory, the grade of a membership function indicates a subjective degree of preference of a decision maker with a given tolerance [23]. Therefore, the analysis of repairable system in this paper is extended to the fuzzy environment and fuzzy membership functions are derived for system characteristics using mathematical programming approach.

2 Reliability Evaluation

2.1 System Modeling

A two-state Markov Model [23] is normally used to define a failure and success state of system operation. The transition between states is governed by probabilistic laws and the transition rates are normally represented by expected (crisp) values. Markov models are frequently used in reliability and maintainability analysis where events, such as the failure or repair of a module, can occur at any point in time. The Markov process can thus easily describe degraded states of operation, where the item has either partially failed or is in a degraded state where some functions are performed while others are not [24]. Also, when the coverage factor (the capacity of failure occurrence detection in a module of a system) is considered, the Markov model becomes a more suitable modeling approach, since the covered and uncovered failures of components are supposed to be mutually exclusive events. Trivedi et al. [25] developed a detailed availability model of the IBM SIP Application Server cluster by incorporating imperfect coverages for detection, failover and recovery. A sensitivity analysis of availability was carried out for important model parameters such as coverage factor and showed that the unavailability of the system decreases as the coverage factor increases.

To begin with the model, we consider a redundant degradable computing system with two identical and independent modules. Each module has only two states: failed or functioning. The time to breakdown for each module follows an exponential distribution with parameter λ . The system has a coverage factor c . When failure comes to same module, the system immediately takes reconfiguration operation with negligible time, to remove logically the faulty module whereas other fault free modules continue to do their work if the reconfiguration operation is performed successfully. The reliability of the system is computed using both the Markov and fuzzy modeling approaches and the results have been compared with respect to the effects of failure, repair and coverage factors.

2.2 Design and Characteristics of the Model

Let S_i represents the system state that two active (operational) modules are available. The system may then have three states: S_0 , S_1 , and S_2 (Fig. 1). The system represents the Markovian transition among the system states, where c represents the system coverage factor, the success probability of a reconfiguration operation.

The differential equations describing the system shown in Fig. 1 are:

$$\begin{aligned}\frac{dP_2}{dt} &= -2\lambda P_2 + \mu P_1 \\ \frac{dP_1}{dt} &= -(\lambda + \mu)P_1 + 2\lambda c P_2 \\ \frac{dP_0}{dt} &= \lambda P_1 + 2\lambda(1 - c)P_2\end{aligned}$$

where $P_i(t)$ is the probability of the system being in i th state ($i = 0, 1, 2$) at time t .

The MTTF. To obtain the reliability behavior of repairable system, the Laplace transform equations of the above equations can be obtained. Assume that state S_0 is systems down state. Let Z be the random variable representing the time to failure of the system; thus, the reliability function can be expressed as,

$$R_Z(t) = 1 - P_0(t) = P_1(t) + P_2(t), \quad t \geq 0.$$

The mean time to system failure (MTTF) is obtained using

$$MTTF = \int_0^{\infty} R_Z(t) dt = \frac{2\lambda c + \lambda + \mu}{2\lambda^2 + 2\lambda\mu - 2\lambda\mu c}$$

The Steady-State Availability. When one of the two working units fails the system goes into state S_1 and from S_1 due to the failure of second unit the system enters state S_0 , which is the down state of the system. At this state a failed unit is repaired with repair rate μ and the system again enters state S_1 . The state transition diagram of the system is shown in Fig. 2. To investigate the availability behavior of the system, the steady-state probabilities are:

$$\begin{aligned}P_0 &= \lambda P_1 + 2\lambda(1 - c)P_2 \\ (\lambda + \mu)P_1 &= 2\lambda c P_2 \\ 2\lambda P_2 &= \mu P_1\end{aligned}$$

where P_i is the probability of the system being in i th state ($i = 0, 1, 2$).

Fig. 1 State transition diagram for the reliability of a 2-unit repairable system

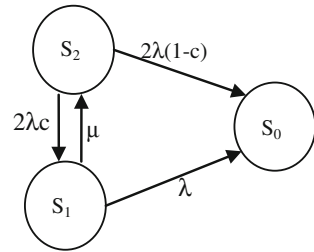
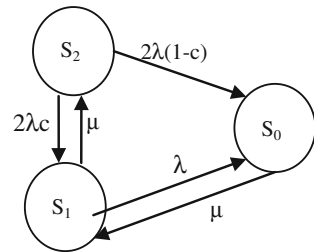


Fig. 2 State transition diagram for the steady-state availability of a 2-unit repairable system



Then the system availability is given by, $A = \frac{1}{2}P_1 + P_2$. On solving the steady-state equations, the expression for availability is obtained as:

$$A = \frac{\lambda\mu + \mu^2}{2\lambda^2 + 2\lambda\mu(2 - c) + \mu^2}$$

3 The Fuzzy Approach

In this section, we derive the explicit membership functions for MTTF and Availability. This approach has been used by several authors for reliability analysis of repairable systems with imperfect coverage and fuzzy parameters [26–29].

Referring to the applicability of degradable systems, we express the fuzzy sets in terms of the crisp universal sets as follows:

$$\begin{aligned} \tilde{\lambda} &= \{(x, \eta_{\tilde{\lambda}}(x)) \mid x \in X\} \\ \tilde{\mu} &= \{(y, \eta_{\tilde{\mu}}(y)) \mid y \in Y\} \end{aligned}$$

Let $f(x, y)$ denote the system characteristic of interest such as MTTF and availability. The function $f(\tilde{\lambda}, \tilde{\mu})$ is a fuzzy number as $\tilde{\lambda}$ and $\tilde{\mu}$ are fuzzy numbers. The membership function of the system characteristic $f(\tilde{\lambda}, \tilde{\mu})$, using Zadeh’s extension principle is defined as:

$$\eta_{f(\tilde{\lambda}, \tilde{\mu})}(z) = \sup_{x \in X, y \in Y} \min\{\eta_{\tilde{\lambda}}(x), \eta_{\tilde{\mu}}(y) \mid z = f(x, y)\}$$

Using this membership functions for MTTF and availability that is $\eta_{\tilde{M}}(z)$ and $\eta_{\tilde{A}}(z)$ respectively, can be written as:

$$\eta_{\tilde{M}}(z) = \sup_{x \in X, y \in Y | x > 0, y > 0} \min\{\eta_{\tilde{\lambda}}(x), \eta_{\tilde{\mu}}(y) | z = f(x, y)\} \quad (1)$$

where $f(x, y) = \frac{2xy+x+y}{2x^2+2xy(1-c)}$

And

$$\eta_{\tilde{A}}(z) = \sup_{x \in X, y \in Y | x > 0, y > 0} \min\{\eta_{\tilde{\lambda}}(x), \eta_{\tilde{\mu}}(y) | z = f(x, y)\} \quad (2)$$

where $f(x, y) = \frac{xy+y^2}{2x^2+2xy(2-c)+y^2}$.

3.1 Finding Alpha-Cuts Using NLP

In this paper we applied parametric non-linear programs (NLP), a mathematical programming technique to find out the alpha-cuts of the system characteristic based on the extension principle. The alpha-cuts $\tilde{\lambda}$ and $\tilde{\mu}$ as crisp intervals are as follows:

$$\lambda(\alpha) = [x_{\alpha}^L, x_{\alpha}^U] = \left[\min_{x \in X} \{x | \eta_{\tilde{\lambda}}(x) \geq \alpha\}, \max_{x \in X} \{x | \eta_{\tilde{\lambda}}(x) \geq \alpha\} \right] \quad (3)$$

$$\mu(\alpha) = [y_{\alpha}^L, y_{\alpha}^U] = \left[\min_{y \in Y} \{y | \eta_{\tilde{\mu}}(y) \geq \alpha\}, \max_{y \in Y} \{y | \eta_{\tilde{\mu}}(y) \geq \alpha\} \right] \quad (4)$$

The bounds of these intervals can be described as functions of α as:

$$x_{\alpha}^L = \min \eta_{\tilde{\lambda}}^{-1}(\alpha), \quad x_{\alpha}^U = \max \eta_{\tilde{\lambda}}^{-1}(\alpha), \quad y_{\alpha}^L = \min \eta_{\tilde{\mu}}^{-1}(\alpha)$$

and $y_{\alpha}^U = \max \eta_{\tilde{\mu}}^{-1}(\alpha)$.

Therefore, the α -cuts of \tilde{M} and \tilde{A} are used to construct their membership functions respectively because the membership functions defined in (1) and (2) are parameterized by α .

Using Zadeh's [30] extension principle $\eta_{\tilde{M}}(z)$ as well as $\eta_{\tilde{A}}(z)$ are the minimum of $\eta_{\tilde{\lambda}}(z)$ and $\eta_{\tilde{\mu}}(z)$. To derive the membership function $\eta_{\tilde{M}}(z)$, at least one of the following cases must hold such that,

$$Z = \frac{2\lambda c + \lambda + \mu}{2\lambda^2 + 2\lambda\mu - 2\lambda\mu c} \text{ satisfies } \eta_{\tilde{M}}(z) = \alpha.$$

Case (i): $(\eta_{\tilde{\lambda}}(z) = \alpha, \eta_{\tilde{\mu}}(z) \geq \alpha)$

Case (ii): $(\eta_{\tilde{\lambda}}(z) \geq \alpha, \eta_{\tilde{\mu}}(z) = \alpha)$

The lower and upper bounds of the α -cuts of $\eta_{\tilde{M}}(z)$ using parametric NLP for Case (i) and (ii) respectively, are:

$$M_{\alpha}^{L_1} = \min_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right) \tag{5}$$

$$M_{\alpha}^{U_1} = \max_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right) \tag{6}$$

$$M_{\alpha}^{L_2} = \min_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right) \tag{7}$$

$$M_{\alpha}^{U_2} = \max_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right) \tag{8}$$

From the definitions of $\lambda(\alpha)$ and $\mu(\alpha)$ in (3, 4), $x \in \lambda(\alpha)$ and $y \in \mu(\alpha)$ can be replaced by $x \in [x_{\alpha}^L, x_{\alpha}^U]$ and $y \in [y_{\alpha}^L, y_{\alpha}^U]$, respectively. Since the α -cuts form a nested structure with respect to α , for given $0 < \alpha_2 < \alpha_1 < 1$ we have,

$$[x_{\alpha_1}^L, x_{\alpha_1}^U] \subseteq [x_{\alpha_2}^L, x_{\alpha_2}^U] \text{ and } [y_{\alpha_1}^L, y_{\alpha_1}^U] \subseteq [y_{\alpha_2}^L, y_{\alpha_2}^U].$$

Thus the lower bounds in both the cases, i.e. (5) and (7) have the same smallest element and the upper bounds in both the cases, i.e. (6) and (8) have the same largest element. The lower and upper bounds of M are as follows:

$$M_{\alpha}^L = \min_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right), \tag{9}$$

suct that $x_{\alpha}^L \leq x \leq x_{\alpha}^U$ and $y_{\alpha}^L \leq y \leq y_{\alpha}^U$

$$M_{\alpha}^U = \max_{x \in X, y \in Y | x > 0, y > 0} \left(\frac{2xc + x + y}{2x^2 + 2xy - 2xyc} \right), \tag{10}$$

suct that $x_{\alpha}^L \leq x \leq x_{\alpha}^U$ and $y_{\alpha}^L \leq y \leq y_{\alpha}^U$

At least one of x and y must hit the boundaries of their α -cuts satisfying that $\eta_{\tilde{M}}(z) = \alpha$. The α -cuts of T are represented by the crisp intervals $[M_{\alpha}^L, M_{\alpha}^U]$ obtained from (9) and (10). By applying the results of Zimmermann [27], Kaufmann [31], and convexity properties to \tilde{M} , we have $M_{\alpha_1}^L \geq M_{\alpha_2}^L$ and $M_{\alpha_1}^U \geq M_{\alpha_2}^U$, where $0 < \alpha_2 < \alpha_1 < 1$. If both the bounds M_{α}^L and M_{α}^U in (9), and (10) are invertible with respect to α , then a left-shape function L(z) and a right-shape function R(z) are defined as $L(z) = (M_{\alpha}^L)^{-1}$ and $R(z) = (M_{\alpha}^U)^{-1}$.

Thus the membership function $\eta_{\tilde{M}}(z)$ is constructed as:

$$\eta_{\tilde{M}}(z) = \begin{cases} L(z), & M_{\alpha=0}^L \leq z \leq M_{\alpha=1}^L, \\ 1, & M_{\alpha=1}^L \leq z \leq M_{\alpha=1}^U, \\ R(z), & M_{\alpha=1}^U \leq z \leq M_{\alpha=0}^U \end{cases}$$

In most the cases a closed form of the membership function of M cannot be obtained because the lower and upper bounds of M i.e.; M_α^L and M_α^U cannot be obtained analytically. However, an approximation to the shapes of $L(z)$ and $R(z)$ can be made by collecting the numerical solutions of M_α^L and M_α^U at different possibility levels of α and hence the shape of $\eta_{\tilde{M}}(z)$ is obtained by the set of intervals $\{(M_\alpha^L, M_\alpha^U) | \alpha \in [0, 1]\}$.

3.2 Numerical Example

Consider a degradable system with two-units. Assume that the failure rate of each unit is a fuzzy number $\tilde{\lambda}$ and repair times are exponentially distributed with a fuzzy parameter $\tilde{\mu}$. A fault is covered with probability c , and is not covered with probability $1-c$. This system can be described by the model given in Fig. 1. Using the proposed approach the fuzzy system characteristics can be constructed.

The Fuzzy MTTF. For the system in Fig. 1, the failure and repair rates are assumed to be triangular fuzzy numbers, represented as:

$$\begin{aligned} \tilde{\lambda} &= [0.1, 0.2, 0.3] \\ \tilde{\mu} &= [0, 0.5, 1.0] \end{aligned}$$

Taking coverage factor $c = 0.9$, the α -cuts of $\tilde{\lambda}$ are constructed as,

$$\begin{aligned} [x_\alpha^L, x_\alpha^U] &= [0.1 + 0.1\alpha, 0.3 - 0.1\alpha] \\ [y_\alpha^L, y_\alpha^U] &= [0.5\alpha, 1 - 0.5\alpha] \end{aligned}$$

The MTTF attains it's minimum value when $x = x_\alpha^U$ and $y = y_\alpha^L$; and attains it's maximum value for $x = x_\alpha^L$ and $y = y_\alpha^U$. Thus the α -cuts of \tilde{M} are:

$$M_\alpha^L = \frac{84 + 22\alpha}{18 - 9\alpha + \alpha^2}, \text{ and } M_\alpha^U = \frac{128 - 22\alpha}{4 + 5\alpha + \alpha^2}$$

Since the inverse functions of M_α^L and M_α^U exists, the membership function is obtained as:

$$\eta_{\tilde{M}}(z) = \begin{cases} \frac{(22+9z)-\sqrt{9z^2+732z+484}}{2z}, & \frac{84}{18} \leq z \leq \frac{106}{10} \\ 1, & z = \frac{106}{10} \\ \frac{-(22+5z)+\sqrt{9z^2+732z+484}}{2z}, & \frac{106}{10} \leq z \leq \frac{128}{4} \end{cases}$$

and is shown in Fig. 3. The overall shape of the membership function turns out as expected. The crisp intervals for the failure-rate, the repair rate and the fuzzy MTTF at different levels of α are presented in Table 1.

The Fuzzy Availability. The membership function for the steady-state availability of the system given in Fig. 2 obtained by applying the proposed approach

Fig. 3 The membership function for fuzzy MTTF

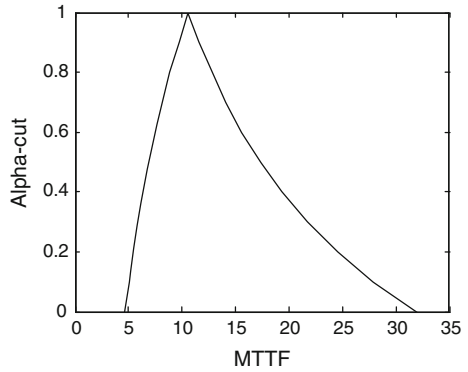
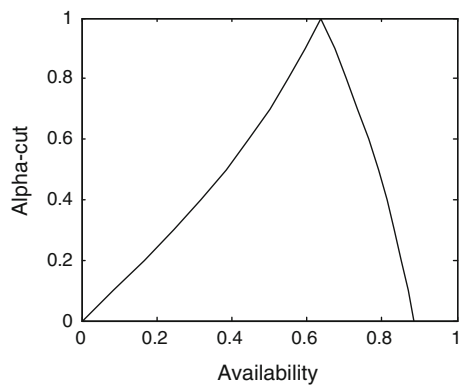


Table 1 α -cuts of failure, repair rate, fuzzy MTTF and fuzzy availability

α	λ_x^L	λ_x^U	μ_x^L	μ_x^U	M_x^L	M_x^U	A_x^L	A_x^U
0.0	0.1	0.3	0.0	1.00	4.6667	32.0000	0.0000	0.8871
0.1	0.11	0.29	0.05	0.95	5.0380	27.8936	0.0839	0.8707
0.2	0.12	0.28	0.10	0.90	5.4433	24.5238	0.1664	0.8528
0.3	0.13	0.27	0.15	0.85	5.8869	21.7174	0.2448	0.8335
0.4	0.14	0.26	0.20	0.80	6.3736	19.3506	0.3177	0.8124
0.5	0.15	0.25	0.25	0.75	6.9091	17.3333	0.3846	0.7895
0.6	0.16	0.24	0.30	0.70	7.5000	15.5978	0.4455	0.7643
0.7	0.17	0.23	0.35	0.65	8.1542	14.0926	0.5007	0.7368
0.8	0.18	0.22	0.40	0.60	8.8811	12.7778	0.5506	0.7064
0.9	0.19	0.21	0.45	0.55	9.6919	11.6219	0.5957	0.6732
1.0	0.20	0.20	0.50	0.50	10.6000	10.6000	0.6364	0.6364

Fig. 4 The membership function for fuzzy availability



and is shown in Fig. 4. The crisp intervals for fuzzy availability at different levels of α are given in Table 1.

Fig. 5 Relationship between failure-rate and MTTF

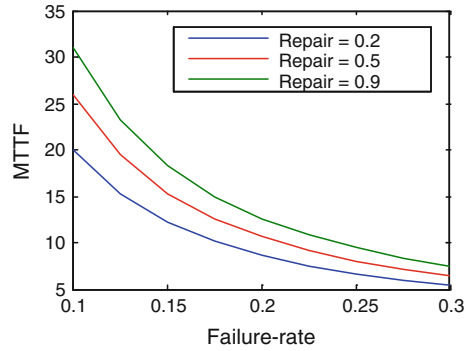
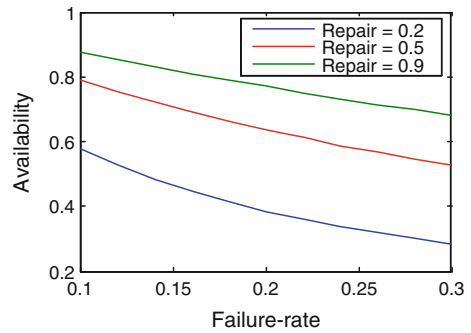


Fig. 6 Relationship between failure-rate and availability



4 Results and Discussion

The α -cuts of the failure rate, repair rate, MTTF and availability of a two-unit degradable system are obtained and given in Table 1. The graphs of traditional MTTF for the system given in Fig. 1 and steady-state availability for the system given in Fig. 2 for various repair rates are shown in Figs. 5 and 6 respectively for the coverage factor $c = 0.9$. It shows that if the traditional approach is used to find the MTTF and the availability for the above stated system, it will be 14.9865 and 0.7698 respectively for $\lambda = 0.2, \mu = 0.9$ and $c = 0.9$. However, if fuzzy parameters are used the MTTF ranges from 4.6667 to 32 and the availability ranges from 0 to 0.8871. The range for fuzzy MTTF \tilde{M} indicates that though the MTTF is fuzzy, its values cannot fall below 4.6667 or exceed 32.000 and the most possible value for the MTTF is 10.600 which is at $\alpha = 1$. From Table 1 the intervals of failure and repair rates are $\lambda_\alpha^L = 0.14, \lambda_\alpha^U = 0.26, \mu_\alpha^L = 0.20$ and $\mu_\alpha^U = 0.80$ respectively at $\alpha = 0.4$. This indicates that if the repair rates can be controlled in the range $[0.25, 0.75]$ time units, the system can get the required availability ranging from 0.3177 to 0.8124.

5 Conclusion

The fuzzy approach provides more effective, realistic and flexible measures as compared with the traditional approach based on crisp parameter values. This paper applies Zadeh's extension principle and the concepts of α -cuts to a two-unit degradable repairable system with coverage factor and using NLP models, the membership functions for the system characteristics are constructed. The proposed approach finds the explicit closed-form expressions for system characteristics by inverting the interval limits corresponding to the α -cuts of the membership functions. However, if the membership function intervals cannot be inverted explicitly to obtain the closed-form expressions of the system characteristics, the numerical solutions at different possibility levels can be collected to approximate the shapes of the membership functions.

References

1. Cai, K.Y., Wen, C.Y., Zhang, M.L.: Fuzzy reliability modeling of gracefully degradable computing systems. *Reliab. Eng. Syst. Saf.* **33**(1), 141–157 (1991)
2. Crk, V.: Reliability assessment from degradation data. Annual reliability and maintainability symposium, pp. 155–161 (2000)
3. Huang, W., Askin, R.G.: A generalized SSI reliability model considering stochastic loading and strength aging degradation. *IEEE Trans. Reliab.* **53**(1), 77–82 (2004)
4. Wang, K.H., Chiu, L.W.: Cost benefit analysis of availability systems with warm standby units and imperfect coverage. *Appl. Math. Comput.* **172**, 1239–1256 (2006)
5. Pham, H.: Reliability analysis of a high voltage system with dependent failures and imperfect coverage. *Reliab. Eng. Syst. Saf.* **37**, 25–28 (2001)
6. Akhtar, S.: Reliability of K-out-of-n: G system with imperfect fault coverage. *IEEE Trans. Reliab.* **43**, 101–106 (1994)
7. Moustafa, M.: Reliability analysis of K-out-of-N: G systems with dependent failures and imperfect coverage. *Reliab. Eng. Syst. Saf.* **58**, 15–17 (1997)
8. Trivedi, K.S.: Probability and Statistics with Reliability, Queuing and Computer Science Applications. Wiley, New York (2002)
9. Ke, J.C., Lee, S.L., Hsu, Y.L.: On a repairable system with detection, imperfect coverage and reboot: Bayesian approach. *Simul. Model. Pract. Theory* **16**, 353–367 (2008)
10. Liu, T.H., Ke, J.C., Hsu, Y.L.: Bootstrapping computation of availability for a repairable system with standby subject to imperfect switching. *Commun. Stat. Simul. Comput.* **40**, 469–483 (2011)
11. Chen, A., Wu, G.S.: Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *Int. J. Prod. Res.* **45**(15), 3351 (2007)
12. Li, N., Xie, W.C., Haas, R.: Reliability-based processing of Markov chains for modeling pavement network deterioration. *Transp. Res. Rec.* **1524**, 203–213 (1996)
13. Welte, T.M., Vatn, J., Heggset, J.: Markov state model for optimization of maintenance and renewal of hydro power components. In: International Conference on Probabilistic Methods Applied to Power Systems, pp. 1–7 (2006)
14. Ei-Hawary, M.E.: Electric Power Applications of Fuzzy System. Wiley-IEEE Press, New York (1998)

15. Mohanta, D.K., Sandhu, P.K., Chakrabarti, R.: Fuzzy reliability evaluation of captive power plant maintenance scheduling incorporating uncertain forced outage rate and load representation. *Elec. Power Syst. Res.* **72**, 73–84 (2004)
16. Cai, K.Y., Wen, C.Y., Zhan, M.L.: Fuzzy variables as a basis for a theory of fuzzy reliability in the possibility context. *Fuzzy Sets Syst.* **42**, 145–172 (1991)
17. Cheng, C., Mon, D.: Fuzzy system reliability analysis by interval of confidence. *Fuzzy Sets Syst.* **56**, 29–35 (1993)
18. Wang, Z., Huang, H.Z., Du, L.: Reliability analysis on competitive failure processes under fuzzy degradation data. *Appl. Soft Comput.* **11**(3), 2964–2973 (2011)
19. Wu, H.C.: Fuzzy reliability analysis based on closed fuzzy numbers. *Inf. Sci.* **103**, 35–159 (1997)
20. Klir, G.J., Yuan, B.: *Fuzzy Set and Fuzzy Logic: Theory and Applications*. Prentice-Hall of India, New-Delhi (2005)
21. Chen, S.: Fuzzy system reliability analysis using fuzzy number arithmetic operations. *Fuzzy Sets Syst.* **64**, 31–38 (1994)
22. Cai, K.Y., Wen, C.Y., Zhang, M.L.: Postbist reliability behavior of typical systems with two types of failure. *Fuzzy Sets Syst.* **43**, 17–32 (1991)
23. Huang, H.I., Lin, C.H., Ke, J.C.: Parametric nonlinear programming approach for a repairable system with switching failure and fuzzy parameters. *Appl. Math. Comput.* **183**, 508–517 (2006)
24. Billinton, R., Bollinger, K.E.: Transmission system reliability evaluation using Markov processes. *IEEE Trans. Power Apparatus Syst.* **87**(2), 538–547 (1968)
25. Trivedi, K., Wang, D., Hunt, D.J., Rindos, S., Smith, W.E., Vashaw, B.: Availability modeling of SIP protocol on IBM® Websphere®. Dependable computing, 14th IEEE Pacific Rim international symposium, pp 323–330 (2008)
26. Zimmermann, H.J.: *Fuzzy Set Theory and Its Applications*, 4th edn. Kluwer Academic, Boston (2001)
27. Ke, J.C., Huang, H., Lin, C.H.: Redundant repairable system with imperfect coverage and fuzzy parameters. *Appl. Math. Model.* **32**, 2839–2850 (2008)
28. Wang, S., Watada, J.: Reliability optimization of a series parallel system with fuzzy random lifetimes. *Int. J. Innovative Comput. Inf. Control* **5**(6), 1547–1558 (2009)
29. Sharma, M.K., Pandey, D.: Fuzzy reliability and fuzzy availability of a three unit degraded systems. *Int. J. Math. Sci. Eng. Appl. (IJMSEA)* **3**(2), 199–213 (2009)
30. Zadeh, L.A.: Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets Syst.* **1**, 3–28 (1978)
31. Kaufmann, A.: *Introduction to the Theory of Fuzzy Subsets*, vol. 1. Academic Press, New York (1975)

An Inventory Model with Time Dependent Demand Under Inflation and Trade Credits

Yogendra Kumar Rajoria, Seema Saini and S. R. Singh

Abstract The traditional inventory model assumes that a retailer accepts the offer of delay in payments since he does not have the capital with him. Even when he has to make the payments at the end of credit limit, he takes a loan to pay off the supplier. The model focuses on commodities having quadratic demand with trade credit policies. The commodities considered in this model are perishable stock whose deterioration starts immediately as soon as you store the items. We have considered all the factors which one retailer must kept in mind deciding his inventory level, the concept of inflation and time value of money is also considered.

Keywords Quadratic demand · Inflation and time value of money

1 Introduction

Inventory is spread throughout the supply chain from raw materials to semi-finished and final products that suppliers, manufacturers, distributors and retailers hold. Implementation of a good inventory management policy is highly effective in reducing the inventory costs. In this study, the problem of determining the optimal replenishment policy for deteriorating items with both constant and stock-

Y. K. Rajoria (✉) · S. Saini

Department of Mathematics, Graphic Era University, Dehradun, Uttarakhand, India
e-mail: yogendrarajo@gmail.com

S. Saini

e-mail: sainiseema1@gmail.com

S. R. Singh

Department of Mathematics, D. N. College, Meerut, Uttar Pradesh, India
e-mail: shivrajpundir@gmail.com

dependent demands is considered. Here shortages are allowed without backlogging with inflationary environment. The model considered in the problem is one in which items considered are deteriorating items with constant rate of deterioration, but these deteriorating items like fruits and vegetables are having time dependent demand rate which means that with passage of time their demand increases rapidly. We have also considered the concept of permissible delay, although this concept is not new but most of the earlier papers don't consider this. This help's big retailer in reducing their inventory cost. Thus, in this chapter, we have considered most of the factors which can affect the total inventory cost. Some latest works to the relevant area are summarized here [1]. Presented an inventory model for deteriorating items with partial backlogging under inflation [2]. Investigated an inventory model for deteriorating item with two types of retailers [3]. Formulated an optimal selling price and lot size with time varying deterioration and partial backlogging [4]. Presented a supply chain model for perishable items with ramp type demand rate [5]. Obtained an optimal price and lot size determination for a perishable product under conditions of finite production, partial backordering and lost sale [6]. Studied an optimal replenishment policy for life time inventory with partial backlogging [7]. Presented reserve money for an EOQ model in an inflationary environment under supplier credits [8]. Formulated perishable inventory model with quadratic demand partial backlogging and permissible delay in payments [9]. Supply chain inventory model with price-dependent consumption rate with ameliorating and deteriorating items and two levels of storage [10]. Studied inventory policy for an item with inflation induced purchasing price, selling price and demand with immediate part payment.

2 Assumptions and Notations

In developing the mathematical model of the inventory system the following assumptions are made:

1. Shortages are allowed.
2. Time value of money and inflation both are considered.
3. Replenishment occurs instantaneously at an infinite rate.
4. The deteriorated units are not repaired or replaced.
5. Interest paid is more than interest earned.
6. A single item is held in stock.
7. The item deteriorates with a fixed rate ϑ .

In addition, the following notations are used:

f	Inflation rate
i	Inventory carrying rate
A	Ordering cost of inventory, Rupees/order
S	Period with shortage

T	Length of inventory cycle, time units
T_1	Length of period with positive stock of the items
θ	Rate of deterioration per unit time
i_p	$I_p - r$
r	Discount rate representing the time value of money
I_p	Nominal interest paid per Rupees per unit time at time
$a + bt + kt^2$	Time dependent demand where a, b and k are constant
i_e	$I_e - r$
I_e	Nominal interest at time $t = 0$
$I_e(t)$	Rate of interest earned at time t Rupees per Rupee per unit time
$I_p(t)$	Interest rate paid at time t Rupees per Rupee per unit time
I_T^i	Total interest earned per cycle with inflation
M	Permissible delay in payments permitted in settling the account
P_T^i	Interest paid per cycle under the effect of inflation
c	Unit cost of per item at time $t = 0$, Rupees/unit
C_0^b	Present value of the inflated backorder cost c_b' Rupees/unit
C_D^i	Total cost of deterioration per cycle with inflation
C_H^i	Total holding cost per cycle with inflation
D_T	Amount of materials deterioration during a cycle time, T

3 Mathematical Formulation of the Model

In this model quadratic demand pattern is considered with constant rate of deterioration. Depletion of the inventory occurs due to demand (supply) as well as due to deterioration which occurs only when there is inventory i.e. during the period $[0, T_1]$. For this period inventory at any time is given by:

$$I'(t) + \theta I(t) = -(a + bt + kt^2) \quad 0 \leq t \leq T_1 \tag{1}$$

At $t = 0$, $I = I_0$, initial inventory

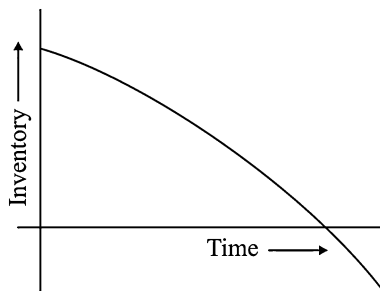
$$I(t)e^{\theta t} = I_0 + \frac{a}{\theta} - \frac{b}{\theta^2} + \frac{2k}{\theta^3} - \frac{ae^{\theta t}}{\theta} - \frac{bte^{\theta t}}{\theta} + \frac{be^{\theta t}}{\theta^2} + \frac{k}{\theta^3} [\theta^2 t^2 - 2\theta t + 2] \tag{2}$$

At, $t = T_1$, we have $I(T_1) = 0$, (Fig. 1)

$$I_0 = \frac{a}{\theta} [e^{\theta T_1} - 1] + \frac{b}{\theta^2} [(\theta T_1 - 1)e^{\theta T_1} + 1] + \frac{K}{\theta^2} [(\theta^2 T_1^2 - 2\theta T_1 + 2)e^{\theta T_1} - 2] = Q - b'$$

where b' being the maximum backorder (shortage) level permitted. Substituting the value of I_0 in Eq. (2), we get

Fig. 1 Depletion of the inventory



$$I(t) = \frac{a}{\theta} \left[e^{\theta(T_1-t)} - 1 \right] + \frac{b}{\theta^2} \left[(\theta T_1 + 1)e^{\theta(T_1-t)} - \theta t + 1 \right] + \frac{k}{\theta^3} \left[(\theta^2 T_1^2 - 2\theta T_1 + 2)e^{\theta(T_1-t)} - \theta t^2 + \theta t - 2 \right] \tag{3}$$

And $I(t) = 0$ when $T_1 \leq t \leq T$. Total demand during T_1 is $\int_0^{T_1} d(a + bt + kt^2) dt$.

Thus it can easily be seen that the amount of items deteriorates during one cycle is given

$$D_T = I_0 - \int_0^{T_1} (a + bt + kt^2) dt \tag{4}$$

For inflation rate f , the continuous time inflation factor for the time period is e^{ft} which means that an item costing Rupees c at time $t = 0$ will cost ce^{ft} at time t , for a discount rate r , representing the time value of money, the present value of an amount at time t , is ce^{-rt} . Hence, the present value of the inflated price of an item at time $t = 0$, $ce^{ft}e^{-rt}$ is given by

$$c_0 = ce^{(f-r)t} = ce^{Rt}, \quad \text{where } R = f - r \tag{5}$$

In this cost c is inflated by the net inflation factor R . Thus R is the present value of the inflation rate similarly, the present value of the inflated backorder cost c_b , C_0^b is given by

$$C_0^b = c_b e^{(f-r)t} = c_b e^{Rt} \tag{6}$$

which gives a backorder cost anytime t under the effect of inflation

3.1 The Inflation Model

There are two distinct cases in the type of inventory system.

1. Payment at on before the total depletion of inventory
2. After-depletion payment

Case 1

In this situation credit period expires on or before the depleted completely to zero. Therefore total variable cost is given by the sum of ordering cost, carrying cost, backorder cost, deterioration cost and interest paid minus interest earned.

1. As the ordering is made at time $t = 0$, therefore it is not affected by inflation. Hence ordering cost is fixed as Rupees A per order.
2. $I_0 = \frac{a}{\theta} [e^{\theta T_1} - 1] + \frac{b}{\theta^2} [(\theta T_1 - 1)e^{\theta T_1} + 1] + \frac{K}{\theta^2} [(\theta^2 T_1^2 - 2\theta T_1 + 2)e^{\theta T_1} - 2]$,

The value of this inventory at time $t = 0$, is cI_0 .

The present value of the items sold is $\int_0^{T_1} c_0(a + bt + kt^2)dt$. Hence the cost of deterioration per cycle time T under inflation, C_D^i is given

$$\begin{aligned}
 C_D^i &= cI_0 - \int_0^{T_1} c_0(a + bt + kt^2)dt = \frac{ca}{\theta} [e^{\theta T_1} - 1] + \frac{cb}{\theta^2} [(\theta T_1 - 1)e^{\theta T_1} + 1] \\
 &+ \frac{ck}{\theta^3} [(\theta^2 T_1^2 - 2\theta T_1 + 2)e^{\theta T_1} - 2] - ca \left(\frac{e^{RT_1} - 1}{R} \right) \\
 &- \frac{cb}{R^2} [RT_1 e^{RT_1} - e^{RT_1} + 1] - \frac{kc}{R^3} [R^2 T_1^2 e^{RT_1} - 2RT_1 e^{RT_1} - 2e^{RT_1} - 2]
 \end{aligned} \tag{7}$$

3. The holding cost under inflation is given by

$$\begin{aligned}
 C_H^i &= i \int_0^{T_1} c_0 I(t) e^{-Rt} dt \\
 &= \frac{ica}{\theta} \left[\left(\frac{e^{RT_1} - e^{\theta T_1}}{R - \theta} \right) - \left(\frac{e^{RT_1} - 1}{R} \right) \right] + \frac{icb}{\theta^2} \left[\begin{aligned} &(\theta T_1 - 1) \left(\frac{e^{RT_1} - e^{\theta T_1}}{R - \theta} \right) \\ &- \frac{1}{R^2} ((R\theta T_1 - R - \theta)e^{RT_1} + R + \theta) \end{aligned} \right] \\
 &+ \frac{ick}{\theta^3} \left[\begin{aligned} &(\theta^2 T_1^2 - 2\theta T_1 + 2) \left(\frac{e^{RT_1} - e^{\theta T_1}}{R - \theta} \right) - \frac{\theta^2}{R^3} \{ (R^2 T_1^2 - 2RT_1 + 2)e^{RT_1} - 2 \} \\ &+ \frac{2\theta}{R^2} \{ (RT_1 - 1)e^{RT_1} + 1 \} \end{aligned} \right] \\
 &- 2 \frac{ick}{\theta^3} \left(\frac{e^{RT_1} - 1}{R} \right)
 \end{aligned} \tag{8}$$

4. The interest payable rate at time t is $ce^{I_p t} - 1$ Rupees. per Rupee, the present value of c at $t = 0$ of interest payable at time is $I_p^i(t) = (e^{I_p t} - 1)e^{-rt}$ Rupees per rupee. Therefore the interest payable per cycle for the inventory not sold after the due date M is given by

$$\begin{aligned}
 p_T^i &= \int_M^T cI_p(t)I(t)dt = \int_M^{T_1} cI_p(t)I(t)dt \\
 &= \frac{ck}{\theta^3} [\theta^2 T_1^2 - 2\theta T_1 + 2] e^{\theta T_1} + \left[\frac{e^{(i_p - \theta)T_1} - e^{(i_p - \theta)M}}{i_p - \theta} \right] + \left[\frac{e^{-(r + \theta)T_1} - e^{-(r + \theta)M}}{r + \theta} \right] \\
 &\quad + \frac{ca}{\theta} \left[e^{\theta T_1} \left\{ \left(\frac{e^{(i_p - \theta)T_1} - e^{(i_p - \theta)T_1}}{i_p - \theta} \right) + \left(\frac{e^{-(r + \theta)T_1} - e^{-(r + \theta)M}}{(r + \theta)} \right) \right\} \right. \\
 &\quad \left. - \left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p} \right) - \left(\frac{e^{-r T_1} - e^{-r M}}{r} \right) \right] \\
 &\quad + \frac{bc}{\theta^2} \left[\theta \left\{ \left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p^2} \right) - \left(\frac{e^{-r T_1} - e^{-r M}}{r^2} \right) \right\} \right. \\
 &\quad \left. + \left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p} \right) + \left(\frac{e^{-r T_1} - e^{-r M}}{r} \right) \right] \\
 &\quad + \frac{cb}{\theta^2} \left[\left((\theta T_1 - 1) e^{\theta T_1} \left\{ \left(\frac{e^{(i_p - \theta)T_1} - e^{(i_p - \theta)T_1}}{i_p - \theta} \right) + \left(\frac{e^{-(r + \theta)T_1} - e^{-(r + \theta)M}}{(r + \theta)} \right) \right\} \right) \right. \\
 &\quad \left. - \theta \left\{ \left(\frac{T_1 e^{i_p T_1} - M e^{i_p M}}{i_p} \right) + \left(\frac{T_1 e^{-r T_1} - M e^{-r M}}{r} \right) \right\} \right] \quad (9) \\
 &\quad - \frac{ck}{\theta} \left[\left(\frac{T_1^2 e^{i_p T_1} - M^2 e^{i_p M}}{i_p} \right) - 2 \left(\frac{T_1 e^{i_p T_1} - M e^{i_p M}}{i_p^2} \right) \right] \\
 &\quad - \frac{ck}{\theta} \left[2 \left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p^3} \right) + \left(\frac{T_1^2 e^{-r T_1} - M^2 e^{-r M}}{r} \right) \right. \\
 &\quad \left. - 2 \left(\frac{T_1 e^{-r T_1} - M e^{-r M}}{r^2} \right) + 2 \left(\frac{e^{-r T_1} - e^{-r M}}{r^3} \right) \right] \\
 &\quad + \frac{ck}{\theta^2} \left[\left(\frac{T_1 e^{i_p T_1} - M e^{i_p M}}{i_p} \right) - \left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p^2} \right) \right. \\
 &\quad \left. + \left(\frac{T_1 e^{-r T_1} - M e^{-r M}}{r} \right) + \left(\frac{e^{-r T_1} - e^{-r M}}{r^2} \right) \right] \\
 &\quad - \frac{2ck}{\theta^3} \left[\left(\frac{e^{i_p T_1} - e^{i_p M}}{i_p} \right) - \left(\frac{e^{-r T_1} - e^{-r M}}{r} \right) \right]
 \end{aligned}$$

5. The present value of the interest earned at time t , $I_e(t)$ is $(e^{I_e t} - 1)e^{-rt}$, considering inflated unit cost at time t as ce^{Rt} , the interest earned per cycle, I_T^i , is the interest earned during the positive inventory, and it is given by

$$\begin{aligned}
 I_T^i &= \int_0^{T_1} c_0 I_e(t) [a + bt + kt^2] t dt \\
 &= ca \left[\frac{T_1 e^{(R+ie)T_1}}{R + ie} - \frac{e^{(R+ie)T_1}}{(R + ie)^2} + \frac{1}{(R + ie)^2} - \frac{T_1 e^{(R-r)T_1}}{R - r} + \frac{e^{(R-r)T_1}}{(R - r)^2} - \frac{1}{(R - r)^2} \right] \\
 &\quad + cb \left[\frac{T_1^2 e^{(R+ie)T_1}}{R + ie} - 2T_1 \frac{e^{(R+ie)T_1}}{(R + ie)^2} + \frac{2e^{(R+ie)T_1}}{(R + ie)^3} - \frac{2}{(R + ie)^3} \right. \\
 &\quad \left. - \frac{T_1^2 e^{(R-r)T_1}}{(R - r)^2} + \frac{2T_1 e^{(R-r)T_1}}{(R - r)^2} - \frac{2e^{(R-r)T_1}}{(R - r)^3} + \frac{2}{(R - r)^3} \right] \\
 &\quad + kc \left[\frac{T_1^3 e^{(R+ie)T_1}}{R + ie} - \frac{3T_1^2 e^{(R+ie)T_1}}{(R + ie)^2} + \frac{6T_1 e^{(R+ie)T_1}}{(R + ie)^3} - \frac{6e^{(R+ie)T_1}}{(R + ie)^4} \right. \\
 &\quad \left. - \frac{6T_1 e^{(R-r)T_1}}{(R - r)^3} + \frac{6e^{(R-r)T_1}}{(R - r)^4} - \frac{6}{(R - r)^4} \right]
 \end{aligned} \tag{10}$$

6. The back order cost per cycle under inflation, C_B^i given by

$$\begin{aligned}
 C_B^i &= \int_0^{T-T_1} c_0 bt(a + bt + kt^2) dt \text{ Since backorder starts at } t = T_1 \\
 &= \frac{c_b a e^{RT_1}}{R^2} \left[(RT - RT_1 - 1)e^{R(T-T_1)} + 1 \right] \\
 &\quad + bc_b \left[\left[\frac{(T - T_1)^2 e^{RT}}{R} \right] - \left[\frac{2(T - T_1)e^{RT}}{R^2} \right] + \left[\frac{2[e^{RT} - e^{RT_1}]}{R^3} \right] \right] - kc_b d \left[\frac{6(e^{RT} - e^{RT_1})}{R^4} \right] \\
 &\quad + kc_b d \left[\frac{(T - T_1)^3 e^{RT}}{R} \right] - kc_b \left[\frac{3(T - T_1)^2 e^{RT}}{R^2} \right] + kc_b d \left[\frac{6(T - T_1)e^{RT}}{R^3} \right]
 \end{aligned} \tag{11}$$

The total variable cost per cycle C_{VT} , is defined as

$$C_{VT}(T_1, T) = A + C_D^i + C_H^i + P_T^i - I_T^i + C_B^i \tag{12}$$

Substituting the values Eq. (12) we have C_{VT} in terms of T_1 and T .

As inflation and time value of money is considered in each cycle of replenishment, Thus we consider the effect over the time horizon, NT suppose there N complete cycles during a year. The total cost during a year is given by.

$$C_T = C_{VT} \times \left\{ 1 + e^{RT} + e^{2RT} + e^{3RT} + \dots + e^{(N-2)RT} \right\} = C_{VT} \times \left[\frac{1 - e^{NRT}}{1 - e^{RT}} \right] \tag{13}$$

The values of T and T_1 which minimize may be obtained by simultaneously solving

$$\begin{aligned} \frac{\partial C_T(T_1, T)}{\partial T} = 0 \text{ and } \frac{\partial C_T(T, T_1)}{\partial T} = 0 \\ \frac{\partial C_T}{\partial T} = \frac{\partial C_{VT}}{\partial T} \times \left(\frac{1 - e^{NRT}}{1 - e^{RT}} \right) + C_{VT} \times \frac{\partial}{\partial T} \left(\frac{1 - e^{NRT}}{1 - e^{RT}} \right) \\ \frac{\partial}{\partial T} \left(\frac{1 - e^{NRT}}{1 - e^{RT}} \right) = \frac{(1 - e^{NRT})NR}{(1 - e^{RT})^2} \end{aligned} \tag{14}$$

$$\begin{aligned} \frac{\partial C_{VT}}{\partial T} = ac_b(T - T_1)e^{RT} + bc_b(T - T_1)^2e^{RT} + kc_b(T - T_1)^3e^{RT} = s \\ \frac{\partial C_T}{\partial T} = s \left(\frac{1 - e^{NRT}}{1 - e^{RT}} \right) + \frac{(1 - e^{NRT})NR}{(1 - e^{RT})^2} C_{VT}(T, T_1) = 0 \end{aligned} \tag{15}$$

In which $C_{VT}(T_1, T)$ is expressed in known quantities from Eq. (12) similarly

$$\frac{\partial C_T}{\partial T_1} = \left(\frac{1 - e^{NRT}}{1 - e^{RT}} \right) \frac{\partial C_{VT}}{\partial T_1} = 0 \tag{16}$$

Equations (15) and (16) can be used to obtain optimal solution.

3.2 Optimal Solutions

By direct search approach it can be shown that given by (12) is convex in feasible domain of T and T_1 . Therefore the optimum value of T and T_1 minimizing C_T can be obtained by simultaneously solving equations $\frac{\partial C_T}{\partial T} = 0$ and $\frac{\partial C_T}{\partial T_1} = 0$. The expression for the total cost involves higher order exponential terms, it is not easy to evaluate the Hessians in closed form, to conclude about its positive definiteness directly, and thus it is not trivial to see whether the total cost function is convex. By solving these equations we can decide our initial inventory level with allowable shortages so that to minimize the total inventory cost.

Case 2 $T_1 < M$ (i.e., after depletion payment)

The deterioration cost C_D^i , carrying cost, C_H^i and the backorder cost C_B^i per cycle are the same as in the Eqs. (7), (8) and (11) respectively. The interest paid P_T^i per cycle is equal to zero when $T_1 < M$ because the supplier can pay in full at the end of permissible delay, M . The interest earned per cycle is the interest earned during the positive inventory period plus the interest earned from the cash invested during the time period (T_1, M) after the inventory is exhausted at time I_T^i , it is given by

$$\begin{aligned}
 I_T^i &= \int_0^T c_o J_e(t) t(a + bt + kt^2) dt + (e^{i_e(M-T_1)} - 1) \int_0^{T_1} c_0 t(a + bt + kt^2) dt \\
 &= ca \left[\frac{T_1 e^{(R+ie)T_1}}{R + ie} - \frac{e^{(R+ie)T_1}}{(R + ie)^2} + \frac{1}{(R + ie)^2} - \frac{T_1 e^{(R-r)T_1}}{R - r} + \frac{e^{(R-r)T_1}}{(R - r)^2} - \frac{1}{(R - r)^2} \right] \\
 &\quad + cb \left[\frac{T_1^2 e^{(R+ie)T_1}}{R + ie} - 2T_1 \frac{e^{(R+ie)T_1}}{(R + ie)^2} + \frac{2e^{(R+ie)T_1}}{(R + ie)^3} \right. \\
 &\quad \left. - \frac{2}{(R + ie)^3} - \frac{T_1^2 e^{(R-r)T_1}}{(R - r)^2} + \frac{2T_1 e^{(R-r)T_1}}{(R - r)^2} - \frac{2e^{(R-r)T_1}}{(R - r)^3} + \frac{2}{(R - r)^3} \right] \\
 &\quad + kc \left[\frac{T_1^3 e^{(R+ie)T_1}}{R + ie} - \frac{3T_1^2 e^{(R+ie)T_1}}{(R + ie)^2} + \frac{6T_1 e^{(R+ie)T_1}}{(R + ie)^3} - \frac{6e^{(R+ie)T_1}}{(R + ie)^4} \right. \\
 &\quad \left. + \frac{6}{(R + ie)^4} - \frac{T_1^3 e^{(R-r)T_1}}{(R - r)} + \frac{3T_1^2 e^{(R-r)T_1}}{(R - r)^2} - \frac{6T_1 e^{(R-r)T_1}}{(R - r)^3} + \frac{6e^{(R-r)T_1}}{(R - r)^4} - \frac{6}{(R - r)^4} \right] \\
 &\quad + (e^{i_e(M-T_1)} - 1) \left[ac \left(\frac{T_1 e^{RT_1}}{R} - \left(\frac{e^{RT_1} - 1}{R^2} \right) \right) \right. \\
 &\quad \left. + cb \left(\frac{T_1^2 e^{RT_1}}{R} - \frac{2T_1 e^{RT_1}}{R^2} + \frac{2(e^{RT_1} - 1)}{R^3} \right) \right] \\
 &\quad + (e^{i_e(M-T_1)} - 1) ck \left(\frac{T_1^3 e^{RT_1}}{R} - \frac{3T_1^2 e^{RT_1}}{R^2} + \frac{6T_1 e^{RT_1}}{R^3} + \frac{6(e^{RT_1} - 1)}{R^4} \right)
 \end{aligned} \tag{17}$$

Incorporating the modification of I_T^i in Eq. (17) and $P_T^i = 0$ into Eq. (12) because of the changes in assumption for Case 2, value in equation has changed from that in equation. The solution structure for the total annual cost remains the same as in previous case.

4 Numerical Illustration

Based on the previous published research work in the standard journals we consider the following data in appropriate units. Taking $a = 100$, $b = 1$, $k = 0.5$, $T = 6$, $R = 0.05$, $T_1 = 5$, we obtain the Table 1, which shows the effect of increasing permissible delay on variable and total cost.

Some observations based on the applications are as follows:

1. This model is helpful in determining the optimal ordering quantity by the purchaser for items having quadratic demand.
2. This model is helpful in study of inventory of perishable items.
3. The inflation which is effecting most of the economies of the world has been consider in this paper.
4. If we put $k = 0$, the above model changes to linear demand model.
5. If we put K and b both equal to zero the given model is converted to constant demand model.

Table 1 Effect of increasing permissible delay on variable and total cost

The table below permissible delay M	Total variable cost C_{VT}	Total cost C_T
1	49,169.69	7,203.35
2	48,658.53	7,128.47
3	48,231.11	7,065.85
4	47,421.26	6,947.21

5 Conclusion

This model will be helpful for purchaser in deciding the optimal order quantity in which demand is increasing rapidly with time with allowable shortages under the effect of inflation. We have considered two cases with respect to depletion time T_1 which is either greater than or less than equal to permissible delay. Thus, this model will help inventory managers in deciding their stock of inventory for perishable items with time dependent demand rate. This model can also be converted to linear demand pattern, plus we can also use this model for constant demand. This model although prepared for retailers but can be used by both suppliers and retailers. This model can be further extended for items having ramp type demand and even for some probabilistic demand pattern. The study can be extended for items having time dependent deterioration.

References

1. Chern, M.S., Yang, H.L., Teng, J.T., Papachristos, S.: Partial backlogging inventory lot-size models for deteriorating items with fluctuating demand under inflation. *Eur. J. Oper. Res.* **191**, 127–141 (2008)
2. Li, J., Mao, J.: An inventory model of permissible item with two types of retailers. *J. Chin. Inst. Ind. Eng.* **26**, 176–183 (2009)
3. Sana, S.: Optimal Selling Price and Lot Size with time varying Deterioration and Partial Backlogging. *Appl. Math. Comput.* **217**, 185–194 (2010)
4. Skouri, K., Konstantaras, I., Papachristos, S., Teng, J.T.: Supply chain models for deteriorating products with ramp type demand rate under permissible delay in payments. *Expert Syst. Appl.* **38**, 14861–14869 (2011)
5. Ghosh, S.K., Khanra, S., Chaudhuri, S.K.: Optimal price and lot size determination for a perishable product under conditions of finite production, partial backordering and lost sale. *Appl. Math. Comput.* **217**, 6047–6053 (2011)
6. Rajoria, Y.K., Saini, S., Singh, S.R.: Optimal replenishment policy for life time inventory with partial-backlogging. *Int. Trans. Math. Sci. Comput.* **5**(2), 207–216 (2012)
7. Singh, S.R., Jain, R.: On reserve money for an EOQ model in an inflationary environment under supplier credits. *Opsearch* **46**, 303–320 (2009)
8. Singh, S.R., Singh, T.J., Singh, C.: Perishable inventory model with quadratic demand, partial backlogging and permissible delay in payments. *Int. Rev. Pure Appl. Math.* **3**, 199–212 (2007)

9. Singh, S.R., Vishnoi, M.: Supply chain inventory model with price-dependent consumption rate with ameliorating and deteriorating items and two levels of storage. *Int. J. Procurement Manag.* **6**(2), 129–151 (2013)
10. Guria, A., Das, B., Mondal, S., Maiti, M.: Inventory policy for an item with inflation induced purchasing price, selling price and demand with immediate part payment. *Appl. Math. Model.* **37**, 240–257 (2013)

Digital Image Processing Method for the Analysis of Cholelithiasis

Neha Mehta, S. V. A. V. Prasad and Leena Arya

Abstract The occurrence of Cholelithiasis is the commonest biliary disease to be reported in India. Our research is aimed to assess the potential association between the gallbladder stone and the blood analysis through a cohort study. An attempt is also made to illustrate the importance of image enhancement technique by grey level mapping in digital image processing for the better analysis of biliary system and its related complications.

Keywords Cholelithiasis · Image enhancement technique · Grey level mapping

1 Introduction

Gallstone diseases are one of the most common biliary diseases and there has been a great progress in understanding the gallstones. The historical background of Cholelithiasis has been made for easy classification of Gallstones [1]. According to Japanese classification of Gallstones, the Gallstones are widely of two types: One is the Cholesterol stone, which is further classified as the Pure Cholesterol stone, the Combination stone and the Mixed stone. Second is the Pigment stone, which is further classified as the Black stone and the Calcium Bilirubinate stone. The division line between Cholesterol and the pigment stones depends upon the

N. Mehta (✉) · S. V. A. V. Prasad
Lingaya's University, Faridabad, Haryana, India
e-mail: nehamehta389@gmail.com

S. V. A. V. Prasad
e-mail: prasad.svav@gmail.com

L. Arya
I.T.S Engineering College, Greater Noida, Uttar Pradesh, India
e-mail: leenaarya18@gmail.com

proportion of Cholesterol. If it is equal to or more than 70 % then the stone will be designated as Cholesterol stone; otherwise it is pigment stone with calcium bilirubinate as its principal constituent. The purpose of this study is to investigate the variation of chemical factors of the patients during the pathogenesis of Gallbladder and Kidney Calculus. Moreover the Gallstones can be classified as Primary stones and the Secondary stones [1]. If the stone is viewed in its original site, it is called as Primary stone and if it has migrated from its original site, it is called as Secondary stone. The common manifestation faced to Gallstones includes biliary pain, gallbladder inflammation, pancreatitis, and even bile duct obstruction [2].

The remainder of this paper is organized as follows: In Sect. 2, we have described the Ultrasonography for gallstones. The Sect. 3 is composed of laboratorial Analysis for the Risk Investigation from the dataset that we have chosen for our research background. In Sect. 4, we have stated the Research dataset and methodology. In Sect. 5, the Digital Image Processing technique for Image Enhancement is discussed. In Sect. 6, we have shown the outline of the proposed algorithm. Finally, conclusions and future directions for research are explained in Sect. 7.

2 Ultrasonography for Gallstones

Ultrasonography is a widely used non invasive and non reactive technique for real time imaging, as a current trend in biomedical and industrial applications [3]. The system consists of a beam former part called as transducer, which generates the ultrasonic sound waves at the time of transmission and detects them at the time of reception. However, the typical working frequency of the sound waves depends particularly on the material to be investigated by the ultrasound. These signals are detected by the electronic part of the system, conditioned, acquired, processed and finally displayed in an appropriate format. The Gallstones are viewed as opacities having a distal shadow with high specific gravity [2].

3 Laboratorial Analysis for the Risk Investigation

The laboratorial analysis of Gallstones includes the tests related to Biochemistry, Hematology and Serology. This analysis allows the clinicians to investigate the risk factors associated with the formation of stone and its anatomical risks. The biochemical analysis is important to discover the underlying risk factors of stone formation and preventing its recurrence. The part of hematology analyzes the blood for its related disorders like bleeding disorders such as hemophilia, lymphoma and leukemia. We have categorized the Total Leukocyte Count (TLC) of the patients in order to measure the percentage of each type of white blood cell (WBC) present in the blood. However, the study of Serology helps to identify the antibodies present in the serum.

Table 1 Classification of the hematology record of the test samples

Classification	Approx. % in adults	Normal diameter	Normal range	Mean of the true value of patients	No. of patients in normal range
Total leukocyte count	–	–	4,000–11,000	9,160	7 out of 10
Neutrophils	62	10–12	40–75	74.6	5 out of 10
Lymphocytes	30	12–15	20–45	21.9	6 out of 10
Eosinophils	2.3	10–12	01–06	03.4	9 out of 10
Basophils	0.4	12–15	0–0	0.0	10 out of 10
Monocytes	5.3	12–20	02–10	0.1	0 out of 10

4 Research Dataset and Methodology

Our test data consisted of a group of ten patients who have been selected on random basis and they underwent complete physical examination along with biochemical analysis of blood giving out the results in terms of Hematology, Biochemistry and Serology. Table 1 shows the Hematology record of the patients in which we have calculated the mean TLC and categorized them according to five types of White Blood Cells, namely Neutrophils, Lymphocytes, Monocytes, Eosinophils, and Basophils. Along with the blood analysis, the patients were also undergone with Ultrasonography to detect the anatomical location of the stone. Patient’s ultrasound status, designated as USG Report was also classified according to different anatomical structures and the results are shown in Table 2.

5 Digital Image Processing for Image Enhancement

Image Enhancement is one of the techniques of digital image processing having great significance in the part of medical electronics [4, 5]. The technique of Image enhancement satisfies the aim of improving the interpretation and perception of information present in the images. The techniques of image enhancement are broadly classified as spatial domain methods and frequency domain methods [6]. In spatial domain methods, the processing refers to the plane of the image itself and manipulation of each pixel in an image. The spatial domain operations include spatial filtering of the image and the intensity transformation. In spatial filtering, the operations are performed at the neighborhood level of each pixel. The Spatial filtering operations include sharpening of the image, Noise reduction and enhancement of the edges.

5.1 Grey Level Mapping

We are working on the grey level mapping of the image, which is one of the type of intensity transformation in spatial domain operations. It is applied to get the

Table 2 Classification of the ultrasound record of the test samples

S. N.	Age	Sex	USG Report	Gb	Pancreas	Rt. Kidney	L. Kidney	Ur. bladder
1.	52	F	Normal, 4 mm gb calculus	Bulky with fluid	Normal	Normal	Normal	Normal
2.	60	F	Few calculi largest 23.7 mm	Normal	Normal	Normal	Normal	Normal
3.	36	M	Normal	Normal	Grade II hydronephrosis with 7 mm calculus in right lower ureter	Normal	Normal	Normal
4.	54	F	Absent	Normal	Normal	Normal	Normal	Normal
5.	55	M	Few gb calculi	Normal	Normal	Normal	Normal	Normal
6.	44	M	Normal	Normal	Gr I hydr. and 25 mm calculus in renal pelvis	Cortical calcification with cysts	Normal	Normal
7.	24	M	Normal	Normal	Mild fullness of pc system	Mild fullness of pc system	9.7 mm calculus	Normal
8.	48	F	Calculus of 24.6 mm	Normal	Calculus of 5.9 mm in middle calyx	Calculus, 5.1 mm in upper calyx	Normal	Normal
9.	60	M	Normal	Normal	Few calculi largest of 5.1 mm in middle calyx w/o hydronephrosis	Few calculi largest of size 15.5 mm in lower calyx	Normal	Normal
10.	42	F	Normal	Normal	A calculus of 8 mm in lower calyx	Normal	Normal	Normal

detail resolution of the tissue patterns for easy pattern analysis and recognition. It also helps to differentiate between the speckle and the tissue patterns on the basis of image intensity to indicate the disease effectively. However, the process of point processing is sometimes also called as pixel processing in the grey scale image. It intends to alter the grey levels at the spatial location of the pixels resulting into new grey scale levels. Finally, the grey levels of the image can be compared among the output image grey levels.

5.2 Grey Level Adjustment

Grey level adjustment is done by using thresholding, in which the object pixel is given a value of “1” and a background pixel with “0” [5]. The process makes the use of threshold filter to set the grey scale images into high contrast black and white images in order to differentiate between the areas of interest. The analysis of grey level adjustment can be easily done by using histograms and allows visualizing the pixels of each value and minor components.

6 Proposed Algorithm

Since the images of the ultrasound have typical ranges of luminance therefore it requires some processing for better visualization and understanding. In order to visualize the dynamic range image and to preserve the subjective perception of an image, an algorithm to map the visualized original range values of any image is given as follows.

- Step 1: To compute the brightness factor of the image. The brightness factor will be in terms of the range $[B_{fmin}, B_{fmax}]_{input}$ and $[B_{fmin}, B_{fmax}]_{output}$.
- Step 2: Represent the grey level image as $U [M \times N]$

$$U [M \times N] = \{u(i, j)\}$$

$$i = 1, 2, \dots, M; j = 1, 2, \dots, N$$

- Step 3: Compute the luminance in terms of $L (cd/m^2)$.
- Step 4: Divide the image into two parts—the stone and the rest of the background image, naming them as the two segments of interest.
- Step 5: Set the threshold level without changing the data based pixel values.

Step 6: Implement grey level adjustment method.

Step 7: Compare the grey level histogram of the input image and that of the output image.

Step 8: Finally analyze the stone.

7 Conclusion

The present synthesis provides the comprehensive assessment of the patients having Cholelithiasis, selecting some patients at a random. The association between the gallstones and the Hematology is strong. However, the casual relationship between Biochemistry and Serology factors is under study.

For future work we are working on the Grey Level Mapping technique of image enhancement, for better understanding the complexities of deep grey scale ultrasound images of the anatomical features. It is a commonly used procedure of image enhancement by changing its intensity levels via different levels of mapping and the better results are expected.

References

1. Kim, I.S., Myung, S.J., Lee, S.S., Lee, S.K., Kim, M.H.: Classification and nomenclature of gallstones revisited. *Yonsei Med. J.* **44**(4), 561–570 (2003)
2. Tseng, M., Tseng, M., Everhart, J.E., Sandler, R.S.: Dietary intake and Gallbladder disease: a review. *Public Health Nutr.* **2**(2), 161–172 (1999)
3. Kouame, D., et al.: Ultrasound imaging: signal acquisition, new advanced processing for biomedical and industrial applications. In: ICASSP, vol. 993 (2005)
4. Mittal, A., Dubey, S.K.: Analysis of MRI images of rheumatoid arthritis through morphological image processing techniques. *IJCS* **10**(2–3), 118–122 (2013)
5. Foghi, M.M., et al.: Application of three-dimensional digital image processing for dynamic pore structure characterization. *Glob. J. Sci. Eng. Technol.* (5), 203–208 (2013), ISSN: 2322-2441
6. Jago, J., Billon, A.C., Chenal, C., Jon, J.-M., Ebeid, S.M.: XRES: adaptive enhancement of ultrasound images. *Medicamundi* **46**(3), 36–41 (2002)
7. Gupta, P., Malik, V., Gandhi, M.: Implementation of multilevel threshold method for digital images used in medical image processing. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **2**(2) (2012)
8. Kapoor, V.K., Mc Michael, A. J.: Gallbladder cancer: an ‘Indian’ disease. *Nat. Med. J. India*, **16**(4) (2003)
9. Jagadeesh, S., et al: Image processing based approach to cancer cell prediction in blood samples. *ME&HWDS Int. J. Technol. Eng. Sci.* **1**(1) (2013), ISSN: 2320-8007

Theoretical Study on Vibration of Skew Plate Under Thermal Condition

Anupam Khanna and Pratibha Arora

Abstract The present study represents a computational prediction for the effect of thermal gradient on the vibrations of non-homogeneous four sided clamped skew plate with variable thickness. Authors assumed that temperature varies bi-linearly, density of the plate's material varies linearly in one direction due to non-homogeneity and thickness of plate varies exponentially in one direction. The general equation of motion and consecutive equations are solved by using the Rayleigh–Ritz method. Calculations are made for natural frequencies for first two modes of vibration of a parallelogram plate (a special type of skew plate) at the different combinations of parameters. Results are shown in graphs.

Keywords Skew plate · Thermal gradient · Non-homogeneity constant · Taper constant · Vibration

1 Introduction

The structural components made by different types of plates are used in varied applications involving aerospace, submarine and even civil engineering constructions always with a special emphasis on good quality material that offer the required strength along with the low cost. Therefore, a study of character and behaviour of these plates becomes imperative to ensure optimum utilization of potential of these plates. These plates may be of any geometrical shape:

A. Khanna (✉) · P. Arora
Department of Mathematics, M.M. Engineering College, M.M. University, Mullana,
Haryana, India
e-mail: rajieanupam@gmail.com

P. Arora
e-mail: pratibha360@gmail.com

rectangular, square, triangular, skew plates etc. with constant or variable thickness which may affect the vibrational behaviour of the plates. Skew plates may be of many types for example rhombic plates, parallelogram plates, trapezoidal plates etc. Parallelogram Plates are widely used in modern structures such as in body of aircraft, vehicles bodies etc. In civil engineering, these shapes often observe at entrance to bridges.

A series of early investigation on vibration characteristics of plates were reviewed in Liessa's monograph. In this, Leissa [1, 2] mentioned some papers dealing with square, rectangular, parallelogram and circular etc. in his excellent survey. Chakraverty [3] has studied natural frequencies for free vibration of non-homogeneous elliptic and circular plates using two dimensional orthogonal polynomials. Singh and Sexena [4] studied the transverse vibration of skew plates with variable thickness. Notable contributions are available on vibrations of rectangular plates with linear variable thickness. In [5] Gupta and Khanna studied vibration of visco-elastic rectangular plates with linearly varying thickness. Lal [6] studied the effect of non-homogeneity on the vibration of orthotropic rectangular plates. Huang and Leissa [7] studied vibration of rectangular plates with side cracks via Ritz method. Alijani and Amabili [8] studied non-linear vibration of imperfect rectangular plates with free edges. Wu and Lu [9] studied free vibration analysis of rectangular plates. Quintana [10] presented general Ritz formulation for free vibrational analysis of thick trapezoidal and triangular laminated plates. Again, Quintana and Nallim [11] gave a vibrational approach to vibration analysis of shear deformable polygonal plates with variable thickness. Zhou [12] studied vibration of skew plates by the MLS-Ritz method. Gupta and Kumar [13] studied thermal effect of the vibration of parallelogram plate of bi-linearly varying thickness.

The objective of paper is to present a study on the effect of thermal gradient on non-homogeneous parallelogram plates with exponential thickness variation in x-direction. Rayleigh-Ritz technique is applied to get the numeric values of frequencies (for both the modes of vibration) for different values of structural parameters and results are shown graphically.

2 Analysis and Equation of Motion

The equation of motion of a visco-elastic plate of variable thickness [5] is

$$\begin{aligned} \tilde{D}_1 [D_1 (w_{,xxxx} + 2w_{,xxyy} + w_{,yyyy}) + 2D_{1,x}(w_{,xxx} + w_{,xyy}) + 2D_{1,y}(w_{,yyy} + w_{,yxx}) \\ + D_{1,xx}(w_{,xx} + \nu w_{,yy}) + D_{1,yy}(w_{,yy} + \nu w_{,xx}) + 2(1 - \nu)D_{1,xy}w_{,xy}] + \rho h w_{,tt} = 0 \end{aligned} \quad (1)$$

Here ν is poisson ratio, \tilde{D}_1 is viscoelastic operator and a comma in the suffix denotes partial differentiation of w with respect to suffix variable.

Also, D_1 is the flexural rigidity of the plate's material defined as

$$D_1 = \frac{Eh^3}{12(1 - \nu^2)} \tag{2}$$

By using variable separation technique, the solution of Eq. (1) can be taken in the form of product of two functions as

$$w(X, Y, t) = W(X, Y)T(t) \tag{3}$$

Substituting Eq. (3) into Eq. (1), one obtains

$$[D_1(W_{,XXXX} + 2W_{,XXYY} + W_{,YYYY}) + 2D_{1,x}(W_{,XXX} + W_{,XY}) + 2D_{1,y}(W_{,YY} + W_{,YXX}) + D_{1,xx}(W_{,XX} + \nu W_{,YY}) + D_{1,yy}(W_{,YY} + \nu W_{,XX}) + 2(1 - \nu)D_{1,xy}W_{,XY}] / \rho h W = -\ddot{T} / \tilde{D}T \tag{4}$$

Here, (.) dot denotes differentiation with respect to t.

The preceding equation is satisfied if both of its sides are equal to a constant say p^2 , one gets:

$$[D_1(W_{,XXXX} + 2W_{,XXYY} + W_{,YYYY}) + 2D_{1,x}(W_{,XXX} + W_{,XY}) + 2D_{1,y}(W_{,YY} + W_{,YXX}) + D_{1,xx}(W_{,XX} + \nu W_{,YY}) + D_{1,yy}(W_{,YY} + \nu W_{,XX}) + 2(1 - \nu)D_{1,xy}W_{,XY}] - \rho p^2 h W = 0 \tag{5}$$

and

$$\ddot{T} + p^2 \tilde{D}T = 0 \tag{6}$$

Equations (5) and (6) are the differential equation of motion and time function for visco-elastic plate respectively.

The expressions for strain energy and kinetic energy are given as [5]

$$\bar{T} = \frac{1}{2} \rho p^2 \iint h W^2 dY dX \tag{7}$$

and

$$\bar{V} = \frac{1}{2} \iint \frac{Eh^3}{12(1 - \nu^2)} \{ (W_{,XX})^2 + (W_{,YY})^2 + 2\nu W_{,XX} W_{,YY} + 2(1 - \nu)W_{,XY^2} \} dY dX \tag{8}$$

A parallelogram plate R (lateral sides a and b) with skew angle (θ) is shown in Fig. 1. The parallelogram plate is assumed to be non-uniform, thin and isotropic.

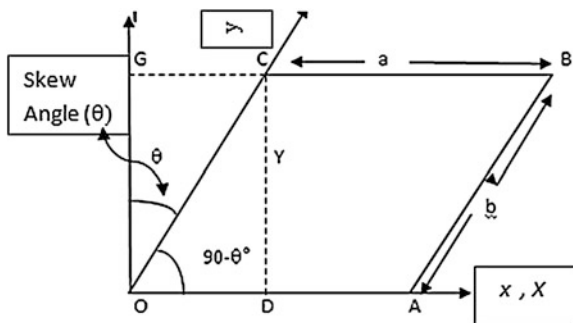
On using the concept of skew coordinates [1], one gets

$$X = x + y \sin \theta \quad \text{and} \quad Y = y \cos \theta \tag{9}$$

The boundaries of the plate in skew coordinates are

$$x = 0, x = a \quad \text{and} \quad y = 0, y = b \tag{10}$$

Fig. 1 A Parallelogram plate of lateral sides a and b



On using Eqs. (2) and (9) in Eqs. (7) and (8), the maximum kinetic energy \bar{T} and strain energy \bar{V} in the plate when it is executing transverse vibration mode shape $W(x, y)$ are obtained as follows:

$$\bar{T} = \frac{1}{2} \rho p^2 \cos \theta \int_0^b \int_0^a h W^2 dx dy \tag{11}$$

and

$$\begin{aligned} \bar{V} = \frac{1}{2 \cos^3 \theta} \int_0^b \int_0^a \frac{Eh^3}{12(1-\nu^2)} & (W_{,xx^2} - 4 \sin \theta W_{,xx} W_{,xy} + 2(\sin^2 \theta + \nu \cos^2 \theta) W_{,xx} W_{,yy} \\ & + 2(1 + \sin^2 \theta - \nu \cos^2 \theta) W_{,xy^2} - 4 \sin \theta W_{,xy} W_{,yy} + W_{,yy^2}) dx dy \end{aligned} \tag{12}$$

Authors assumed that the plate is subject to study of bi-linear temperature variations i.e. x -direction and y -direction as

$$\tau = \tau_0 \left(1 - \frac{x}{a}\right) \left(1 - \frac{y}{b}\right) \tag{13}$$

where τ denotes the temperature excess above the reference temperature at any point on the plate and τ_0 denotes the temperature excess above the reference temperature on the boundary of plate.

The modulus of elasticity which depends upon temperature τ is given by

$$E(\tau) = E_0(1 - \gamma\tau) \tag{14}$$

where E_0 is the value of Young modulus at the reference temperature i.e. at $\tau = 0$ with the temperature at the end of the plate as reference

Using Eq. (13) in Eq. (14) one obtains

$$E(\tau) = E_0 \left(1 - \alpha \left(1 - \frac{x}{a}\right) \left(1 - \frac{y}{b}\right)\right) \tag{15}$$

where $\alpha = \gamma\tau_0$ ($0 \leq \alpha < 1$) is temperature parameter, known as thermal gradient.

The thickness variation of the parallelogram plate is assumed to be exponentially in x-direction as:

$$h = h_0 e^{\beta x} \tag{16}$$

where β is the taper constant in x-direction and $h = h_0$ when $x = 0$

Since the plate material is non-homogeneous, authors considered that density of plate varies in x-direction i.e.

$$\rho = \rho_0 \left(1 - \alpha_1 \left(\frac{x}{a} \right) \right) \tag{17}$$

where ρ_0 is density at the center of plate and α_1 is non-homogeneity constant.

Using Eqs. (15), (16), (17) in Eqs. (11) and (12), one gets

$$\bar{T} = \frac{1}{2} h_0 \rho_0 P^2 \int_{y=0}^b \int_{x=0}^a e^{\beta x} \left(1 - \alpha_1 \left(\frac{x}{a} \right) \right) W^2 dx dy \tag{18}$$

and

$$\begin{aligned} \bar{V} = & \frac{E_0 h_0^3}{24(1 - \nu^2) \cos^4 \theta} \int_{y=0}^b \int_{x=0}^a \left(1 - \alpha \left(1 - \frac{x}{a} \right) \left(1 - \frac{y}{b} \right) \right) (e^{\beta x})^3 \left[W_{,xx^2} - 4 \left(\frac{a}{b} \right) \sin \theta W_{,xx} W_{,xy} \right. \\ & + 2 \left(\frac{a}{b} \right) (\sin^2 \theta + \nu \cos^2 \theta) W_{,xx} W_{,yy} + 2 \left(\frac{a}{b} \right) (1 + \sin^2 \theta - \nu \cos^2 \theta) W_{,xy^2} \\ & \left. - 4 \left(\frac{a}{b} \right)^3 \sin \theta W_{,xy} W_{,yy} + \left(\frac{a}{b} \right)^4 W_{,yy^2} \right] dx dy \end{aligned} \tag{19}$$

3 Solution and Frequency Equation

In Rayleigh–Ritz technique, one of the requirements is that the maximum strain energy must be equal to maximum kinetic energy. Hence, for arbitrary variations of W satisfying geometric boundary conditions, it is observed that

$$\delta(\bar{V} - \bar{T}) = 0 \tag{20}$$

Using Eqs. (18) and (19) in Eq. (20) yields

$$\delta(V_1 - \lambda^2 T_1) = 0 \tag{21}$$

where

$$\begin{aligned}
 V_1 = & \frac{1}{\cos^4 \theta} \int_{y=0}^b \int_{x=0}^a \left(1 - \alpha \left(1 - \frac{x}{a}\right) \left(1 - \frac{y}{b}\right)\right) (e^{\beta x})^3 \left[W_{,xx^2} - 4 \left(\frac{a}{b}\right) \sin \theta W_{,xx} W_{,xy} \right. \\
 & + 2 \left(\frac{a}{b}\right) (\sin^2 \theta + v \cos^2 \theta) W_{,xx} W_{,yy} + 2 \left(\frac{a}{b}\right) (1 + \sin^2 \theta - v \cos^2 \theta) W_{,xy^2} \\
 & \left. - 4 \left(\frac{a}{b}\right)^3 \sin \theta W_{,xy} W_{,yy} + \left(\frac{a}{b}\right)^4 W_{,yy^2} \right] dx dy \tag{22}
 \end{aligned}$$

$$T_1 = \int_{y=0}^b \int_{x=0}^a e^{\beta x} \left(1 - \alpha_1 \left(\frac{x}{a}\right)\right) W^2 dx dy \tag{23}$$

$$\text{and } \lambda^2 = \frac{12a^4 p^2 \rho_0 (1 - v^2)}{E_0 h_0^2} \text{ is frequency parameter.} \tag{24}$$

For a parallelogram plate which is clamped along all the four edges, the boundary conditions are

$$W = W_{,x} = 0 \text{ at } x = 0, a \text{ and } W = W_{,y} = 0 \text{ at } y = 0, b \tag{25}$$

Corresponding two term deflection function which satisfy above boundary conditions is taken as [2]

$$W(x, y) = \left(\frac{x^2}{a^2}\right) \left(\frac{y^2}{b^2}\right) \left(1 - \frac{x}{a}\right)^2 \left(1 - \frac{y}{b}\right)^2 \times \left[A_1 + A_2 \left(\frac{x}{a}\right) \left(\frac{y}{b}\right) \left(1 - \frac{x}{a}\right) \left(1 - \frac{y}{b}\right) \right] \tag{26}$$

Substituting the value of W from Eq. (24) in Eq. (21), an equation involving A₁ and A₂ is obtained whose values is to be determined using Eq. (21) for which

$$\frac{\partial}{\partial A_n} (V_1 - \lambda^2 T_1) = 0, n = 1, 2 \tag{27}$$

The above equation simplifies as

$$b_{n1} A_1 + b_{n2} A_2 = 0, n = 1, 2 \tag{28}$$

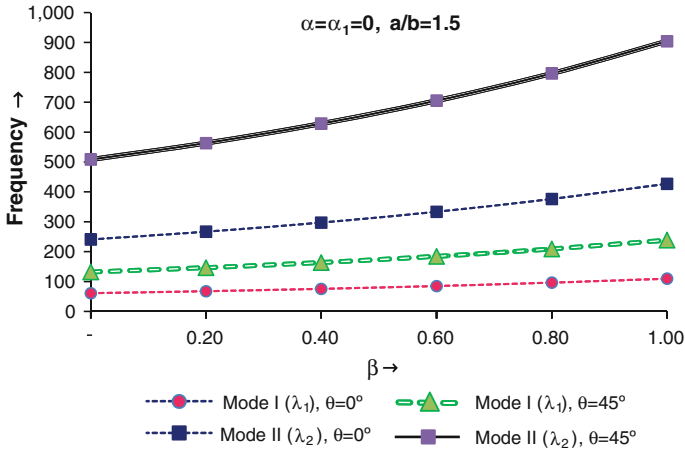
where b_{n1}, b_{n2} (n = 1, 2) involve parametric constants and frequency parameter.

For a non trivial solution the determinant of the coefficients of Eq. (28) must be zero. Therefore one gets the frequency equation as

$$\begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} = 0 \tag{29}$$

From Eq. (29), one can obtain the quadratic equation in λ² from which two values of λ² can be found.

The frequency parameter (λ) corresponding to first two modes of vibration of clamped parallelogram plate have been computed for various values of temperature gradient (α), non-homogeneity (α₁), skew angle (θ), aspect ratio (a/b) and



Graph 1 Frequency versus taper constant

taper constant (β). The value of poisson ratio (ν) is taken as constant i.e. 0.345. All results are presented in graphs.

4 Results and Discussion

Graph 1: It represents the values of frequency for both the modes of vibration with increasing values of taper constant (β) from 0.0 to 1.0 for the following cases:

- Case (i): $\alpha = \alpha_1 = 0.0, a/b = 1.5, \theta = 0^\circ$
- Case (ii): $\alpha = \alpha_1 = 0.2, a/b = 1.5, \theta = 45^\circ$

It is easily inspected that frequency increases continuously as taper constant increases from 0.0 to 1.0. It is also observed as angle increases from 0° to 45° , frequency also increases.

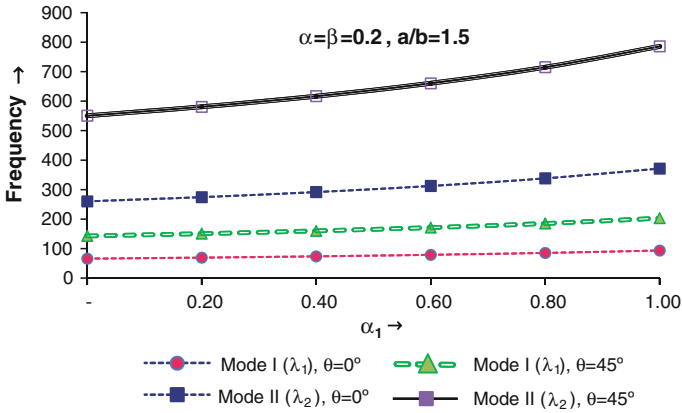
Graph 2: In this graph, variation of frequencies (for both modes of vibration) with respect to non-homogeneity is shown for following cases:

- Case (i): $\alpha = \beta = 0.2, a/b = 1.5, \theta = 0^\circ$ (Mode I and Mode II)
- Case (i): $\alpha = \beta = 0.2, a/b = 1.5, \theta = 45^\circ$ (Mode I and Mode II)

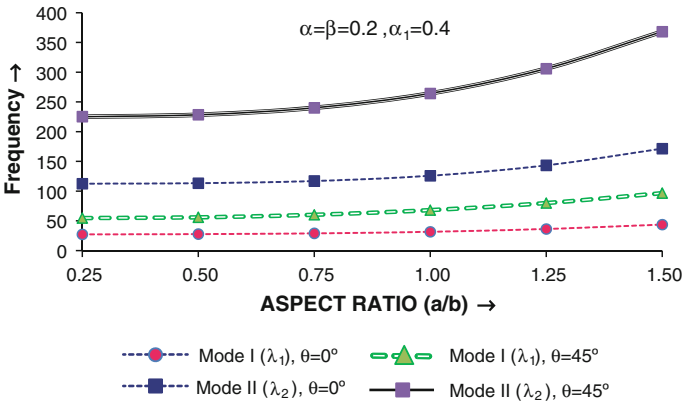
Here, frequency increases as the values of non-homogeneity constant increases from 0.0 to 1.0. Also it is interesting to note that as θ increases from 0° to 45° , frequency also increases for both the modes of vibration.

Graph 3: It represents variation of frequency with respect to aspect ratio under the following considerations:

- Case (i): $\alpha = \beta = 0.2, \alpha_1 = 0.4, \theta = 0^\circ$ (Mode I and Mode II)



Graph 2 Frequency versus non homogeneity



Graph 3 Frequency versus aspect ratio

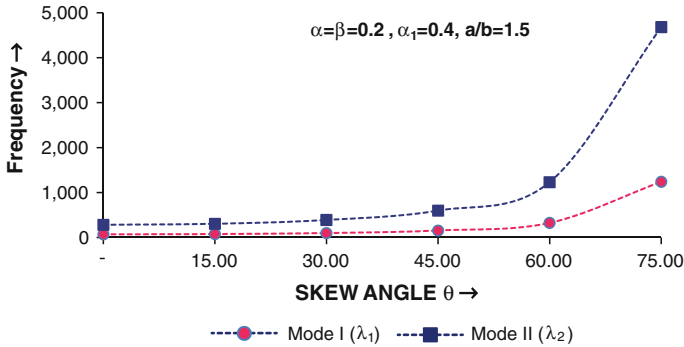
Case (ii): $\alpha = \beta = 0.2, \alpha_1 = 0.4, \theta = 45^\circ$ (Mode I and Mode II)

It is observed that frequency (for both the modes of vibration) continuously increases for increasing values of aspect ratio from 0.25 to 1.5 and θ s from 0° to 45° .

Graph 4: This shows the variation in frequency (for both the modes of vibration) with respect to increasing values of aspect ratio for the following case:

$$\alpha = \beta = 0.2, \alpha_1 = 0.4, a/b = 1.5 \text{ (Mode I and Mode II)}$$

It is obvious to note that frequency (for both the modes of vibration) increases slowly but continuously as skew angle increases from 0° to 60° but an immediate jump is found as skew angle changes from 60° to 75° .



Graph 4 Frequency versus skew angle

Table 1 Frequency versus taper constant

β	$\alpha = \alpha_1 = 0, \theta = 30^\circ$		$\alpha = 0.2, \alpha_1 = 0, \theta = 30^\circ$		$\alpha = 0.0, \alpha_1 = 0.2, \theta = 30^\circ$	
	Mode I	Mode II	Mode I	Mode II	Mode I	Mode II
0.00	84.61 _{84.61}	329.84 _{329.84}	82.53 _{82.53}	321.79 _{321.79}	89.19 _{89.19}	347.68 _{347.68}
0.20	93.86 _{96.01}	365.65 _{375.28}	91.66 _{93.75}	357.07 _{366.44}	98.99 _{101.26}	385.83 _{395.96}
0.40	104.90 _{107.77}	407.84 _{421.96}	102.56 _{105.31}	398.66 _{412.26}	110.69 _{113.70}	430.79 _{445.54}
0.60	118.08 _{119.77}	457.66 _{496.46}	115.58 _{117.11}	447.82 _{458.89}	124.65 _{126.39}	483.92 _{496.01}
0.80	133.80 _{131.93}	516.63 _{517.56}	131.11 _{129.06}	506.05 _{506.10}	141.32 _{139.27}	546.85 _{547.11}
1.00	152.54 _{144.22}	586.57 _{566.09}	149.63 _{141.14}	575.16 _{553.73}	161.18 _{152.27}	621.52 _{598.67}

Note Frequency in subscript is for [14]

5 Comparison and Conclusion

Authors compare the frequencies of present paper with [14] at the corresponding values of varying parameter and results are given in tabular form as follows:

From Table 1, it is evident that values of frequencies in present paper are slightly less than [14] for corresponding values of varying parameters for both the modes of vibration. Hence, authors conclude that frequency (for both the modes) decreases in case of exponential varying thickness (present paper) as compared to sinusoidal varying thickness [14]. Therefore, authors provide a mathematical model for scientists and engineers so that they can achieve more authentic and variable mechanical structures or designs.

References

1. Leissa, A.W.: Vibration of plates. NASA SP-160 (1969)
2. Leissa, A.W.: Recent studies in plate vibration 1981–1985, part-II complicating effect. Shock Vib. Dig. **19**, 10–24 (1987)

3. Chakraverty, S., Petyt, M.: Natural frequencies for free vibration of non-homogeneous elliptic and circular plate using two dimensional orthogonal polynomials. *Appl. Math. Modell.* **21**, 399–417 (1997)
4. Singh, B., Sexena, V.: Transverse vibration of skew plates with variable thickness. *J. Sound Vib.* **206**(1), 1–13 (1997)
5. Gupta, A.K., Khanna, A.: Vibration of visco-elastic rectangular plate with linearly varying thickness vibration in both directions. *J. Sound Vib.* **301**(3–5), 450–457 (2007)
6. Lal, R., Dhanpati: Effect of non-homogeneity on the vibration of orthotropic rectangular plates of varying thickness resting on a Pasternak foundation *J. Vib. Acoust.* **131**, 1–9 (2009)
7. Huang, C.S., Leissa, A.W.: Vibration analysis of rectangular plates with side cracks via the Ritz method. *J. Sound Vib.* **323**(3–5), 974–988 (2009)
8. Alijani, F., Amabili, M.: Theory and experiments for nonlinear vibrations of imperfect rectangular plates with free edges. *J. Sound Vib.* **332**, 3564–3588 (2013)
9. Wu, L.H., Lu, Y.: Free vibration analysis of rectangular plates with internal columns and uniform elastic edge supports by pb-2 Ritz method. *Int. J. Mech. Sci.* **53**, 494–504 (2011)
10. Quintana, M.V., Nallim L.G.: a general Ritz formulation for the free vibration analysis of thick trapezoidal and triangular laminated plates resting on elastic supports. *Int. J. Mech. Sci.* **69**, 1–9
11. Quintana, M.V., Nallim, L.G.: A variational approach to free vibration analysis of shear deformable polygonal plates with variable thickness. *Appl. Acoust.* **71**, 393–401 (2010)
12. Zhou, L., Zheng, W.X.: Vibration of skew plates by the MLS-Ritz method. *Int. J. Mech. Sci.* **50**, 1133–1141 (2008)
13. Gupta, A.K., Kumar, M.: Thermal effect of vibration of a parallelogram plate of bi-direction linearly varying thickness. *Appl. Math* **2**, 33–38 (2011)
14. Khanna, A., Arora, P.: Effect of sinusoidal thickness variation on vibrations of non-homogeneous parallelogram plate with bi-linearly temperature variations. *Indian J. Sci. Technol.* **6**(9) (2013)

Efficient Approach for Reconstruction of Convex Binary Images Branch and Bound Method

Shiv Kumar Verma, Tanuja Shrivastava and Divyesh Patel

Abstract In this paper reconstruction algorithm of convex binary image in discrete tomography made efficient by implementing branch and bound method. We focus on diagonal and anti-diagonal (dad) projections and comparison done with the conventional horizontal and vertical (hv) projections. It was shown that proposed strategy is computationally strong and gives fast reconstruction.

Keywords Discrete tomography · Convexity · Branch and bound · Anti-diagonal and diagonal projections

1 Introduction

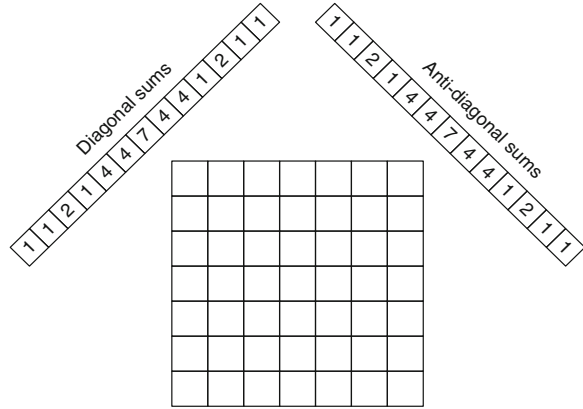
Reconstruction of binary matrices from few projections is the field of study in discrete tomography. We can provide information about any unknown binary images from the projections in few directions and some prior information about the object such as convexity and connectivity may give close approximate reconstruction result of unknown objects [1–3]. Reconstruction of binary matrix were based on horizontal and vertical projections and using branch and bound method it was first implemented by Miklós and Csongor in [4] and discussed that constructing of tree exponential grow if the size of matrix increase but work well for matrix of order less than 10. In present approach considering different view of binary images,

S. K. Verma (✉) · T. Shrivastava · D. Patel
Department of Mathematics, Indian Institute of Technology-Roorkee,
Roorkee, India
e-mail: shivkdma@gmail.com

T. Shrivastava
e-mail: tanujfma@iitr.ernet.in

D. Patel
e-mail: divyshpatel@gmail.com

Fig. 1 Binary matrix and projection set in anti-diagonal and diagonal directions



restricting to only two orthogonal projections anti-diagonally and diagonally (Detail description given for anti-diagonal and diagonal characteristics can be find in [5]) shown in Fig. 1. Let $\mathcal{P}(D, A)$ is the given projection set such that,

$$A = \{ 1 \ 1 \ 2 \ 1 \ 4 \ 4 \ 7 \ 4 \ 4 \ 1 \ 2 \ 1 \ 1 \}$$

$$D = \{ 1 \ 1 \ 2 \ 1 \ 4 \ 4 \ 7 \ 4 \ 4 \ 1 \ 2 \ 1 \ 1 \}$$

where A represents the anti-diagonal vector sum, and D represents the diagonal vector sum. If solution exists, then we can reconstruct the binary matrix from the given two projection vectors shown in Fig. 1. The condition for existence of the solution given in [5] for an $m \times n$ binary matrix, where a_k and d_k are number of 1's in anti-diagonal and diagonal projection are as follows,

- $\sum_{k=1}^{m+n-1} a_k = \sum_{k=1}^{m+n-1} d_k$ and
- $\sum_{k=(2l-1)}^{m+n-1} a_k = \sum_{k=(2l-1)}^{m+n-1} d_k$ and $\sum_{k=2l}^{m+n-1} a_k = \sum_{k=2l}^{m+n-1} d_k$:

for $l = 1, \dots, m + n - 1$.

If suppose there exist exactly single 1 in every row (column), then the possibility to place single one in any row (column) can be perform by $n!$ number of ways and for simplicity if the matrix is square then there will be $n! \times n!$ ways to place a one in binary matrix. In our approach, if the matrix is viewed diagonally and anti-diagonally then placing of 1's accordingly in anti-diagonal (diagonals) sums, since diagonal and anti-diagonal are not intersect each other as in case of horizontal and vertically, thus we can separate those intersecting diagonals with anti-diagonal and observed that odd and even indexing of anti-diagonals (diagonal), then only odd index diagonals (anti-diagonal) will have a single one while the even diagonals (anti-diagonal) will have no 1's. Thus, the total possibility to place single 1's in every row (Column) will be $n!$ which is nothing but the main anti-diagonal (diagonal) and can be considered as best case for anti-diagonal (diagonal). Hence, the total number possibility having exactly single one's in every anti-diagonal (diagonal) will have $n!((n - 1)!)$.

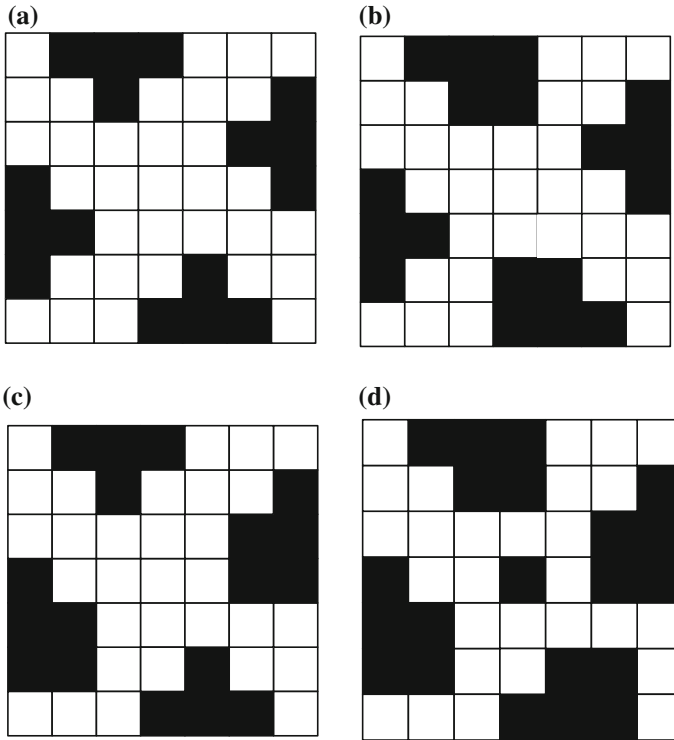


Fig. 2 a Dad-convex image, b a-convex image, c d-convex image, and d non-convex image

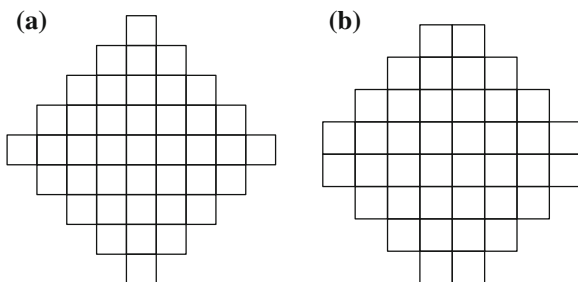
1.1 Dad-Convex Matrix

A binary matrix with respect to anti-diagonal and diagonal projection will be called as *ad-convex* if there are no 0's between the sequence of 1's in all the anti-diagonal sums as shown in Fig. 2b, similarly *d-convex* if there are no 0's between the sequence 1's in every diagonals as shown in Fig. 2c. A binary matrix will be dad-convex if it is a-convex as well as *d-convex* as shown in Fig. 2a otherwise non-convex shown in Fig. 2d.

1.2 Branch and Bound Method

The *Branch and Bound* is a convex optimization method for global optimization, and is not always fast (indeed, are often slow). The method is based on the observation that the enumeration of integer solutions has a tree structure. The main idea in branch and bound is to restrict unbound growth of tree by calculating

Fig. 3 Building of two tree for anti-diagonal and diagonal projection **a** odd-odd indexing, and **b** even-even indexing



bound at every branching point and at intermediate node point the current version of the branch and bound tree will be available and consists of nodes labeled with their bounding values.

The *node selection policy* governs how to choose the next node for expansion. There are three popular and basic policies for node selection:

- *Best-first* or *global-best node selection*: in this process best bounding value of node is selected for maximizing while for minimizing select the lowest bounding value of nodes.
- *Depth-first*: selection from the current set of nodes i.e. first anti-diagonal sums and go to one step deeper into the branch and bound tree after each iteration by checking the corresponding diagonal sums. Hence, it reaches the last nodes quickly. If it cannot proceed to any deeper into the tree, back track one level and choose another child node from that level.
- *Breadth-first*: expanding nodes in the same order in which they were created.

2 Method of Building Tree

Building tree is done by two separate odd index Fig. 3a binary matrix and even index Fig. 3b binary matrix. For simplicity, square odd order matrix considered.

The development of tree is started with anti-diagonal line sum and the corresponding intersecting diagonal line sums, complying with the *dad-convexity*. If any of the corresponding diagonal sum and *dad-convexity* properties are not conflicting, then that sub tree will be expanded further, up to leaves of the tree or till diagonal sums and *dad-convexity* matches else the sub tree will be cut out narrowing the tree. All possible solutions will be obtained upon reaching to the leaves of the tree. The leaves node of the tree will represent the number of solutions.

3 Reconstruction Strategy

The basic idea for constructing the tree will be started with the very first contents of the anti-diagonal sums of a binary matrix. If at any point tree construction is not possible that some of the diagonal sums or the dad-convexity condition may not satisfy, hence these sub trees will be cut out, narrowing the solution. Whenever we reached to the leaf node we get a solution.

Consider a 7×7 binary matrix with anti-diagonal sums, red and black color indicating the intersecting diagonal with the corresponding anti-diagonal. Let the anti-diagonal sums is given by

$$A = [1 \ 1 \ 2 \ 1 \ 4 \ 4 \ 7 \ 4 \ 4 \ 1 \ 2 \ 1 \ 1]$$

and diagonal sums

$$D = [1 \ 1 \ 2 \ 1 \ 4 \ 4 \ 7 \ 4 \ 4 \ 1 \ 2 \ 1 \ 1]$$

Separate all anti-diagonal and diagonal sums into even and odd indexing as follows

$$A_{odd} = [1 \ 2 \ 4 \ 7 \ 4 \ 2 \ 1], \quad A_{even} = [1 \ 1 \ 4 \ 4 \ 1 \ 1],$$

Similarly,

$$D_{odd} = [1 \ 2 \ 4 \ 7 \ 4 \ 2 \ 1], \quad D_{even} = [1 \ 1 \ 4 \ 4 \ 1 \ 1],$$

Now check for consistency or existence of the solutions

$$\begin{aligned} \sum A &= 33 = \sum D \text{ also } \sum A_{odd} = 21 = \sum D_{odd} \text{ and } \sum A_{even} = 12 \\ &= \sum D_{even} \end{aligned}$$

Starting with the first anti-diagonal sums $A_{odd}(1) = 1$ and placed in top most block of odd-odd index of Fig. 3a, next for $A_{odd}(2) = 2$ and the available positions are 3 so we can place two 1's in three places by $(P_n - k + 1)$ way for the convex sums, while for the non convex there will be $C_k^{P_n}$ possibility to place 1's where $(P_n = \{1, 2, \dots, n - 1, n, n - 1, \dots, 2, 1\})$ is a set containing the number of pixel in diagonally or anti-diagonally. The tree will be expanded for 110 and 011. The corresponding diagonal sums matches up to $A_{odd}(4) = 7$, when $A_{odd}(5) = 4$, the diagonal sums became 1345311 which is out of bound to the diagonal sums 1247421, hence tree will not expanded from this node and this node will be pruned. In the last $A_{odd}(7) = 1$, we get four possible matches with the corresponding sums, hence the four different solution as shown in Fig. 4. In the similar manner we will build the tree for the even anti-diagonal and intersecting diagonal sums showing in Fig. 5a and 5b.

In all Figs. 4, 5a and 5b, red box indicating not matching with the diagonal sum, while green box matches with the diagonal sum and represent the possible solution. Thus, we get six possible solution of the given projection set.

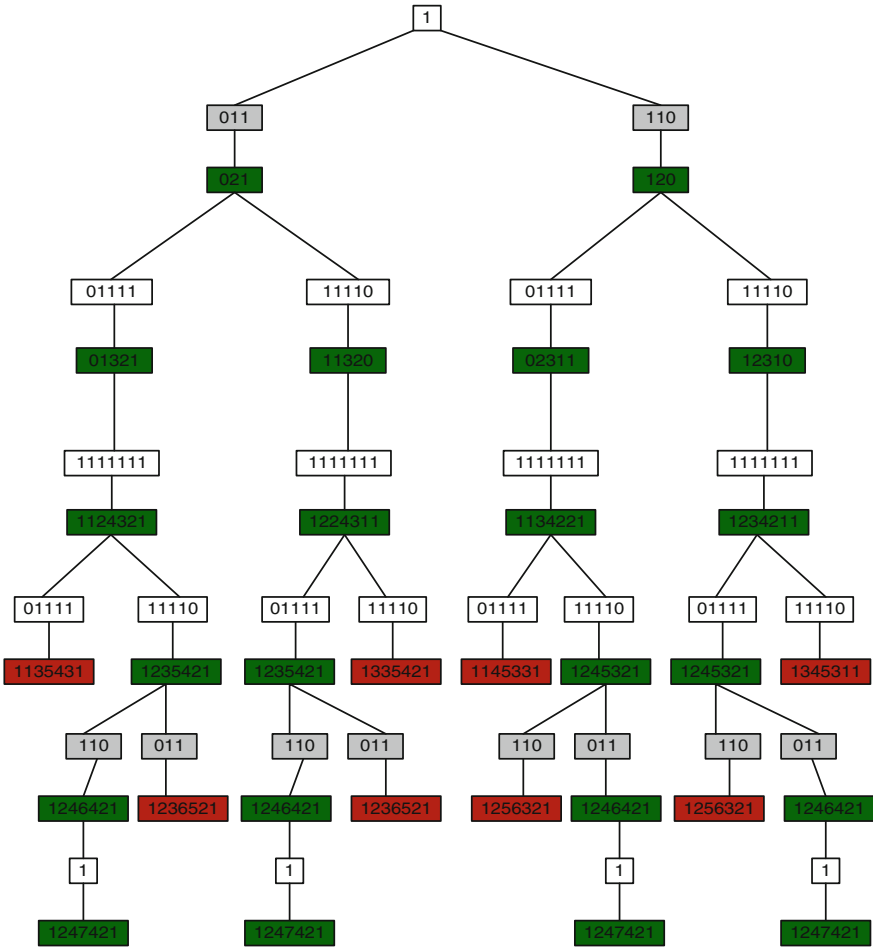


Fig. 4 First solution odd–odd indexed of anti-diagonal and diagonal sums

4 Computational Results

4.1 Unique Solution

If the anti-diagonal and diagonal vector sums are consistent then it is possible to reconstruct unique binary image [6] and from the experiments it was verified that, branch and bound method give fast solution. If we consider the following anti-diagonal and diagonal vector sum

$$A = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 2 & 3 & 2 & 5 & 1 & 0 & 1 \\ & 1 & 1 & 0 & 5 & 2 & 3 & 2 & 1 & 2 & 1 & 2 & 1 \end{bmatrix}$$

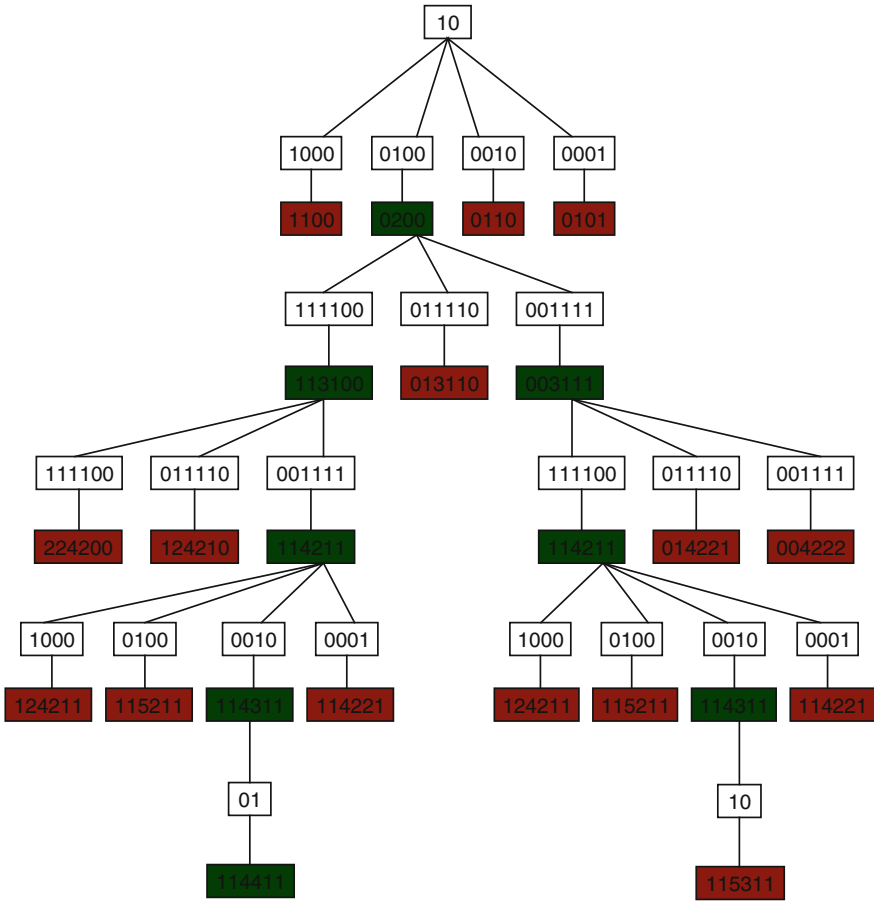


Fig. 5a Second solution for even index of anti-diagonal and diagonal

$$D = \begin{bmatrix} 1 & 2 & 1 & 2 & 1 & 2 & 3 & 2 & 5 & 1 & 0 & 1 \\ 1 & 1 & 0 & 5 & 2 & 3 & 2 & 1 & 2 & 1 & 2 & 1 \end{bmatrix}$$

We get the following unique solution shown in Fig. 6.

4.2 Other Solutions

There are possibilities to get many solutions of same given projections. It is due to presence of switching elements presents and we follow the dad-convexity condition then we get many valid solution of the same anti-diagonal and diagonal vector

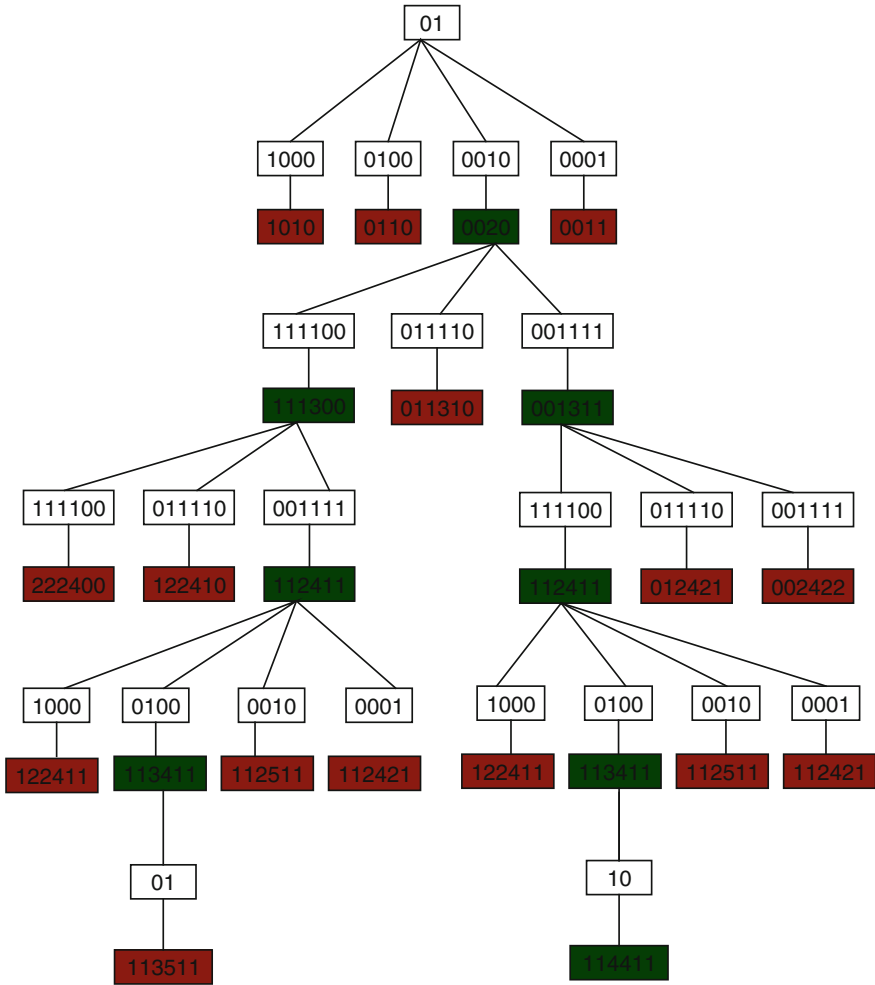


Fig. 5b Third solution for even indexing of diagonals and anti-diagonals

sums as shown in Fig. 7a–d, if we are not following the convexity condition then we get more solution from Sect. 3, shown in Fig. 7e–i.

4.3 Comparison with Horizontal and Vertical Projections

Experiment done on various synthetic binary images and compared our computational result with the *hv*-convex images [4]. It was observed that execution time and number of nodes increases exponentially in the tree in case of *hv*-convex images, while in case of *dad*-convex, execution time is drastically decreases, although the height of tree is more as compare to *hv*-convex images.

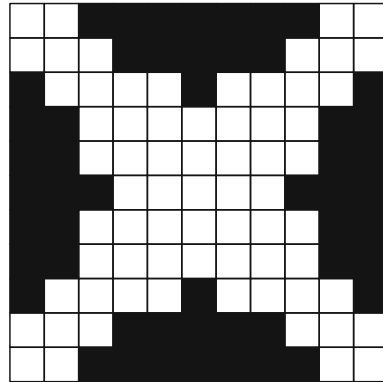


Fig. 6 Unique reconstruction from the given anti-diagonal and diagonal sums

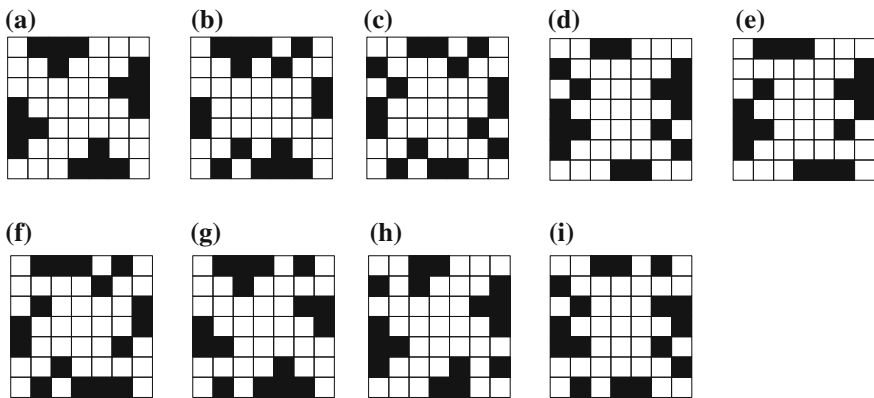


Fig. 7 Other possible solution of **a** due to switching, **b–d** with *dad*-convex and from **e–i** without *dad*-convexity

Consider an example, the row and column sum of a binary image as $R = [2 \ 4 \ 4 \ 4 \ 2]$, $C = [2 \ 4 \ 4 \ 4 \ 2]$, then we get four different solution with an average of 38 nodes of all in Fig. 8.

Since the above binary images are *hv*-convex as well as *dad*-convex, if we reconstruct using our approach, unique reconstruction is possible for all images as shown in Fig. 8, an in the tree 8 nodes for Fig. 8a, 7 nodes for Fig. 8b, 6 nodes for Fig. 8c and 7 nodes for Fig. 8d respectively are appeared.

For comparison purpose the proposed strategy and method given in [4]. We have created a database of more than 200 images of various sizes of 5×5 to 30×30 , having different ratio of 1's against 0's in every images and varies from 30, 60 and 80 %. Programs were developed in MATLAB 2009, and executed on AMD (Phenom) II Quad Core with 4 GB RAM.

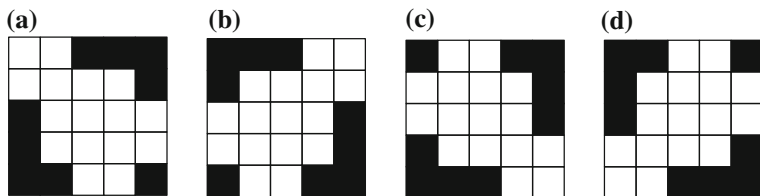


Fig. 8 Four solutions for the given $R = [2 \ 4 \ 4 \ 4 \ 2]$ and $C = [2 \ 4 \ 4 \ 4 \ 2]$ sums

Table 1 Comparison of 7×7 binary images

I's (%)	Avg. time (s)		Avg. solution		Avg. nodes	
	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>
30	1.0205	0.4644	1.5	2	84.5	13.84
60	1.0917	0.5023	1.3	1.3	109	11.33
80	0.5416	0.5108	1.3	1.7	61	10

Table 2 Comparison of 11×11 binary images

I's (%)	Avg. time (s)		Solution		Nodes	
	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>
30	24.1593	1.6065	1.6	3	755.4	32
60	22.36	1.3198	1.4	1.2	1,377	58
80	2.3462	0.7129	1.2	1.8	284	30

Table 3 Comparison of 15×15 binary images

I's (%)	Avg. time (s)		Solution		Nodes	
	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>
30	146.32	12.788	2	10	3,360	148.5
60	665.473	8.0052	1	1	18,561	176
80	124.6752	6.0142	4	1	15,632	128

Table 4 Comparison of 21×21 binary images

I's (%)	Avg. time (s)		Solution		Nodes	
	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>	<i>hv</i>	<i>dad</i>
30	245.32	42.365	4	1	42,532	253
60	186.423	14.436	1	1	36,452	187
80	145.471	5.146	1	1	25,732	145

Tables 1, 2, 3, 4, and 5, represents comparison, on the basis of the execution time of *hv*-convex and *dad*-convex, number of node exist in the tree and the number of possible solution exists for the binary image consistent data.

5 Conclusions

Result obtained and displayed in Tables 1, 2, 3, and 4, are clearly indicating that the proposed method is better in execution time, number of nodes presented in tree as compare to horizontal and vertical approach given in [4].

We have implemented our approach for the reconstruction of binary images through branch and bound method using only two orthogonal projections in different directions. Although the branch and bound method is comparatively slow converges rate comparing to other optimization methods. For large size of the binary images it is not feasible for *hv*-convex method of solutions, while it works well for *dad*-convex up to 50×50 order of matrices.

This reconstruction method can be applied for in crystalline tomography, data hiding and security by using the coding and decoding of the transmission data for which further research is in progress.

Acknowledgments This research was supported by the MHRD (Ministry of Human Resource Development) INDIA under the QIP (Quality Improvement Program) sponsored by AICTE (All Indian Council for Technical Education) and Subhash Institute of Software Technology Kanpur.

References

1. Chrobak, M., Durr, C.: Reconstructing *hv*-convex polyominoes from orthogonal projection. *Inf. Process. Lett.* **69**(6) (1999)
2. Del Lungo, A.: Reconstructing permutation matrices from diagonal sums. *Theor. Comput. Sci.* **281**(1–2), pp. 235–249 (2002)
3. Kuba, A.: Reconstruction of two directionally connected binary patterns from their orthogonal projections. *Comput. Vis. Graph. Image Process.* **27**, 249–265 (1984)
4. Miklós, P., Csongor, G.: Discrete tomographic reconstruction of binary matrices using branch and bound method. In: 7th International Symposium on Intelligent Systems and Informatics (SISY 2009), Subotica, Serbia, 25–26 Sept 2009
5. Srivastava, T., Verma, S.K., Patel, D.: Reconstruction of binary images from two orthogonal projections. *IJTS* **21**(2), 105–114 (2012)
6. Srivastava, T., Verma, S.K.: Uniqueness algorithm with diagonal and anti-diagonal projections. *IJTS* **23**(2), 22–31 (2013)

Construction of Enhanced Sentiment Sensitive Thesaurus for Cross Domain Sentiment Classification Using Wiktionary

P. Sanju and T. T. Mirnalinee

Abstract Sentiment classification is classification of reviews into positive or negative depends on the sentiment words expressed in reviews. Automatic sentiment classification is necessary in various applications such as market analysis, opinion mining, contextual advertisement and opinion summarization. However, sentiments are expressed differently in different domain and annotating label for every domain of interest is expensive and time consuming. In cross domain sentiment classification, a sentiment classifier trained in source domain is applied to classify reviews of target domain, always produce low performance due to the occurrence of features mismatch between source domain and target domain. The proposed method develops solution to feature mismatch problem in cross domain sentiment classification by creating enhanced sentiment sensitive thesaurus using wiktionary. The enhanced sentiment sensitive thesaurus aligns different words in expressing the same sentiment not only from different domains of reviews and from wiktionary to increase the classification performance in target domain. In this paper, the proposed method describes the method of construction of enhanced sentiment sensitive thesaurus which will be useful for cross domain sentiment classification.

Keywords Cross domain sentiment classification • Enhanced sentiment sensitive thesaurus • Domain adaptation • Data mining

P. Sanju (✉)

University College of Engineering, Anna University, Villupuram, Tamil Nadu, India
e-mail: Sanjupani_79@yahoo.co.in

T. T. Mirnalinee

SSN College of Engineering, Anna University, Chennai, Tamil Nadu, India
e-mail: mirnal23@gmail.com

1 Introduction

Nowadays purchasing of goods through online has been increased. In connection with online purchase, customers' opinions are shared to manufacturer through online. Such kind of opinions and sentiment information of the customers are overwhelming and growing exponentially which becomes a tedious work for the manufacture to classify the reviews manually. An automatic sentiment classifier is classification of reviews into positive or negative based on the sentiment words expressed in documents which is necessary to be developed for the manufacturer and the customer in order to analyze the reviews of the customers. The goal of sentiment classification is to discover customer opinion on a product. Sentiment classification has been applied in various tasks such as opinion mining [1], market analysis [2], opinion summarization [3] and contextual advertising [4].

Constructing general sentiment classifier is complex because in all domains the sentiment words have different connotation. Sentiment classification problem is more challenging because sentiments can be expressed in a more subtle manner. For example, the sentence, "How could anyone sit through this movie!" [5] contains not even a single sentiment words to express the sentiment but it is obviously negative. These kinds of sentences are available plenty while conveying the reviews of the product purchased. To develop sentiment classifier for these types of sentences is very complex.

Generally two types of approaches are being used for sentiment classification. The first approach is machine learning [5], in which classifier is trained using feature vectors and it produce more accurate results. The Second approach is semantic orientation [6, 7], it does not require prior training, instead it measures semantic orientation of sentences used in documents. Existing machine learning approaches [5, 8] rely on supervised learning models and performance of these models depends on manually labeled training data. However, such labeled data are not always available in practical applications and it is well known that sentiment classifier trained in one domain may not produce satisfactory results when it is used in another domain, since sentiments are expressed differently in different domain. Table 1 shows user review sentences from two domains books and electronics. In the books domain, the words "*well researched*" and "*interesting*" are used to express positive sentiments and "*disappointed*" is used to express negative sentiments. While in electronics domain "*fast*", *reliable*, *compact* and "*sharp*" are used to express the positive sentiment, "*blurry*" is negative sentiments. Due to the mismatch between domain specific words of source and target domain, classifier trained in one domain may not work well when directly applied to other domain.

In literature, Pan et al. [9] proposed spectral feature alignment algorithms to solve feature mismatch problem by aligning domain specific words from different domains into unified cluster with the help of domain independent words and then unified cluster is used to train a classifier in target domain. Bollegale et al. [10] proposed cross domain sentiment classification by creating sentiment sensitive

Table 1 Cross domain sentiment classification examples: Reviews of books and electronics products. Bold faces are domain specific words and italic words are domain independent words

Label	Books source domain	Electronics target domain
+	This book is <i>excellent</i> and well researched	SanDisk products are <i>excellent</i> , fast and reliable
+	This is an interesting and <i>Good</i> story book	<i>Good</i> Compact , easy to operate, looks sharp
-	When I read this book, I really disappointed , so <i>never buy</i> this book	It is blurry in dark settings, I would <i>never buy</i> this product

thesaurus which aligns different words that express the same sentiments. They expanded feature vector using created sentiment sensitive thesaurus while training a binary classifier. In existing approaches [10], they didn't give much more importance to the adjectives used in the reviews and also they collected more semantically similar sentiment elements from the reviews itself. In this paper, the proposed method creates an enhanced sentiment sensitive thesaurus (ESST) which aligns semantically similar words from various domains as well as sentiment elements from wiktionary. The idea behind this approach is, it is well known that adjectives are used to express sentiment in all domains, so, it is necessary to collect more adjectives or sentiment elements from any lexical knowledge base to solve the feature mismatch problem. The proposed method creates an enhanced sentiment sensitive thesaurus which collects more sentiment elements from wiktionary knowledge base because it has more semantic relatedness information.

To create an enhanced sentiment sensitive thesaurus, first, domain independent features and domain specific features are collected from the reviews. Second, co-occurrence matrix is computed between each domain independent features with domain specific features. Third, features are weighted using point wise mutual information. Finally, semantically similar words are collected for each domain specific word by computing similarity measure between two domain specific elements based on the PMI weighted ratio. At the same time semantically similar sentiment words for each adjective of the given reviews are collected from wiktionary with the help of java wiktionary library [11] and these sentiment words are added to the already created thesaurus.

In rest of the paper is organized as follows: [Sect. 2](#) describes related work. [Section 3](#) describes definitions and Methodology used in proposed work. [Section 4](#) describes experimental setup and solution. [Section 5](#) concludes the proposed work.

2 Related Works

Many machine learning techniques have been deployed for sentiment classification. The existing methods are helpful in classifying the sentiments at various levels i.e. Document level [9, 10], sentence level [12] and word level [13, 14]. The

proposed work is based on document level that classifies the documents as positive or negative, based on overall sentiment words expressed in the documents. Two approaches have been used in existing sentiment classification studies; machine learning and semantic orientation. In machine learning approaches, documents are represented as feature vector and it is trained by various classification algorithms such as Naive Bayes, Maximum entropy and SVM [8]. Two types of semantic orientation approaches [15]: Corpus based and dictionary based techniques used in existing research work. Corpus based techniques aim to find co-occurrence patterns of the words to determine the sentiments. Dictionary based methods utilize synonyms, antonyms and hierarchies in wordnet or sentiwordnet to determine the sentiments.

Turney [6] used supervised learning techniques with mutual information to predict overall document sentiment by averaging out the semantic orientation of phrases in document. Turney [5] used Naive Bayes, Maximum Entropy and SVM for classifying movie reviews and received best results using SVM. Kennedy and Inkpen [16] used Contextual valence shifter to predict sentiment of the sentences. The Work on the papers [5, 6, 16] focuses only on classify the given reviews in particular domain. Lin et al. [17] proposed a novel probabilistic modeling framework called joint sentiment topic model which detects sentiments and topic simultaneously from text.

In sentiment classification, many researchers have been concentrated on online lexical resources such as sentiwordnet, wiktionary and Wikipedia which are publically available for research purpose. SentiWordNet [18] is a lexical resource, containing opinion information on terms extracted from the wordnet database where each term is associated with numerical scores indicating positive and negative information. Liu et al. [19] proposed movie rating and review summarization in mobile environment using Latent semantic analysis. Khan et al. [20] proposed sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and blog which classify the reviews into objective or subjective sentences. The semantic score of the subjective sentences are extracted from sentiwordnet to calculate their polarity as positive or negative.

Zesch et al. [11] proposed extracting lexical semantic knowledge from Wikipedia. They proposed two application programming interfaces for Wikipedia and wiktionary which are designed for mining the lexical semantic information from knowledge bases. Chesley et al. [21] proposed automatic classification of blog posts using verbs, adjectives and information from wiktionary. They used wiktionary to determine polarity of the adjectives in the text. Chihli et al. [22] proposed using objective words in sentiwordnet to improve sentiment classification for word of mouth. Wiebe [23] proposed learning subjective adjectives from corpora. They identified subjectivity of sentences form corpora by clustering the words according to distributional lin's similarity measure, seeded by a small amount of detailed manual annotation. Hatzivassiloglou et al. [24] proposed predicting the semantic orientation of adjectives which describes conjunction between adjectives provides indirect information about semantic orientation of sentences. The proposed method is different from existing approaches [23, 24], more

sentiment elements are extracted from the wiktionary with the help of seed adjectives of the given reviews which will be useful to predict the sentiment of target domain. Muller et al. [25] proposed Domain-Specific Information Retrieval using wiktionary and Wikipedia.

Many researchers have addressed the problem in cross domain sentiment classification [9, 10, 18, 26]. Blitzer et al. [26] addressed the problem in cross domain sentiment classification using structural correspondence learning algorithm where frequent words in both source and target domain were selected as candidate pivot features and linear predictors are trained to predict the occurrences of those pivot features. In structural corresponding learning-mutual information approach, the mutual information between a feature and the domain label is used to select pivot features instead of co-occurrence frequency. Pan et al. [9] proposed a spectral feature alignment algorithm to align domain specific words from different domain into unified clusters with the help of domain independent words as a bridge and bipartite graph is constructed between domain specific and domain independent features and then the domain specific clusters can be used to train a sentiment classifier in target domain. Yan et al. [27] used self growth algorithm to generate a cross domain sentiment word list which is used in sentiment classification of web news. Bollegale et al. [10] proposed cross domain sentiment classification by creating sentiment sensitive thesaurus which aligns different words that express the same sentiments. They expanded feature vector using created thesaurus while training a binary classifier.

All existing cross domain sentiment classification solution, identify the sentiment of unlabeled data of target domain with the help of labeled data in source domain by applying various methodology. The proposed solution for cross domain sentiment classification predict the unlabeled data of target domain by creating extended sentiment sensitive thesaurus using reviews and wiktionary which increases the classification performance in target domain.

3 Methodology

3.1 Definitions

In this section, some definitions are given to clarify basic terminology.

Domain-A domain D denotes a class of entities in the real world. For example, different types of product such as DVD, Kitchen appliances, books and electronics.

Sentiment-Given a specific domain D , sentiment data are the text documents which express the user opinion about the entities of the domain.

Generally, in sentiment classification tasks, single word (unigram), bigram, ngram are used as features. In this work, unigram and bigram are used as features.

Cross Domain Sentiment Classification-Given a set of labeled reviews $D_s = \{(x_i, y_i)\}$ from source domain where x_i represents features and y_i represent

sentiment label $y_i \in \{+1, -1\}$. To predict the label of unlabeled target domain $D_t = \{x_j\}$ where x_j represent features in target domain. Classifier is trained by labeled reviews source domain and it is applied to classify the unlabeled reviews of target domain.

3.2 *Enhanced Sentiment Sensitive Thesaurus (ESST) Using Wiktionary*

In cross domain sentiment classification, sentiment classifier trained from one domain (source) is applied to another (target) domain. Obviously it produces poor performance because trained features are mismatched with target domain features. This feature mismatch problem in cross domain sentiment classification is solved by creating enhanced sentiment sensitive thesaurus using all domain reviews and knowledge from Wiktionary. Wiktionary is online dictionary which has glosses and synset for each word. Every word in the wiktionary is clearly defined with list of synonyms. Wiktionary also contains more information which is found in linguistic knowledge base (wordnet), like definitions, synonyms, and hyponyms, and also has additional types of information, e.g. abbreviations, compounds or contractions, which are usually not found in LKBs. Another difference to LKBs is that Wiktionary contains not only lexical knowledge for particular language, but also for other languages. Figure 1 shows the construction of enhanced sentiment sensitive Thesaurus. The Enhanced sentiment sensitive Thesaurus is constructed using the knowledge from the Wiktionary, labeled/unlabeled reviews from source domain and unlabeled reviews from target domain. To interact with wiktionary, JWKTl (java wiktionary library) is used which is released in Ubiquitous Knowledge Processing Lab [11].

Given a labeled or unlabeled review, first split the reviews into set of sentences, and then perform parts of speech (POS) tagging and Lemmatization is performed using RASP [28] system. Lemmatization is used to reduce the features by removing inflectional endings only and to return the base or dictionary form of a word, which is known as the *lemma*. By using a stop word removal algorithm based on the POS tagging to filter out unwanted words and retaining nouns, verbs, adjectives, adverbs. Table 2 shows how to extract unigrams, bigrams and sentiment elements from reviews. First, model the review as bag of words and then extract unigrams, bigrams from each sentence. Bigrams are necessary in sentiment classification, since semantic orientations of sentences are identified by bigrams. Generally unigrams, bigrams are called as lexical elements or sentiment elements. Next, from each source domain labeled reviews, sentiment elements are created by appending the label of the review to each lexical element. The notation *p to indicate positive sentiment elements and *N to indicate negative elements.

Domain independent features are extracted from all domains which occur frequently in all domains. Domain specific features from various domains are

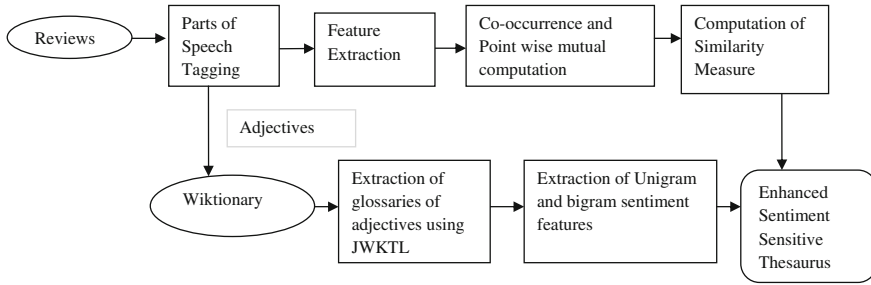


Fig. 1 Construction of enhanced sentiment sensitive thesaurus

Table 2 Example for extracting unigram, bigram and sentiment elements from positive review sentence

Sentence	An excellent workbook full of delicious recipes
POS tags	An_AT1 excellent_JJ workbook_NN1 full_JJ of_IO delicious_JJ recipes_NN2
Unigrams	Excellent, workbook, full, delicious, recipes
Bigrams	Excellent + workbook, workbook + full, full + delicious, delicious + recipes
Sentiment elements	Excellent * p, workbook * p, full * pl, delicious * p, recipes * p Excellent + workbook * p, workbook + full * p, full + delicious * p, delicious + recipes * p

aligned in ESST with the help of domain independent features. Features other than domain independent features are considered as domain specific features. Here, each domain independent feature is represented as feature vector based on the domain independent feature co-occurring with its distributional contexts or feature. Later on the co-occurrence matrix is computed between domain independent features with domain specific features. Semantic meanings of the words are found based on co-occurrences of the terms used in the documents. Co-occurrence [29] is used to measure the similarity between meanings of the two words based on words distributional contexts.

After the computation of co-occurrence between domain independent features and domain specific features, value of the domain specific features in the co-occurrence matrix are weighted using point wise mutual information [13]. Point wise Mutual information is an information theoretic measure of association between two lexical elements. Point wise mutual information is useful in numerous tasks such as similarity measurement [30], word classification [31] and word clustering [32]. To align domain specific features, transpose PMI weighted matrix and find similarity between domain specific words based on its distributional contexts.

Point wise mutual information (PMI) is used to weight the features based on the co-occurrence value in the co-occurrence matrix. For each domain specific feature s , domain independent feature v that co-occurs with domain specific features s contributes feature vector \mathbf{s} . The value of the feature v in vector \mathbf{s} is denoted by $M(\mathbf{s}, v)$.

$$M(s, v) = \log \left(\frac{\frac{c(s,v)}{N}}{\frac{\sum_{i=1}^n C(i,v)}{N} \times \frac{\sum_{j=1}^m C(s,j)}{N}} \right) \quad (1)$$

$M(s, v)$ is the point wise mutual information between a domain independent features s and domain specific feature v . In Eq. 1, $C(s, v)$ represents number of review sentences in which both domain independent feature s and domain specific feature v co-occur, $C(s, j)$ represents number of times domain independent feature s occur in review sentences. $C(i, v)$ represents number of times domain specific feature v occur in the review sentences. n and m denotes the total no of domain independent elements and domain specific features and $N = \sum_{i=1}^n \sum_{j=1}^m C(i, j)$.

After the computation of PMI values, the semantically similar features for each domain specific feature is computed by finding similarity score between domain specific features with every other domain specific features of PMI weighted matrix. For example, similarity score between two domain specific features s, t (both s and t have feature vectors \mathbf{s}, \mathbf{t}) is computed by following formula

$$T(s, t) = \frac{\sum_{v \in \{x | M(s,x) > 0\}} M(\mathbf{t}, v)}{\sum_{v \in \{x | M(t,x) > 0\}} M(\mathbf{t}, v)} \quad (2)$$

The similarity score $T(s, t)$ is the proportion of PMI weighted values of the domain specific feature t that are shared with domain specific feature s . The distributional hypothesis states that words that have similar distributions are semantically similar [30] i.e. Two words are semantically similar if two words occur with same distributional contexts.

Enhanced Sentiment sensitive thesaurus (ESST) aligns many domain specific features that are semantically similar from various domains by computing similarity measure Eq. (2) for every domain specific feature in PMI weighted matrix. For every domain specific feature s , ESST list up many domain specific features t based on descending order of the similarity score.

Moreover, Enhanced Sentiment Sensitive Thesaurus (ESST) collect more semantically similar features from wiktionary using java wiktionary library (JWKTL) which is very useful tool to extract more semantic relatedness from Wiktionary. To enhance sentiment sensitive thesaurus, first, seed adjectives are extracted from the given reviews. Second, XML dump file of wiktionary is downloaded from web. Third, using JWKTL, the XML file is parsed and data is extracted from wiktionary. Next, the extracted adjectives are given as input to the JWKTL and the corresponding glossaries of each adjective are extracted from the wiktionary. Finally, unigram and bigrams from the extracted glossaries are collected for each adjective and then these are appended to the ESST with similarity score of 0.9. This enhanced sentiment sensitive thesaurus is entirely different from normal thesaurus because it finds the similar meaning of the words based on the co-occurrence terms used in the sentences. The created ESST is very useful to classify the reviews of target domain in cross domain sentiment classification and also it will increase the classification performance in target domain.

3.3 Algorithm: Creation of Enhanced Sentiment Sensitive Thesaurus

Input: labeled source Domain data $D_{sr} = \{X_{sr}, Y_{sr}\}$ and unlabeled source domain $D_{sr} = \{X_i\}$ and unlabeled target Domain $D_{tr} = \{X_{tr}\}$ and Wiktionary dump file
Output: Enhanced Sentiment Sensitive Thesaurus

1. Extract Domain Independent features and domain specific features from the given Reviews.
2. Create co-occurrence matrix between domain independent features with domain specific features.
3. Compute Point Wise Mutual Information for each features using Eq. (1).
4. Compute similarity measure using Eq. (2) for every domain specific features with every other domain specific features.
5. Enhanced sentiment thesaurus aligns many domain specific features from various domains for every domain specific feature.
6. Extract seed adjectives from the given Reviews and these seed adjectives are given as input to the wiktionary parser.
7. Glossaries for each seed adjectives are extracted using java wiktionary library(JWKTLL)
8. Unigram and bigrams are generated from the glossaries that are added with Enhanced Sentiment sensitive thesaurus.

4 Experiments

4.1 Data Sets

Amazon product reviews are bench mark data set which consist of four different product types: Books, DVDs, electronics and kitchen appliances are chosen for the proposed work. Each review is assigned with rating (0–5 stars). Review with rating >3 are labeled as positive and review with rating <3 are labeled as negative. The data sets structure is shown in Table 3. For each domain; there are 1,000 positive reviews with 1,000 negative reviews. Each domain also has some unlabeled reviews.

For experiments, among four domains, three domains are selected as source domain and another domain as target domain. To create enhanced sentiment sensitive thesaurus, 800 positive reviews and 800 negative reviews are selected from each source domain and some unlabeled reviews are selected from all domain. Wiktionary dump file is downloaded and stored as separate file. More sentiment features are extracted from wiktionary with the help seed adjectives of the given reviews. Here enhanced sentiment sensitive thesaurus is created automatically by considering each domain as target domain. So four different types of

Table 3 Amazon product data sets

Domain	Positive	Negative	Unlabeled
Kitchen	1,000	1,000	16,746
DVDs	1,000	1,000	34,377
Electronics	1,000	1,000	13,116
Books	1,000	1,000	5,947

thesaurus is created to solve the feature mismatch problem in cross domain sentiment classification

The enhanced sentiment sensitive thesaurus (ESST) is automatically created by giving reviews of source and target domain. To study the effect of multiple thesauruses, different enhanced sentiment sensitive thesaurus is created for one source domain/one target domain and two source domain and one target domain. From the analysis of multiple created thesauruses, enhanced sentiment sensitive thesaurus created from three source domains may be effective because more sentiment features are identified from the reviews as well as from the wiktionary. Moreover, this enhanced sentiment sensitive thesaurus highly effective to classify the reviews of unseen target domain features because it collect more semantically similar features from wiktionary. For example, ESST align 880 features for the adjective good from reviews as well as from wiktionary

5 Conclusions

Enhanced Sentiment Sensitive Thesaurus is created automatically based on the given source domain and target domain reviews. The proposed ESST which aligns more semantically similar sentiment features from source and target domain as well as additional sentiment features from wiktionary in order to increase the classification performance in target domain. This ESST is suitable to the applications when source domain has labeled data and target domain has unlabeled data. Here the gap between source and target domain are bridged by leveraging additional knowledge from wiktionary. In future, cross domain sentiment classification will be performed by feature vector augmentation using created Enhanced Sentiment Sensitive Thesaurus during training and testing time of sentiment classification.

References

1. Pang, B., Lee, L.: Opinion mining and sentiment analysis. *Found. Trends Inf. Retrieval* 2(1–2), 1–135 (2008)
2. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: *Proceeding of the 10th ACM SIG KDD International Conference on Knowledge Discovery and Data Mining*, pp. 168–177. New York, USA (2004)

3. Lu, Y., Zhai, C., Sundaresan, N.: Rated aspect summarization of short comments. In: *The Proceedings of the 18th Conference on World Wide Web*, pp. 131–140. ACM, New York (2009)
4. Fan, T.-K., Chang, C.H.: Sentiment-oriented contextual advertising. *Knowl. Inf. Syst.* **23**(3), 321–344 (2010)
5. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? sentiment classification using machine learning techniques. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 79–86. Philadelphia, July 2002
6. Turney, P.D.: Humbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: *ACL'02 Proceedings of the 40th Annual Meeting on association for Computational Linguistics*, pp. 417–424 (2002)
7. Hatzivassiloglou, V., McKeown, K.R.: Predicting the semantic orientation of adjectives. In: *ACL 1997*, pp. 174–181 (1997)
8. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: *Proceedings 10th European Conference on Machine Learning (ECML)*, pp. 137–142. Springer Verlag, Berlin (1998)
9. Pan, S.J., Ni, X., Sun, J.T., Yan, Q., Chen, Z.: Cross-domain Sentiment classification via spectral feature alignment. In: *WWW 2010, Raleigh, North Carolina, USA*. ACM 978-1-60558-799-8/10/04, 26–30 April 2010
10. Bollegala, D., Weir, D., Carroll, J.: Cross domain sentiment classification using a sentiment sensitive thesaurus. *IEEE Trans. Knowl. Data Eng.* **25**(8) (2013)
11. Zesch, T., Muller, C., Gurevych, I.: Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary (2008)
12. Yu, H., Hatzivassiloglou, V.: Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In: *EMNLP 2003*, pp. 129–136 (2003)
13. Pantel, P., Ravichandran, D.: Automatically labeling semantic classes. In: *NAACL-HLT'04*, pp. 321–328 (2004)
14. Breck, E., Choi, Y., Cardie, C.: Identifying expressions of opinion in context. In: *IJCAI 2007* (2007)
15. Turney, P.: Similarity of semantic relations. *Comput. Linguist.* **32**(3), 379–416 (2006)
16. Kennedy, A., Inkpen, D.: Sentiment classification of movie reviews using contextual valence shifters. *Comput. Intell.* **22**, 110–125 (2006)
17. Lin, C., He, Y., Everson, R., Ruger, S.: Weakly supervised joint sentiment- topic detection from text, J. Latex class files, 24 (2011)
18. Esuli, A., Sebastiani, F.: Sentiwordnet: a publicly available lexical resource for opinion mining. In: *LREC 2006*. pp. 417–422 (2006)
19. Liu, C.L., Hsaio, W.H., Lee, C.H., Lu, G.C., Jou, E.: Movie Rating and review summarization in mobile environment. In: *IEEE 2012*, vol. 42, pp. 397–407 (2012)
20. Khan, A., Baharudin, B.: Sentiment classification by sentence level semantic orientation using sentiwordnet from online reviews and Blogs. *Int. J. Comput. Sci. Emerging Tech.* **2**(4) (2011)
21. Chesley, P., Bruce Vincent, B., Xu, L., Srihari, R.K.: Using verbs and adjectives to automatically classify blog sentiment. In: *American Association for Artificial intelligence* (2005)
22. Hung, C., Lin, H.: Using objective words in sentiwordnet to improve the sentiment classification word of mouth. In: *IEEE Intelligent Systems* (2013)
23. Wiebe, J.: Learning subjective adjectives from corpora. In: *American Association for AI* (2001)
24. Hatzivassiloglou, V., McKeown, K.R.: Predicting the Semantic Orientation of Adjectives .pp. 176–182 (1995)
25. Muller, C., Gurevych, I.: Using Wikipedia and Wiktionary in Domain-Specific Information Retrieval (2008)

26. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: EMNLP 2006 (2006)
27. Yan, L., Zhang, Y.: News sentiment Analysis based on cross domain sentiment word lists and content classifiers. ADMA 2012.LNAI 7713, pp. 577–588. Springer-verlag, Berlin (2012)
28. Briscoe, T., Carroll, J., Watson, R.: The second release of the RASP system. In: COLING/ACL 2006 Interactive Presentation Sessions (2006)
29. Weeds, J., Weir, D.: Co-occurrence retrieval: a flexible framework for lexical distributional similarity. *Comput. Linguist.* **31**(4), 439–475 (2006)
30. Harris, Z.: Distributional structure. *Word* **10**, 146–162 (1954)
31. Turney, P.: Similarity of semantic relations. *Comput. Linguist.* **32**(3), 379–416 (2006)
32. Lin, D.: Automatic retrieval and clustering of similar words. In: ACL, pp. 768–774 (1998)
33. Hu, K., Zhang, Y., Hu, X.: A multi-domain adaptation for sentiment classification algorithm based on class distribution. In: IEEE International Conference on Granular Computing (2012)
34. Remus, R.: Domain adaptation using domain similarity- and domain complexity-based instance selection for cross-domain sentiment analysis. In: IEEE 12th International Conference on Data Mining Workshops (2012)

Comparative Analysis of Energy Aware Protocols in Wireless Sensor Network Using Fuzzy Logic

Rajeev Arya and S. C. Sharma

Abstract Wireless sensor network is crowded network, which consists lots of nodes. There is some constrained in wireless sensor network like awareness of energy, environmental constraints like temperature, pressure, sound. The aim of this research is to analyze the energy efficient operation in the sensor node. For this purpose, Low energy adaptive clustering hierarchy (LEACH), Stable Election Protocol (SEP) and Gateway Cluster Head Election-FL (GCHE-FL) protocols were analyzed and compared. Most of the researchers have checked topologies and architectures that allow energy efficient operation in the wireless sensor node. Clustering is one of the most popular techniques to reduce the energy in the sensor network. Various node clustering methods have been reported in the literature. In the present research article, the fuzzy logic technique is used in Cluster Head Election and Protocol Gateway for heterogeneous sensor network. It is observed that in GCHE-FL, the sensor node being alive for much more time as compared to LEACH, SEP.

Keywords Wireless sensor network · Fuzzy logic · Clustering · Energy

1 Introduction

Wireless communication is one of the rapidly growing technologies. The demand for connecting device without the use of cable is in increasing everywhere. Wireless network consisting of spatially distributed autonomous device to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants, at different locations. Each sensor

R. Arya (✉) · S. C. Sharma
Wireless Computing Research Lab, IIT Roorkee, SRE Campus, Roorkee, India
e-mail: rajeev.arya.iit@gmail.com

S. C. Sharma
e-mail: scs60fpt@iitr.ernet.in

has capability of wireless communication and a certain level of intelligence for signal processing and networking of the data and communicates using radio link. A sensor network can be quickly and easily deployed and thus is suitable and much attractive for many environments, commercial and military applications [1–4].

A general-purpose sensor network is commonly a dense network that consists of a large number of energy-constrained nodes; it is likely to be deployed in difficult accessible regions and to be remotely operated by only a few operators. Therefore energy becomes the most critical resource for efficient use of sensor deployment. As a result, conserving energy should be a primary demand of the protocols designed for such network. The research interest in different aspects related to deploying and further deployment of sensor network has been increasing in the last few years.

2 Research Review

Heinzelman et al. suggested low energy adaptive clustering (LEACH) protocol [5–7] is separated into different rounds. The each round consists two phases, such as study state phase and setup phase. Gupta et al. [8], Fuzzy logic protocol, select Cluster Head by fuzzy logic approach. Fuzzy logic interference scheme designer considered three descriptors Centrality, concentration and energy level and each descriptor is three parts, and chance output. Smaragdakis et al. [9] cluster heterogeneous WSN by using a stable election protocol. Tashtoush et al. [10–12] suggested the Clustering distributed WSN based on fuzzy inference scheme. The first election is Gateway Election and the capable node is selected based on their energy, centrality and proximity to the base station. The second election is cluster head election used three parameters. These parameters are efficiency, cluster distance and concentration. It is the ratio between the residual energy of each node and the average energy of the cluster. The cluster distance is the sum of the distance between the node and the others nodes within the cluster. The purpose of this protocol is increasing the lifetime of the wireless sensor network through consistently distributing the clusters on the WSN. In the present paper, we analyzed the different protocol (LEACH, SEP, GCHE-FL) using fuzzy logic scheme and Matlab simulation. The objective of the present work is to compare performance of different energy aware routing algorithms to study under specified constraints using fuzzy logic scheme.

3 LEACH, SEP Protocol

Heinzelman et al. suggested low energy adaptive clustering hierarchy (LEACH) protocol [5–7], LEACH is a TDMA-based MAC protocol. The round consists two phases, such as study state phase and setup phase. The steady state phase consists

of transmit and sending of data to the cluster head (CH) and after that to the base stations (BS). The setup phases consist of cluster formation and different Cluster head. In every round and start of the first setup phase and all sensor node generates 0 and 1 as random number to check that it will cluster head or not.

SEP a heterogeneous-aware protocol which extend the time period earlier than the death of the first node, which is crucial for lots of applications wherever the advice as of the sensor network have to be consistent [9]. SEP is supported on probabilities of every node to become cluster head according to the left over energy in every node. SEP protocol effectively extends the constant area by being conscious of heterogeneity through transmission probabilities of cluster-head election biased through the relation opening energy of nodes. SEP is extra flexible than LEACH in thoughtfully consuming the additional energy of higher nodes SEP yields a longer fidelity area for superior values of additional energy.

4 Gateway Cluster Head Election-Fuzzy Logic (GCHE-FL)

Gateway cluster head election-Fuzzy logic uses if then rule to maximize the life time of wireless sensor network. It used two elections fuzzy logic to evaluate the merit of sensor to become gateway and cluster head. The first election (Gateway Election) capable node is selected based on their energy and closeness. The second election is cluster head election.

In the gateway election, the node’s possibility for being a gateway is evaluated according to its physical properties such as its interior energy and its proximity to the base station. Such as traffic load is intense at most on cluster heads [8]. As on one hand, it gathers the data upcoming from the cluster associate nodes, and on the next step. It’s onward the aggregated data to sink. To decrease the energy expenditure of cluster head and reduce probability of failure, gateways had used to send out data from cluster heads to the sink.

$$T(n) = \begin{cases} \frac{q}{1 - q * (r \bmod \frac{1}{q})} & \times \text{Chance if } n \in G_s \\ 0 & \text{otherwise} \end{cases}$$

Every wireless sensor node elects itself to be a gateway at the start of every round with probability q_g .

Gateway Cluster Head Election-FL configures clusters in all rounds [8]. In all round, every non-gateway wireless sensor node creates a random number between 0 and 1, if these random numbers are lesser than q_{opt} , the wireless sensor node analyzes the chance with fuzzy if then rule and presented an applicant Message through the chance.

K_{opt} = optimal number of the cluster head, δ = constant Value, b = fraction of the all node

$$q_{opt} = \delta(k_{opt} / n \cdot (1 - b)),$$

$$k_{opt} = (\sqrt{n \cdot (1 - b) / 2\pi}) * \sqrt{\epsilon fs / \epsilon mp} * \left(\sqrt{\frac{area}{0.75 \times 0.5 \times \sqrt{area} 2}} \right).$$

Gateway Cluster Head Election-FL analyzes a chance with fuzzy if-then rule. The better chance means that the node has extra chance to be a cluster head.

5 Fuzzy Logic Approach

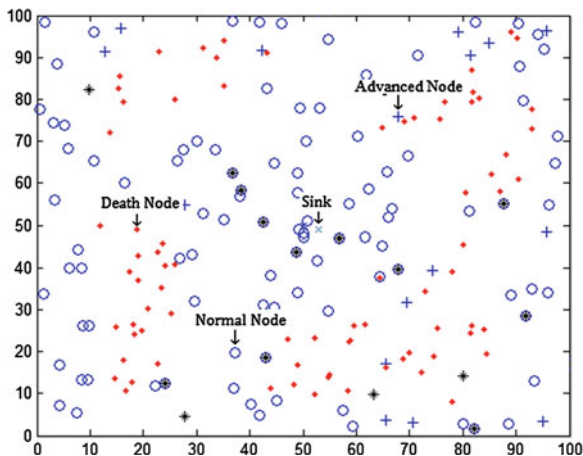
The system of fuzzy logic consists of a fuzzifier, fuzzy system, fuzzy inference, and a defuzzifier. In the present analyses we have used the commonly fuzzy inference method called Mamdani due to its ease. The fuzzy logic (FL) approach consists of four step that is fuzzification of the input variables, Rule assessment (inference), Aggregation of the rule outputs (composition) and Defuzzification [10, 11]. In this part, we initiate GCHE-FL which uses fuzzy if-then rule to exploit the lifetime of WSN. GCHE-FL uses two-election fuzzy logic to judge the qualification of sensors to become the cluster head and a gateway. There is the first election (Gateway Election), the qualified nodes are elected based on their energy and proximity to the base station. And second election (Cluster Head Election) used on tow fuzzy parameters. The ratio between the residual energy (standard energy of the cluster) of every node and Cluster distance, equal the sum of distances among the node and the others nodes which is in cluster. Routing in GCHE-FL works in around and every round is separated into a fist step phase Gateway electorate, cluster head election and cluster formation and second step steady state phase.

6 System and Model (Characteristics of Model)

- Sensor node is at check extending in the surroundings node.
- BS (Base station) is positioned in the focal point of the Surroundings (WSN).
- One arranged, the node does not travel.
- All wireless sensor nodes have the same ability and energy.
- Base station arranged the id and location for all nodes.
- Every node is dispersed homogeneously over the sensor area (Fig. 1).

Energy spending model has been used in this paper similar to energy model in the LEACH protocol [5–7] in which every node to throw 1 byte data to space of it consumes a lot E_s energy, which is gated from this equation. Also the full quantity of energy that is used in the receiver for getting k bit node is gated from equation.

Fig. 1 Snap shot of wireless sensor network area and indicator and nodes



$$E_s = \left\{ \begin{array}{ll} (l \times E_{elect} + l \times \varepsilon_{fs} \times d^2), & d > d_{co} \\ (l \times E_{elect} + l \times \varepsilon_{mp} \times d^4), & d \geq d_{co} \end{array} \right\} \text{ and } (E_r = l \times E_{elect})$$

7 Fuzzy Analysis of Gateway Election

Therefore, in the gateway election, these three parameters, energy, centrality and proximity to BS are fuzzy logic input and sensor chance output. These are input parameter first on energy (poor, good excellent) second on proximity to base (poor, good excellent) and third on the centrality (close, adequate, far). Then $3^3 = 27$ rules are used for fuzzification to sensor chance output. The centroid based defuzzification is used. Show in the gateway election Figs. 2, 3. The if then rules for gate way election given in Table 1.

8 Fuzzy Analysis of Cluster Head Election

To calculate a chance, we have used three fuzzy logics variables concentration Cluster Distance and efficiency. These are input parameter first on efficiency (Low, medium, high) second in the Cluster Distance (low, medium, high) and third on concentration (Low medium, high). Then $3^3 = 27$ rules are used for fuzzification to sensor chance output. The centroid based defuzzification is used. Show in the Cluster head election (Figs. 4 and 5). The if then rules for Cluster head election given in Table 2.

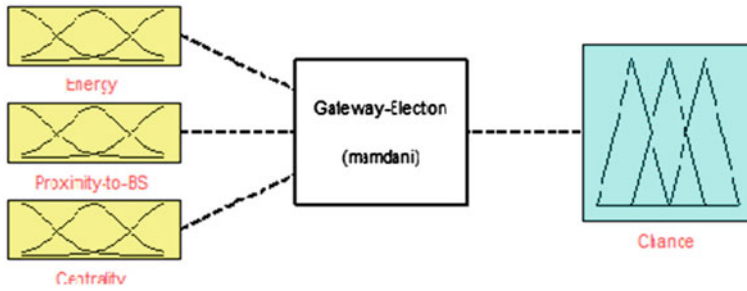


Fig. 2 Gateway-election system architecture

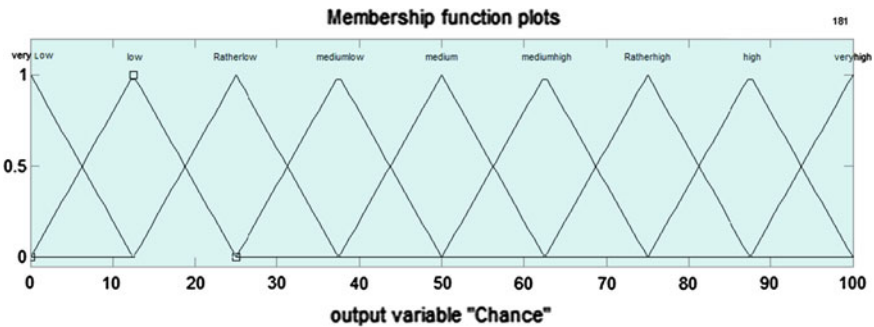


Fig. 3 Membership functions of output variable chance

9 Simulation Parameter

They have been considered for Low Energy Adaptive Clustering Hierarchy, Stable Election protocol and Gateway Cluster Head Election-FL used fuzzy tool (MATLAB R2008b). The area of wireless sensor network is chosen as of 100 m × 100 m and sink located at center of sensing field. The simulation parameters in wireless network area.

Total Energy ($E_{elect.}$) = 50 nJ/bit, Data Aggregation Energy (E_{DA}) = 5 nJ/bit/message, Initial Energy (E_o) = 0.5 J, Maximum number of rounds (R) = 4,999, Number of Nodes in the field (n) = 200, Transmit Energy (E_{TX}) = 50 nJ/bit, Receive Energy (E_{RX}) = 50 nJ/bit, Transmit Amplifier types (ϵ_{fs}) = 10 pJ/bit/(m_2), Transmit Amplifier types (ϵ_{mp}) = 0.0013 pJ/bit/(m_4).

Table 1 If then rule under fuzzy for gateway election

Input variable			Output
Energy	Proximity-to-BS	Centrality	Chance
Poor	Poor	Close	Low
Poor	Poor	Adequate	Low
Poor	Poor	Far	Very low
Poor	Good	Close	Low
Poor	Good	Adequate	Low
Poor	Good	Far	Low
Poor	Excellent	Close	Rather low
Poor	Excellent	Adequate	Low
Poor	Excellent	Far	Very low
Good	Poor	Close	Rather low
Good	Poor	Adequate	Medium
Good	Poor	Far	Low
Good	Good	Close	High
Good	Good	Adequate	Medium
Good	Good	Far	Rather low
Good	Excellent	Close	High
Good	Excellent	Adequate	Rather high
Good	Excellent	Far	Rather low
Excellent	Poor	Close	Rather high
Excellent	Poor	Adequate	Medium
Excellent	Poor	Far	Rather low
Excellent	Good	Close	High
Excellent	Good	Adequate	Rather high
Excellent	Good	Far	Medium
Excellent	Excellent	Close	Very high
Excellent	Excellent	Adequate	Rather high
Excellent	Excellent	Far	Medium

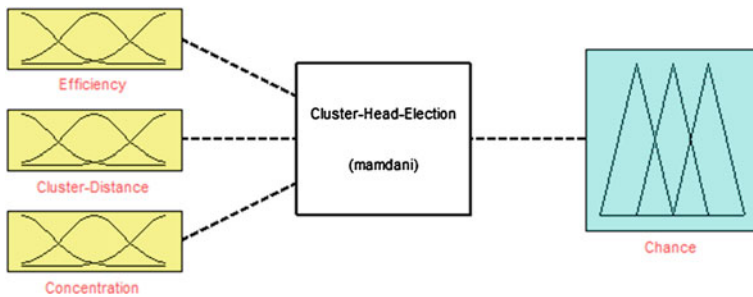


Fig. 4 Cluster-head-election system architecture

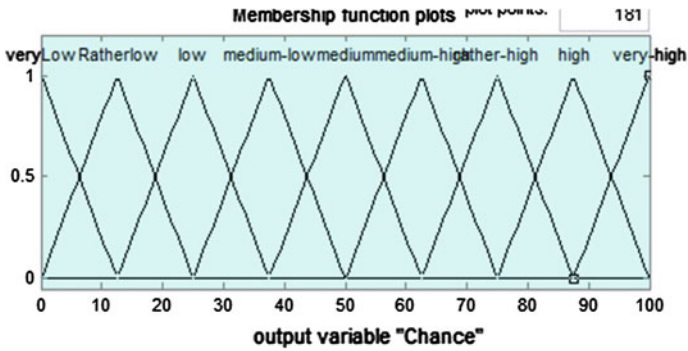


Fig. 5 Membership functions of output variable chance

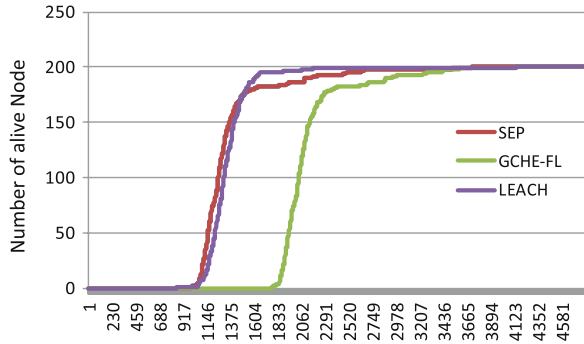
Table 2 If then rule under fuzzy for cluster head election

Input variable			Output
Efficiency	Cluster-distance	Concentration	Chance
Low	Low	Low	Low
Low	Low	Medium	Rather low
Low	Low	High	Very low
Low	Medium	Low	Low
Low	Medium	Medium	Rather low
Low	Medium	High	Very low
Low	High	Low	Low
Low	High	Medium	Rather low
Low	High	High	Very low
Medium	Low	Low	Medium low
Medium	Low	Medium	Medium
Medium	Low	High	Medium high
Medium	Medium	Low	Medium low
Medium	Medium	Medium	Medium
Medium	Medium	High	Medium high
Medium	High	Low	Medium low
Medium	High	Medium	Medium
Medium	High	High	Medium high
High	Low	Low	Rather high
High	Low	Medium	High
High	Low	High	Very high
High	Medium	Low	Rather high
High	Medium	Medium	High
High	Medium	High	Very high
High	High	Low	Rather high
High	High	Medium	High
High	High	High	Very high

Table 3 Comparative analyses of different protocols (With $m = 0.1, a = 1$)

	LEACH	SEP	GCHE-FL
First node dies	946	991	1,775
Half of the node alive	1,373	1,298	2,148
10 % alive	1,505	1,683	2,721

Fig. 6 Number of death node per round for different protocols



10 Result and Analysis

In this paper, we compare performance of three protocols based on fuzzy logic using Matlab. Wireless sensor network area (100 m × 100 m), 200 nodes and 4,999 numbers of rounds (Table 3). Figure 6 based on attribute of nodes show that GCHE-FL far better than stable election protocol and LEACH due to change in fuzzy rules because of extra fuzzy input.

It is observed from the fuzzy analysis that first node die after 946, 991 and 1,775 numbers of rounds in LEACH, SEP, GCHE-FL respectively. Half of the node being alive in LEACH, SEP, GCHE-FL after 1,373, 1,298, 2,148 number of rounds respectively. In wireless sensor network areas only 10 % nodes being alive after 1,505, 1,683 and 2,721 numbers of rounds in LEACH, SEP, GCHE-FL respectively. On the base of simulation result we found that GCHE-FL is better than LEACH and SEP.

11 Conclusions

In this paper we have analyzed three different protocols (LEACH, SEP, GCEH_FL) for Energy of sensor node. It is observed that in GCEH_FL, the sensor node being alive for much more time as compared to LEACH, SEP, which indicate that energy expenditure of node is less as compared with LEACH and SEP. which implies that GCHE-FL of performed well in heterogeneous wireless sensor

network. The simulation results of protocol consumes less energy and prolong lifetime of the network compared with others protocol.

The future direction of this work will be to calculate the optimal energy expenditure using fuzzy logic system.

References

1. Chong, C.Y., Kumar, S.P.: Sensor networks: evolution opportunities, and challenges. *Proc IEEE* **91**(8), 1247–1256 (2003)
2. Haenggi, M.: *Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems*. CRC Press (Chapter 1 Opportunities and challenges in wireless sensor networks, pp. 1–11) (2005)
3. Estrin, D., Girod, L., Pottie, G., Srivastava, M.: Instrumenting the world with wireless sensor networks. In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2001)*, pp. 2033–2036, 2001
4. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. *IEEE Commun. Mag.* **40**, 102–114 (2002)
5. Heinzelman, W.R., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wireless Commun.* **1**(4), 660–670 (2002)
6. Said, B.A., Abdellah, E., Hssane, A.B., Hassnaoui, M.L.: Improved and balanced LEACH for heterogeneous wireless sensor networks. *Int. J. Comput. Sci. Eng.* **2**(8), 2633–2640 (2010)
7. Qing, L., Zhu, Q., Wang, M.: Design of a distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks. *Comput. Commun.* **29**, 2230–2237 (2006)
8. Gupta, I., Riordanand, D., Sampalli, S.: Cluster-head election uses fuzzy logic for wireless sensor networks. In: *Communication Networks and Services Research Conference*, pp. 255–260, May 2005
9. Smaragdakis, G., Malta, I., Bestavros, A.: A stable election protocol for clustered heterogenous wireless sensor network. In: *Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004)*, 2004
10. Tashatoush, Y.M., Okour, M.A.: Fuzzy self—clustering for wireless sensor network. In: *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing*, 2008
11. Daliri, Z., et al.: Railway security through the use of wireless sensor networks based on fuzzy logic. *Int. J. Phys. Sci.* **6**(3), 448–458 (2011)
12. Attea, B.A., Khalil, E.A.: A new evolutionary based routing protocol for clustered heterogeneous wireless sensor network. *Appl. Soft Comput.* **12**, 1950–1957 (2012)

Domain Specific Search Engine Based on Semantic Web

Shruti Kohli and Sonam Arora

Abstract Currently search engines are using keyword based approach which becomes problematic when the user is not aware about the way to write query such that desirable results only appear because for that he must know the semantic concepts that are used in that particular domain in which he is interested. To overcome this problem, we introduced the concept of Semantic Web which is an extension of the current Web that allows the meaning of information to be precisely described in terms of well-defined vocabularies that are understood by people and computers. Ontology is one of the fundamental ingredients used in the semantic web infrastructure. This paper focuses on the problem of internet users who most of the time looks for the good hotels available beforehand whenever they plan any trip. At times they go for image search option to have a view of hotels in a particular location. But problem is user gets confused whether search results displayed online by currently available search engines are reliable or not. This paper aims to implement a tool which is based on Semantic Web for searching the images of hotels and displaying the results. Semantic web refines the search in such a way that only relevant images are returned. The crux is to develop ontology for various hotels along with their locations and the facilities they provide. User enters his query from an interface and corresponding SPARQL query is generated. This query with the help of JENA API is searched in the data present in Ontology.

Keywords SPARQL · RDF · OWL · Semantic Web

S. Kohli (✉)

Department of Computer Science, BIT, Noida, Uttar Pradesh, India
e-mail: kohli.shruti@gmail.com

S. Arora

Department of Computer Science, ABES Engineering College, Ghaziabad, Uttar Pradesh, India
e-mail: sonam.2206@gmail.com

1 Introduction

Nowadays, most of the hotel information is available on the search engine databases online. Tourist will enter keywords in the search engines and will be flooded with the image results. Because of the inherent weakness of internet about precision and recall, many of these images may not be reliable.

The solution is with Semantic Web. The Semantic Web works by extending the current Web with semantics that provides well-defined information, thus enabling machine learning such that more precise search results are returned, bringing together data from different sources, and automating human tasks. On the Semantic Web, information is represented using a new W3C standard called the Resource Description Framework (RDF). Semantic Web Search is a search engine for the Semantic Web. Ontology is one of the most important concepts used in the semantic web infrastructure, and RDF(S) (Resource Description Framework/Schema) and OWL (Web Ontology Languages) are two W3C recommended data representation models which are used to represent ontology. Semantic Web will surely support efficient information learning process and data reusability which was not found in current web search engines. Currently research on semantic web search engines are in the beginning stage, as the traditional search mechanisms used in search engines such as Google, Yahoo, and Bing (MSN), still dominate the present markets of search engines. First, they do not provide the factor of reliability to the extent user demands. For example, when a any user wants to query that “give me the images of all Taj Mahal” the search engine although provide thousands of result to the user which may be of monument Taj Mahal in Agra, India or Taj Mahal casino in USA. So it’s difficult for the user to find out which source is reliable. The user has to sift through all the retrieved images to find only the reliable results. Secondly, the relevancy of provided results is not up to the mark.

The main intent of the Semantic Web is to give machines much better access to information resources so they can be information intermediaries in support of humans. Information retrieval in current services relies only on keyword searches using Google or based on simple metadata such as that of an RSS. Moreover, there is no function to generate personalized searches easily, so users need to consider and enter search keywords that suit their own requirements correspondingly. Such a process of searching is time consuming and requires lot of effort on human part. That means if users does not understand the keywords to be used for searching, then he can’t perform reliable information retrieval even if users might become interested in a topic. He may get some undesirable results by the search engine. To overcome some of these problems an alternative approach in web searching is given in this paper, wherein the user need not think of appropriate keyword that might give them the result they want, instead the user can simply provide the search engine with whatever information it has by selecting domain. The way search engine accepts user query is very user friendly and easy to understand both by the user as well as by the machine.

As domain specific search is a search interface paradigm based on a long running library tradition of faceted classification. And efficient search systems have proved the paradigm both powerful and intuitive for end—users, particularly in drafting complex queries. Thus, Domain—specific search presents a promising direction for semantic search interface design, if it can be successfully combined with Semantic Web Technologies. Domain based search engines will use semantic web documents (RDF, XML, OWL). With these documents one can describe virtually anything.

The core idea of domain based search engine is to describe query in the form of domain description. For example, a user query is “images of all hotels in Agra”.

For this user need to build a query type which Semantic Web understands. In this paper we propose that queries are not build using natural langue, but an easy to use user interface that help users to build the queries they want.

2 Problem Definition or Formulation

Keyword based search is useful especially to a user who knows what keywords are used to index the images as they can easily define queries. This approach becomes problematic when the user is not aware about the way to write query such that desirable results only appear because for that he must know the semantic concepts that are used in that particular domain in which he is interested. Experts have developed various technologies for refined search but unfortunately till now irrelevant links are got.

Consider Fig. 1.

- *Precision: Percentage of returned pages that is relevant.* Or in other words the capability of minimizing the number of irrelevant links returned to the users.
- *Recall: Percentage of relevant pages that is returned.* Or in other words the capability of maximizing the number of relevant links returned to the users.

All search mechanism till date performs the function where precision and recall percentage is too low.

Domain searching using Semantic Web aims to provide better precise and recall rates as compared to keyword based search. User need not think of appropriate keyword that might give them the result they want, instead the user can simply provide the search engine with whatever information it has by selecting options.

For example, consider a situation when User enters “images of all hotels in Delhi” query on Google, now search results may contain lots of unwanted image result which are of no interest to the user (Fig. 2).

The challenge is to create a topic based semantic web search engine which is highly user friendly and provide advance search options with the help of topics. A user shouldn't be aware of the concepts supporting the semantic web to use it. Their experience should be as close as the one they currently have with the current web and the search engines they use daily.

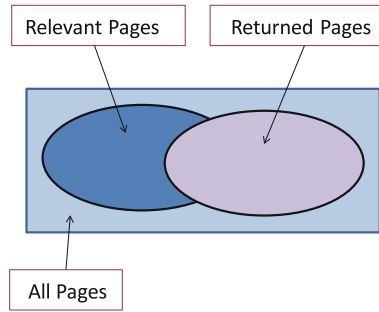


Fig. 1 Precision versus recall

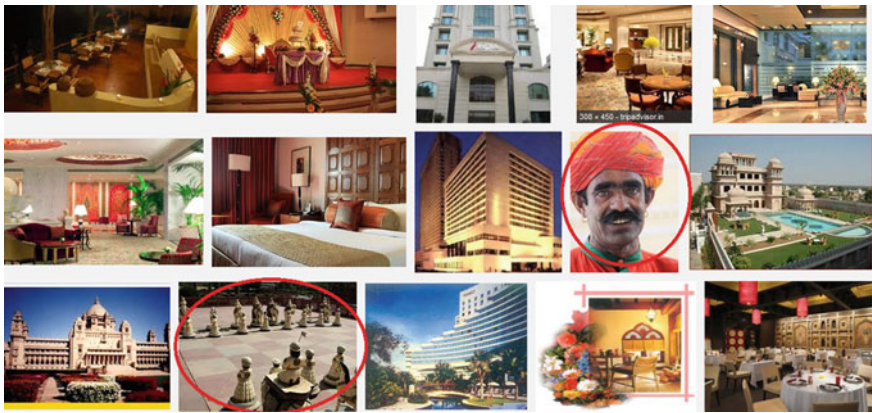


Fig. 2 Google images results

Some of the latest works relating semantic areas are:

Yong-Gui and Zhen [1] gave introduction on Web Mining and Semantic Web-related knowledge and then integrated both of them to improve the effectiveness of Web Mining. They claim that knowledge of Semantic Web makes Web Mining easier. They gave a 5 step process under the framework of Agent for integrating them.

Huiping [2] proposed a semantic web search model to enhance efficiency and accuracy of IR for unstructured and semi structured documents. He used Ranking Evaluator to measure the similarity between documents with semantic information for rapid and correct information retrieval. He introduced Search Arbiter to judge whether the query is answered by Keyword based Search Engine or Ontology Search Engine. He gave just a conceptual architecture of Semantic Web Information Retrieval System.

Kamran and Crestani [3] proposed a method for building a reliable ontology around specific concepts by using the immense potential of active volunteering

collaboration of detected knowledgeable users on social media. They defined a seven step model for creating the semantic relationships between related concepts by using user's input from reliable well known social networking sites. They used automatic information retrieval methods called Wikipedia Link Based Measure (WLM) and Cosine Similarity for computing semantic relatedness score. This model is under construction and there is not any evaluation available yet.

Hyvönen et al. [4] in their paper considered the situation when a user is faced with an image repository whose content is complicated and semantically unknown to some extent. Here ontologies can be of help to the user in formulating the information need, the query and the answers. This paper has used the ontology of promotion event to annotate the promotion event images. They have proposed the in depth structure of Promotion event ontology and the interface developed for handling users query.

Ahmad and Faisal [5] proposed context based search of Personality Images, which helps user to find out required images efficiently. Images were gathered from Google, selecting 300 pictures of 5 personalities which were manually annotated. Main context for finding images related to personalities include activities like playing, attending meeting etc.

Hasany and Selamat [6] presented a system that populates hotel related information in the ontology and use a natural language query platform to retrieve the information from a common interface for decision making. User will get information from a single interface based on user selected parameters. This paper provides the detailed construction of the Hotel Ontology using knowledge base of Malaysian hotels.

Cao et al. [7] presented an ontology based approach for developing STAAR (Semantic Tourist Information Access and Recommending). It helps tourist search information by providing a various semantic search feature in a mobile phone application using web services. In addition they also proposed an algorithm for recommending travel route relevant to both criterions: itinerary length and user interest.

Palaniammal et al. [8] in their paper are concerned with the development of the model towards the semantic search and the result which is based on user's priority while searching the tourism domain of interest. Their system makes use of user's profile ex age, user's current status etc. to understand his behavior and tells the probability of his highest interest on the type of places he/she wishes to tour.

Alice et al. [9] in their paper proposed a tool enhancing a refined search retrieving only the most relevant links eliminating the other links using semantic web technologies. A user's searched text is stemmed and compared with attributes defined in ontology.

Bast et al. [10] in their paper discussed the advantages and shortcomings of full-text search on the one hand and search in ontologies on the other hand. They say that full-text query work well when relevant documents contain the keywords or simple variations of them in a prominent way. Entity oriented queries work well in ontologies. This paper also discussed the challenges while obtaining the facts for the ontologies.

3 Goal of Research

The goal proposed in this research work is to develop system architecture for semantic web search engine for images over a specific domain that is Hotels.

Search engine will be trained with the knowledge base in the form of Ontology which can be in any specification like RDF, OWL or Turtle.

Various classes of Hotels are prepared on the basis of Location, Ratings, and Rate etc. User will be provided with easy to use interface to enter his query. Searching will be done in the ontology itself after which images of the relevant Hotels will be displayed based on their ranking, thus giving a self learning framework to user similar to that of Google image search.

4 Current System Issues

The current search system based on semantic web does not takes into consideration the user behaviour while displaying the search results, which is the important requirement in order to make user aware of the most popular as well as reliable hotels images among Internet users. So this semantic search system focus on using the image popularity as well as the rank of image's host Web Site registered with Google.

5 Proposed System

Here in this proposed system we design and develop a semantic web architecture that can relieve the users from the overburden of doing a lot of keyword based search before getting the desired result.

Data in the ontology is in Turtle format. In this system, latest data about hotel gets populated inside ontology using Google Ajax API. It is basically Google JavaScript API which helps in loading the online search results which includes metadata as well as images directly into the web application.

This system takes the user query in the form of parameters related to that domain in a user friendly environment, develop a SPARQL query and using JENA API will give the reliable results to the user.

On clicking over the image user is redirected to the corresponding host website to which this image belongs. This internally updates the user's preference in the knowledge base. Next time in the results, images will be displayed in the order of user's preferences.

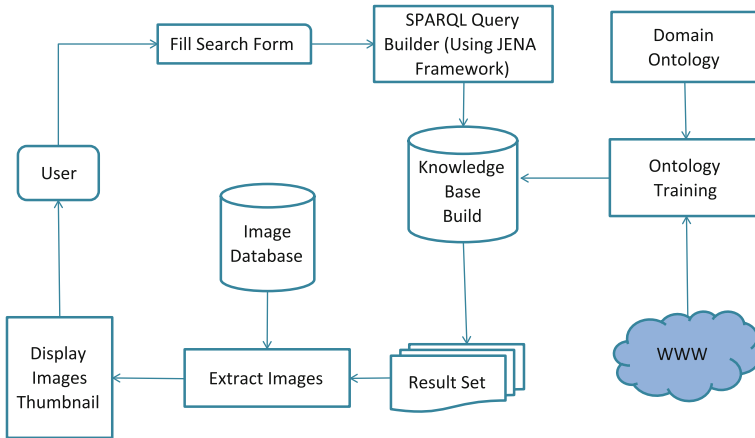


Fig. 3 Architecture

6 System Architecture

See Fig. 3.

7 Training the System

The data of various users is collected and organized around ontology of hotel. The system has a large database of images belonging to various categories. The process of collecting latest data for hotel in the ontology is done automatically using Google API.

All the details along with the URL of image file and its category is stored in the ontology. The category of an image is identified manually and it can be anything like Location, Ratings, and Rate etc.

8 Conclusion

In this research, a model is proposed which will solve the problem of irrelevancy on the search results displayed by current image search engines using the semantic web technologies focusing on single domain that is of Hotels. This model can be extended for multiple domain searches.

References

1. Yong-Gui, W., Zhen, J.: Research on semantic web mining. In: 2010 International Conference on Computer Design And Applications (ICCCA 2010), 978-1-4244-7164-5, IEEE (2010)
2. Huiping, J.: Information retrieval and semantic web, 2010 Technology (ICEIT 2010), 78-1-4244-8035-7/10 2010 IEEE V3-461
3. Kamran, S., Crestani, F.: Defining ontology by using users collaboration on social media, CSCW 2011, Hangzhou, China. ACM 978-1-4503-0556-3/11/03, 19–23 March 2011
4. Hyvönen, E., Styrman, A., Saarela, S.: Ontology based image retrieval, University of Helsinki, Department of Computer Science
5. Ahmad, W., Faisal, C.M.S.: Context based image search, 978-1-4577-0657-8/11/\$26.00 © 2011 IEEE
6. Hasany, N., Selamat, M.H.: Answering user queries from hotel ontology for decision making, 978-1-4577-1481-8/11/\$26.00 ©2011 IEEE 123
7. Cao, T.-D., Phan, T.-H., Nguyen, A.-D.: An ontology based approach to data representation and information search in smart tourist guide system, 978-0-7695-4567-7/11 \$26.00 © 2011 IEEE. doi:[10.1109/KSE.2011.33](https://doi.org/10.1109/KSE.2011.33)
8. Palaniammal, K. Indra Devi, M., Vijayalakshmi, S.: An unfrangled approach to semantic search for E-tourism domain, 978-1-4673-1601-9/12/\$31.00 ©2012 IEEE 130 ICRTIT-2012
9. Alice, P.S., Abirami, A.M., Askarunisa, A.: A semantic based approach to organize eLearning through efficient information retrieval for interview preparation, 978-1-4673-1601-9/12/\$31.00 ©2012 IEEE ICRTIT-2012
10. Bast, H., Baurle, F., Buchhold, B., Haussmann, E.: A case for semantic full-text search, JIWES'12 August 12 2012, Portland, OR, USA Copyright 2012 ACM 978-1-4503-1601-9/12/08...\$15.00
11. <http://jena.apache.org/documentation/ontology/>

1-Error Linear Complexity Test for Binary Sequences

Hetal G. Borisagar, Prasanna R. Mishra and Navneet Gaba

Abstract This paper presents 1-Error Linear Complexity Test (1-ELCT) which is based on Linear Complexity Test—LCT described in (Rukhin et al., NIST Special Publication 800–822, 2001). 1-ELCT is improved version of Bit Flipping Linear Complexity Test (BFLCT). In BFLCT, it is checked that whether the sequence remains random or not after flipping one bit with respect to the LCT. 1-ELCT is for block length of the form $M = 2^q$, $q \in \mathbb{N}$ & $q > 8$ and it is of practical use for binary sequences of length 10^6 .

Keywords Linear complexity · k-error linear complexity · Linear complexity test

1 Introduction

For a periodic binary sequence linear Complexity is defined to be the length of the shortest linear recurrence which generates the sequence. Binary sequences with low linear complexity are cryptographically weak because the entire sequence can be efficiently computed given knowledge of a few initial bits using the Berlekamp-Massey algorithm [1]. It is also seen that some sequences are having high linear complexity but if few bits in the sequence are flipped the linear complexity decreases drastically. If such thing happens, a very close approximation to the

H. G. Borisagar (✉)
DRDO, Delhi, India
e-mail: hetalunique@gmail.com

P. R. Mishra
SAG, Delhi, India
e-mail: prasanna.r.mishra@gmail.com

N. Gaba
DRDO Ministry of Defence, New Delhi, India
e-mail: navneetgaba2000@yahoo.com

sequence can be obtained from a linear recurrence relation. This led to the introduction of a new complexity measure named k -error linear complexity.

The concept of k -error Linear Complexity for periodic binary sequences of period 2^n , was introduced by Stamp and Martin in 1993 [2]. He adopted the approach followed by Games and Chan [3] in their algorithm for finding Linear Complexity for period 2^n . The Stamp-Martin Algorithm was further extended to compute the whole k -error linear complexity profile by Lauder and Paterson [4]. Algorithms for computing the linear complexity and the k -error linear complexity of a sequence, for periodic sequences which have as period a power of the characteristic of the field have been given by Ding et al. [5], Kaida et al. [6], Kaida [7], Wei [8], Alecu and Salagean [9], Zhou [10] and Salagean [11]. All these algorithms, unlike the Berlekamp-Massey Algorithm, need a whole period as input, which means that the whole sequence is already known, which would not be possible in practical applications. As such there is no general algorithm to compute the k -error linear complexity profile of an arbitrary sequence over an arbitrary finite field, other than the exhaustive search. For a stream cipher to be secure, the pseudo-random sequence should have high linear complexity or it must be unpredictable, i.e., it must be computationally infeasible to recover more of the sequence from a short captured segment, i.e. the next captured bit should appear to be drawn from a uniform distribution. The keystream should have uniform characteristics of M -bit blocks ($M = 1$ or 2 or more) or in other words the captured sequence should satisfy randomness properties. That is, randomness reflects the impossibility of predicting next bit from all the previous bits of the sequence. NIST publication [12], defines a randomness test based on the concept of linear complexity for binary sequences. We have used this concept to check that the given sequence continues to remain random and unpredictable when k -errors are introduced (k -bits are flipped) in the sequence. Below we give definitions of linear complexity and k -error linear complexity.

Definition 1 *Linear Complexity of binary sequences is the minimum number of stages of an LFSR, which can generate the sequence [2].*

Definition 2 *k -error linear Complexity of the periodic sequence $(s) = (s_0, s_1, \dots, s_{n-1})$ denoted $c_k(s)$, is the smallest linear complexity that can be obtained when any k or fewer of the s_i 's are altered [2].*

The following sections present a test to indicate the change in randomness of the given binary sequence having high linear complexity, when one bit error ($k = 1$) is introduced in the sequence.

2 Linear Complexity Test

As mentioned in the previous section, Linear complexity indicates the unpredictability of a given binary sequence, i.e., given a small segment of a sequence; the next captured bit should appear to be drawn from a uniform distribution. This means that the key stream should be random. In [12] the LCT for randomness

of sequences is described. It is based on the principle that given a binary sequence of length $n(n = (M \times N))$ where M indicates the length of block and N is the number of blocks, the linear complexity of each block should on the average be $M/2$ if the sequence is random.

Steps for Linear Complexity Test are as follows:

Here length of the sequence is $n \geq 10^6 = N \times M$

where, M is length of the block and it is in the range $500 \leq M \leq 5,000$,

N is number of blocks and $N \geq 200$, K is degree of freedom and $K = 6$ and level of significance is 1 %.

Input: Binary sequence of finite length n .

Output: Sequence is random or sequence is non-random.

Test Description:

Step I: Partition the n -bit sequence into N independent blocks of M bits.

Step II: Using the Berlekamp-Massey algorithm, determine the linear complexity L_i each of the N blocks ($i = 1, \dots, N$).

Step III: Assuming that the sequence is random calculate the mean μ by the formula

$$\mu = \frac{M}{2} + \left(\frac{9 + (-1)^{M+1}}{36} \right) - \left(\frac{\left(\frac{M}{3} + \frac{2}{9}\right)}{2^M} \right) \tag{1}$$

Step IV: For each sub string, calculate a value of T_i , where

$$T_i = (-1)^M * (L_i - \mu) + 2/9 \text{ for } i = 0 \text{ to } N - 1 \tag{2}$$

Step V: Record the T_i values for $i = 0$ to $N - 1$ in v_0, \dots, v_6 as follows

- If $-2.5 \leq T_i$ Increment v_0 by one
- If $-2.5 < T_i \leq -1.5$ Increment v_1 by one
- If $-1.5 < T_i \leq -0.5$ Increment v_2 by one
- If $-0.5 < T_i \leq 0.5$ Increment v_3 by one
- If $0.5 < T_i \leq 1.5$ Increment v_4 by one
- If $1.5 < T_i \leq 2.5$, Increment v_5 by one
- If $T_i > 2.5$ Increment v_6 by one

Step VI: Compute

$$\chi^2 = \sum_{i=0}^6 \frac{(v_i - N\pi_i)^2}{N\pi_i} \tag{3}$$

where, $\pi_0 = 0.01047$, $\pi_1 = 0.03125$, $\pi_2 = 0.125$, $\pi_3 = 0.5$, $\pi_4 = 0.25$, $\pi_5 = 0.0625$, $\pi_6 = 0.02078$ are the probabilities.

Step VII: Compute P-value = $\text{igmac}\left(\frac{K}{2}, \frac{\psi^2(\text{obs})}{2}\right)$

Step VIII: If P-value < 0.01 then conclude that sequence is non-random otherwise random.

3 Our Approach

3.1 0-error Linear Complexity Test

We have reconfigured the LCT for $M = 2^q$, $q \in \mathbb{N}$ in which the step IV and V are clubbed by redefining the frequencies with respect to linear complexities directly. This test is named as 0-Error Linear Complexity Test (0-ELCT).

Observe that, the last term in (1) is too small as compared to the first two terms because $M \geq 500$. Therefore, by neglecting this value the mean value μ will be $2^{q-1} + \frac{2}{9}$.

Now in Eq. (2) if we substitute $M = 2^q$ and the value of μ then T_i will be $L_i - 2^q$ $i = 0$ to $N - 1$ and this will be an integer value only. Accordingly we have changed bound on T_i 's $i = 0$ to $N - 1$.

There are total 7 frequencies and for each frequency we have a bound according to [12].

These seven bounds can be calculated using three cases as shown below:

For any r , r_1 and r_2 in \mathbb{R} , $q \in \mathbb{N}$ and for $i = 0$ to $N - 1$ we have

Case 1

$$\begin{aligned} r &\leq T_i : \\ r &\leq T \Rightarrow [r] \leq T_i \\ &\Rightarrow [r] \leq L_i - 2^{q-1} \\ &\Rightarrow 2^{q-1} + [r] \leq L \end{aligned}$$

Case 2

$$\begin{aligned} r_1 &< T_i \leq r_2, r_2 = r_1 + 1 : \\ r_1 &< T_i \leq r_2 \ \& \ r_2 = r_1 + 1 \\ &\Rightarrow [r_1] + 1 \leq T_i \leq [r_2] \\ &\ \& \ [r_2] = [r_1] + 1 \\ &\Rightarrow [r_1] + 1 \leq T_i \leq [r_1] + 1 \\ &\Rightarrow T_i = [r_1] + 1 \\ &\Rightarrow L_i - 2^{q-1} = [r_1] + 1 \\ &\Rightarrow L_i = 2^{q-1} + [r_1] + 1 \end{aligned}$$

Case 3

$$\begin{aligned}
r &< T_i : \\
[r] + 1 &\leq T_i, \\
\Rightarrow [r] + 1 &\leq L_i - 2^{q-1} \\
\Rightarrow 2^{q-1} + [r] + 1 &\leq L_i
\end{aligned}$$

From case 1 and taking $r = -2.5$ the range for v_0 will be

$$v_0 := \text{Number of blocks for which } 2^{q-1} - 3 \leq L_i$$

Similarly using case 2 and case 3 we can redefine the ranges for T_i given in step V of LCT algorithm as follows:

$$\begin{aligned}
v_1 &:= \text{Number of blocks for which } L_i = 2^{q-1} - 2 \\
v_2 &:= \text{Number of blocks for which } L_i = 2^{q-1} - 1 \\
v_3 &:= \text{Number of blocks for which } L_i = 2^{q-1} \\
v_4 &:= \text{Number of blocks for which } L_i = 2^{q-1} + 1 \\
v_5 &:= \text{Number of blocks for which } L_i = 2^{q-1} + 2 \\
v_6 &:= \text{Number of blocks for which } L_i \geq 2^{q-1} + 3
\end{aligned}$$

Steps for 0-error Linear Complexity Test are as follows:

Input: Binary sequence of finite length n .

Output: Sequence is random or sequence is non-random.

Test Description:

Step I: Partition the n -bit sequence into N independent blocks of M bits, where $n = M \times N$.

Step II: Determine the linear L_i of the i th block ($i = 1, \dots, N$) using Berlekamp-Massey algorithm.

Step III: Assuming that the sequence is random calculate mean μ by the formula (1).

Step IV: Calculate frequencies v_0, \dots, v_6 as described above.

Step V: Compute χ^2 value using formula (3).

Step VI: If $\chi^2 \leq 16.812$ then conclude that sequence is random otherwise non-random (level of significance 1 %).

3.2 Bit Flipping Linear Complexity Test

In bit flipping linear complexity test (BFLCT), if a sequence is random with respect to 0-ELCT then 1 bit at a time is flipped in the given sequence and every time 0-ELCT is applied until 0-ELCT fails or 0-ERLCT pass for last bit of the sequence. In short BFLCT checks randomness of the modified sequences obtained by flipping one bit at a time in the original sequence.

Steps for Bit Flipping Linear Complexity Test are as follows:

Input: Binary sequence of finite length n .

Output: Sequence is random or sequence is non random.

Test Description:

- Step I: Apply 0-ELCT on the sequence. If the sequence is non random algorithm terminates here otherwise go to step II.
- Step II: Initialise $i = 0$.
- Step III: Flip i th bit and apply 0-ELCT.
- Step IV: If the sequence does not passes 0-ELCT declare sequence does not passes BFLCT and stop here otherwise set $i = i + 1$ and check if $i = n$. If $i = n$ then declare sequence passes BFLCT otherwise go to Step III.

3.3 1-Error Linear Complexity Test

We state and prove the following theorem which is used to modify BFLCT and give algorithm for 1-Error Linear Complexity Test (1-ELCT). This reduces the computation complexity of BFLCT drastically.

Let us take a sequence of length $N \times M$ where N is number of blocks and M is block size. We have taken block size M is of the form 2^q , $q \in \mathbb{N}$, $q > 8$ (in bits). Denote w_i and $b_j^{(i)}$ as the blocks and its bits respectively where $i = 0, 1, \dots, N - 1$ and $j = 0$ to $2^q - 1$. We define the sets S_l as sets containing indices of blocks corresponding to the frequencies v_l , $l = 0, 1, \dots, 6$ as follows:

$$\begin{aligned}
 S_0 &= \{i | L_i \geq 2^{q-1} - 3\}, & S_1 &= \{i | L_i = 2^{q-1} - 2\}, & S_2 &= \{i | L_i = 2^{q-1} - 1\} \\
 S_3 &= \{i | L_i = 2^{q-1}\}, & S_4 &= \{i | L_i = 2^{q-1} + 1\}, & S_5 &= \{i | L_i = 2^{q-1} + 2\} \\
 S_6 &= \{i | L_i \geq 2^{q-1} + 3\} & & & & \text{here } i = 0 \text{ to } N - 1
 \end{aligned}$$

Observation: If only one bit in the sequence is flipped then change in the frequencies of the given sequence will be $\sum_{i=0}^6 |v_i - v'_i| = 2$ only as one bit change at most changes two frequencies.

Theorem 1 *If a sequence is random (with respect to Linear complexity test) with frequency distribution $v_i, i = 0, 1, \dots, 6$ and χ^2 for all other distributions $v'_i, i = 0, 1, \dots, 6$ satisfying $\sum_{i=0}^6 |v_i - v'_i| = 2$ is greater than χ_e^2 , then sequence is non-random for 1-error linear complexity test.*

Proof If for any 1-bit change, the distribution changes and corresponding χ^2 value is greater than χ_e^2 then it is obvious that the sequence will be 1-error non-random.

Therefore, it suffices to prove that there exists a bit in the sequence whose flipping changes the frequency distribution such that $\chi^2 > \chi_e^2$ then we are through.

Claim: The required bit will be last bit of any block belonging to S_2 .

Observe that, If a sequences is random with respect to 0-ELCT then S_2 cannot be empty set.

If possible suppose $v_2 = 0 \Rightarrow S_2 = \phi$.

The χ^2 is calculated according to (3)

$$\chi^2 = \sum_{i=0}^6 \frac{(v_i - N\pi_i)^2}{N\pi_i}.$$

We try to minimize χ^2 when $v_2 = 0$. The constraint function $g(v_0, v_1, v_3, v_4, v_5, v_6)$ is given as,

$$g(v_0, v_1, v_3, v_4, v_5, v_6) := v_0 + v_1 + v_3 + v_4 + v_5 + v_6 = N$$

The condition for extrema is given as

$$\begin{aligned} \nabla \chi^2 &= \lambda \nabla g \\ \Rightarrow \frac{2(v_i - N\pi_i)}{N\pi_i} &= \lambda, \quad i = 0, 1, 3, 4, 5, 6 \\ \Rightarrow v_i &= N\pi_i \left(\frac{\lambda}{2} + 1 \right) \end{aligned}$$

Putting this value in the constraint function we get,

$$N \left(\frac{\lambda}{2} + 1 \right) (\pi_0 + \pi_1 + \pi_3 + \pi_4 + \pi_5 + \pi_6) = N \Rightarrow \lambda = \frac{2}{7}$$

It can be shown that the extrema condition corresponds to the minimum value. The minimum value thus corresponds to

$$v_i = \begin{cases} \frac{8}{7}N\pi_i, & i = 0, 1, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

and the minimum value of χ^2 is given as

$$\chi_{\min}^2 = \sum_{i=0,1,3,4,5,6} \frac{\left(\frac{8}{7}N\pi_i - N\pi_i\right)^2}{N\pi_i} + \frac{(0 - N\pi_2)^2}{N\pi_2} = \frac{1}{7}N$$

Sequence is random with respect to 0-ELCT so it has number of blocks at least 200. Therefore, $\chi^2 \geq \chi_{\min}^2 = \frac{1}{7}N > 28 > \chi_e^2$ which is contradiction as sequence is random with respect to 0-ELCT means $\chi^2 \leq \chi_e^2$. So our assumption $S_2 = \phi$ is not possible.

Now since S_2 is non-empty, there will be at least one block index in S_2 . Let i_0 be that index. Therefore linear complexity of block w_{i_0} is $2^{q-1} - 1$, $q \in \mathbb{N}$. When Berlekamp -Massey Algorithm is applied to the block w_{i_0} there will be two cases for the linear complexity with respect to the last bit of w_{i_0} :

Case 1 $b_{2^q-1}^{(i_0)}$ is a point of discrepancy.

There are two possibilities for the linear complexity L at this point

1. $L \leq N/2$:

If $L \leq N/2$ then L will be increased as follows:

$$L = N - L + 1 = (2^q) - (2^{q-1} - 1) + 1 = 2^{q-1} + 2$$

which is not possible.

2. $L > N/2$:

In this case the previous value of linear complexity is taken as L which is $2^{q-1} - 1$ but in this case too,

$$L > N/2 \Rightarrow 2^{q-1} - 1 > 2^{q-1}$$

which is again contradiction.

Therefore, for w_{i_0} is in S_2 , $b_{2^q-1}^{(i_0)}$ cannot be a point of discrepancy.

Case 2 $b_{2^q-1}^{(i_0)}$ is not a point of discrepancy.

Here the value of d can be given as,

$$d = 0 = b_{2^q-1}^{(i_0)} + c_1 b_{2^q-2}^{(i_0)} + c_2 b_{2^q-3}^{(i_0)} + \dots + c_{2^q-1-1} b_{2^q-1}^{(i_0)}$$

Now, flipping $b_{2^q-1}^{(i_0)}$ will make it a point of discrepancy i.e., $d = 1$. At this stage the complexity will be

$$N - L + 1 = 2^q - (2^{q-1} - 1) + 1 = 2^{q-1} + 2 > N/2 = 2^{q-1}$$

Therefore if we flip the bit $b_{2^q-1}^{(i_0)}$ then this block w_{i_0} will not remain in S_2 .

So if the last bit of any block of S_2 is flipped then it will change the frequency distribution such that $\sum_{i=0}^6 |v_i - v'_i| = 2$. Hence the theorem is proved.

In BFLCT every time we flip a bit we have to apply Berlekamp Algorithm. The above theorem says if we just change the frequencies and see its effect on the χ^2 value we can directly conclude that the sequence is random or not if it is satisfying the assumption of the theorem. And empirically we have seen that out of 100 sequences there are only for 3–5 sequences which does not satisfies the assumption of the above theorem. So this reduces the computational complexity of bit flipping linear complexity test drastically.

Steps for 1-error Linear Complexity Test are as follows:

Input: Binary sequence of finite length n .

Output: Sequence is random or sequence is non-random.

Test Description:

Step I: Apply 0-error linear complexity test on the sequence. If the sequence is non random algorithm terminates here.

- Step II: Set $i = 0$.
- Step III: Set $j = \min S$, where $S = \{0, 1, 2, 3, 4, 5, 6\} - \{i\}$.
- Step IV: If $v_i \neq 0$ first decrease v_i by one and then check $j = 6$ or not. If $j = 6$ store the indices of blocks corresponding to the frequency v_{i-1} and go to step VI.
- Step V: Increase i by one and go to step IV.
- Step VI: If $j = 7$ or $j = i$ increase v_i and i by one and go to Step III otherwise increase v_j and calculate χ^2 .
- Step VII: If $\chi^2 \leq 16.812$ decrease v_j by one, set $j = j + 1$ and go to Step VI. Otherwise increase v_i and i by one and go to Step III.
- Step VIII: If indices of all the blocks corresponding to all the v_i are stored then declare sequence is random.
- Step IX: If no index of blocks are stored then declare sequence is non-random otherwise go to Step X.
- Step X: Apply bit flipping Linear Complexity Test on remaining blocks whose indices are not stored.

We have applied 1-ELCT on data collected from various random and pseudorandom sources. The parameters selected are $N = 1,000$, $M = 1,024$ and the number of sequences ≤ 100 .

S. N.	Data	No. of binary sequences	Block size (bits) M	Number of blocks N	Rand. seq. (k = 0)	Rand. seq. (k = 1)
1	Salsa	100	1,024	1,000	100	97
2	Rabbit	100	1,024	1,000	100	100
3	LEX	100	1,024	1,000	100	100
4	TRNG data	100	1,024	1,000	95	94

4 Complexities

For a binary sequence of length $n = N \times M$, where $N =$ Number of blocks and $M =$ block size. The complexity for finding 1 error linear complexity is

$$\begin{aligned}
 O(n^2) + O\left(\binom{n}{1} \times n^2\right) &= O(n^2) + O(n \times n^2) = O(n^3) \\
 &= O(N^3 \times M^3) \text{ (by substituting } n = N \times M)
 \end{aligned}$$

This shows that the time complexity for the computation of k-error Linear Complexity for $k = 1$ is quite high even for moderate size sequences e.g., of size 10^6 . Therefore it is not of much practical use. On the other hand the BFLCT provides fairly good information about distribution of 1-error Linear Complexity

in the sequence, in a much lesser time. The average computational complexity of BFLCT is as follows:

$$\begin{aligned} O(N \times M^2) + O\left(\binom{M}{1} \times N/2 \times M^2\right) &= O(N \times M^2) + O(M \times N/2 \times M^2) \\ &\approx O(N \times M^3) \end{aligned}$$

The complexity here reduces by factor of N^2 . Time complexity for 1-error LCT is very small compared to the complexity of BFLCT as shown below: N' = Number of block on which BFLCT is applied $\leq N$.

$$\begin{aligned} O(N \times M^2) + O\left(\binom{M}{1} \times N' \times M^2\right) &= O(N \times M^2) + O(N' \times M \times M^2) \\ &= O((N + N' \times M) \times M^2) \end{aligned}$$

We have carried out experiments for $n = 1,000 \times 1,024$. The time complexity for finding 1-ELCT is $\approx O(1,000^3 \times 1,024^3) \approx 2^{60}$. For the same sequence BFLCT time complexity will be $\approx O(1,000 \times 1,024^3) \approx 2^{40}$ while for the 1-ELCT the time complexity is $\approx 2^{30}$ this is because it has been empirically tested that for most of the sequences $N' = 0$ or N' very small compared to N .

5 Conclusion

LCT proposed in [12] requires the sequence length to be greater than $\geq 10^6$. We have modified this test to measure the randomness when 1 bit of the sequence is flipped and observed that the naive approach to this test has even a greater complexity for sequences of length $\geq 10^6$ bits. We improved the algorithm and illustrated that the complexity for this modified algorithm is in the realistic range. We are further working on extending this test to work for $k = 2$ and different block lengths.

Acknowledgments We would like to thank Dr. P. K. Saxena, Dr. S. S. Bedi and Ms. Neelam Verma for their support. Also we are thankful to Dr. Indivar Gupta for his valuable suggestions.

References

1. Massey, J.L.: Shift register synthesis and BCH decoding. *IEEE Trans. Inf. Theor.* **15**(1), 122–127 (1969)
2. Stamp, M., Martin, C.F.: An algorithm for the k -error linear complexity of binary sequences with period 2^n . *IEEE Trans. Inf. Theor.* **39**(4), 1398–1401 (1993)
3. Games, R.A., Chan, A.H.: A fast algorithm for determining the complexity of a binary sequence with period 2^n . *IEEE Trans. Inf. Theor.* **29**(1), 144–146 (1983)

4. Lauder, G.B., Paterson, K.G.: Computing the error linear complexity spectrum of a binary sequence of period 2^n . *IEEE Trans. Inf. Theor.* **49**(1), 273–2803 (2003)
5. Ding, C., Xiao, G., Shan, W.: *The Stability Theory of Stream Ciphers*. Springer, Heidelberg (1992)
6. Kaida, T., Uehara, S., Imamura, K.: An algorithm for the k-error linear complexity of sequences over $GF(p^m)$ with period pn , p a prime. *Inf. Comput.* **151**, 134–147 (1999) Academic Press
7. Kaida, T.: On the generalized Lauder-Paterson algorithm and profiles of the k-error linear complexity for exponent periodic sequences. In: *Proceedings of SETA 2004, Volume LNCS 3486*, pp. 166–178. Springer, Berlin (2005)
8. Wei, S.: An efficient algorithm for determining the k-error linear complexity of binary sequences with periods $2p^n$. *JCSNS Int. J. Comput. Sci. Netw. Sec.* **8**(4), 221 (2008)
9. Alecu, A., Salagean, A.M.: A genetic algorithm for computing the k-error linear complexity of cryptographic sequences. In: *2007 IEEE Congress on Evolutionary Computation*, pp. 3569–3576 (2007)
10. Zhou, J.: An algorithm for the k-error linear complexity of a sequence with period $2p^n$ over $GF(q)$. In: *IEEE Proceedings of IWSDA'07*, pp. 104–108 (2007)
11. Salagean, A.: On the computation of the linear complexity and k-error linear complexity of binary sequences with period a power of two. *IEEE Trans. Inf. Theor.* **51**(3), 1145–1150 (2005)
12. Rukhin, A., et al.: A statistical test suite for random and psuedo random number generators for cryptographic applications. *NIST Spec. Publ.* 800–822 (with revisions dated 15 May 2001)

Neural Method for Site-Specific Yield Prediction

Pramod Kumar Meena, Mahesh Kumar Hardaha, Deepak Khare and Arun Mondal

Abstract In the recent years, a variety of mathematical models relating to crop yield have been proposed. A study on Neural Method for Site –Specific Yield Prediction was undertaken for Jabalpur district using Artificial Neural Networks (ANN). The input dataset for crop yield modeling includes weekly rainfall, maximum and minimum temperature and relative humidity (morning, evening) from 1969 to 2010. ANN models were developed in Neural Network Module of MATLAB (7.6 versions, 2008). Model performance has been evaluated in terms of MSE, RMSE and MAE. The basic ANN architecture was optimized in terms of training algorithm, number of neurons in the hidden layer, input variables for training of the model. Twelve algorithms for training the neural network have been evaluated. Performance of the model was evaluated with number of neurons varied from 1 to 25 in the hidden layer. A good correlation was observed between predicted and observed yield ($r = 0.898$ and 0.648).

Keywords Artificial neural network · Crop yield · Site specific

P. K. Meena (✉) · D. Khare · A. Mondal
Department of Water Resources and Management, IIT, Roorkee, India
e-mail: pramodcae@gmail.com

D. Khare
e-mail: kharefwt@gmail.com

A. Mondal
e-mail: arun.iirs@gmail.com

M. K. Hardaha
Department of Soil and Water Engineering, JNKVV, Jabalpur, India
e-mail: mkhardaha@rediffmail.com

1 Introduction

Wheat (*Triticum* spp.) is a major crop of rabi season in India. It is cultivated in 28.52 Mha areas with the total production of about 80.71 million tons. The area, production and productivity of wheat in the states of Madhya Pradesh are 1.815 (Mha), 7.2796 MT, and 18.35 q/ha respectively [1]. The yield of this crop is very sensitive to temperature variation when soil moisture is less during the critical stages, which affect the productivity of wheat.

Yield patterns in fields may change annually, due to spatial variations in soil properties and weather. Climatic factors like temperature, solar radiation and rainfall affect crop yield. Changes in climatic variables like rise in temperature and decline in rainfall have been reported by Intergovernmental Panel on Climate Change [2]. Pre and post-anthesis high temperature and heat had massive impacts on wheat growth, whereas stress reduced its photosynthetic efficiency [3]. You et al. [4] observed a significant reduction in yield caused by a rise in temperature of 1.8 °C caused 3–10 % reduction in wheat yield. A few days of temperature above a threshold value, if coincident with anthesis, can significantly reduce yield by affecting subsequent reproductive processes [5].

A variety of nonlinear techniques for investigating yield response have also been examined, including boundary line analysis [6], state-space analysis Bayesian networks, and regression trees [7]. However, many nonlinear methods can be difficult to implement, and comparison of the results from these vastly different methods is problematic. Clearly, nonlinear methods that are relatively easy to implement and can be readily compared to one another would be highly desirable. A relatively new branch of nonlinear techniques, artificial neural networks (ANN or NN), has been applied not only to artificial intelligence [8] and classification applications [9] but also as general, non-parametric ‘regression’ tools.

Present study has been carried out to Estimate the yield of wheat crop in Jabalpur district based on climatic factors i.e. temperature, rainfall and relative humidity. The objective of the study is to develop an Artificial Neural Network model for wheat productivity for the Jabalpur district.

2 Study Area

Present study was carried out at Jabalpur district in Madhya Pradesh, India. It is located at latitude from 23009'36''N to 23037'N and longitude from 79057'E to 79095'E at the average MSL of 408 m within the Agro-Climatic zone of Kymore Plateau and Satpura hills. Jabalpur district has a humid subtropical climate, typical of North-Central India.

2.1 Collection of Data

Crop yield: Yield data of wheat and paddy crop of Jabalpur district was collected from the Department of Agricultural Economics and Farm Management, JNKVV, Jabalpur for the year 1969–2010. Yield data from 1969 to 1998 includes the yield of Katni district at present. The productivity of these crops has been computed from yield and area under the specific crop of the district.

Rainfall, temperature and humidity: The weekly data of rainfall, maximum and minimum temperature and relative humidity (morning, evening) of Jabalpur were collected from The Department of Physics and Agro-meteorology, College of Agricultural Engineering, JNKVV Jabalpur for the years 1969–2010.

2.2 Randomization of Data

Out of 41 years of data considered for the analysis, initially 25 years of data have been used for model development and rest for validation of the model. Due to separation of Katni district, from the Jabalpur in the year 1999, large variations in the productivity have been observed. In order to reduce the temporal effect on productivity, the total 41 years data have been randomized for the purpose of development and validation of the model.

2.3 Predictor Variables for Wheat

There are large variations in the duration of wheat crop in the district. To generalize the model, it is assumed that the wheat crop has an average crop period of twenty-two weeks that is from 44th Standard Metrological Week (SMW) to 13th SMW. The total number of predictor variable for this period becomes 110, hence to reduce the number of predictor variables model was developed with Principle Component Analysis (PCA). The predictor variables selected by PCA wheat crop and the model criteria set for (set of suitable input parameters based on correlation analysis between wheat yield and the selected input parameter [10]) is shown in Table 1.

3 Materials and Methods

The functional diagram of an artificial neuron is shown in Fig. 1. There are weighted input connections to the artificial neurons. These input signals get added up, and are fed into the activation function. The reaction signals of the neuron

Table 1 ANN Models with varying input variable Model Predictors Criteria (PCA)

Model	Predictors	Criteria (PCA)
WM-20	T47 M47 M48 M51 M13 R47 R48 R50 T49 M6 R44 R46 R8 RHM2 RHM5 RHM6 RHE44 RHE45 RHE46 RHE11	If r value is greater than 0.2
WM-11	T47 M47 M48 M13 R48 T49 R44 R46 RHM6 RHE44 RHE45	If r value is greater than 0.25
WM-4	M47 M48 M13 R44 R48 RHM6	If r value is greater than 0.3

Where, Tn = Average maximum temperature of nth week (°C)

Mn = Average minimum temperature of nth week (°C)

Rn = Rainfall during nth week (mm)

RHMn = Relative humidity (morning) nth week (%)

RHEn = Relative humidity (evening) nth week (%)

would then pass through a transfer function, which decided the strength of the out signal [11, 12]. Finally, the output signal is send through all the output connections to other neurons.

$$y_j = \int \{W_j \times X_j\} - \theta_j \tag{1}$$

The function f(x) is called as an activation function, the activation function enable a network to map any non-linear process. The most commonly used function is the sigmoidal function expressed as:

$$f(x) = \frac{1}{1 + e^{(-x)}} \tag{2}$$

The variables were selected according to the model WM-20, WM-11, WM-8 for developing and evaluating the ANN models. The ANN model architecture is a single layer feed forward network, which is one of the simplest neural network and has been successfully used the prediction of the nonlinear process [13, 14]. The number of hidden layer is one. The transfer function from input to hidden layer is Tan-Sigmoid Transfer Function (Tansig) and from hidden layer to output layer is Linear Transfer function (Purelin). The Back propagation training function has been selected, which is the most common and accurate as reported by many workers. The performance function for training and testing of networks used are Mean Squared Error (MSE). The various combinations of hidden nodes and training function were done to arrive at optimum combinations to give less error [15]. The network iterations (Epochs) were kept at 500. Architecture for ANN models shown in Table 2. The neural network utility file is edited in MATLAB (7.6 Version). The input variable selection, input data source file, network option, training function, setting for the data for training, validation, plotting the predicting values and saving the network is created and run in MATLAB software.

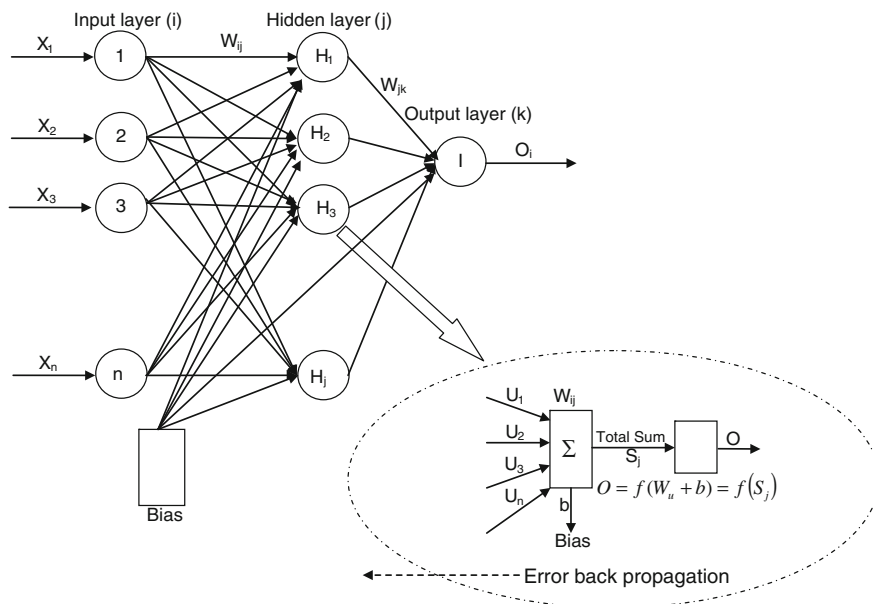


Fig. 1 An artificial neuron showing its function

Table 2 ANN algorithms, architecture and parameters used in ANN model

Algorithm	Back-propagation-L.M
Training functions	Trainlm
Number of Neurons	7
Number of hidden layers	1
Scaling method	Normalization
Activation function of hidden layer	Tan-sigmoid
Activation function of output layer	Purelin
Number of Epochs	500

3.1 Performance Indicators

Correlation coefficient (R), Mean square error (MSE), Root mean square error (RMSE) and Mean Absolute error (MAE), were used as the model development parameters as well as the criteria for evaluation.

4 Result and Discussion

Initial ANN base model has been developed with Levenverg-Marquardt training algorithm with 7 neurons and a single hidden layer. Three models with 20,11,6 predictor variables have been developed.

Table 3 Performance of various wheat ANN models

Model	R		MSE		RMSE		MSE	
	Trg	Val	Trg	Val	Trg	Val	Trg	Val
WM-20	0.87	0.37	2.69E-23	1.71	375	620	297	417
WM-11	0.31	0.48	1.76E-10	1.61	576	719	423	333
WM-6	0.87	0.51	1.76E-10	2.78	576	719	423	394

Training with 60 % of dataset and rest of data set use for validation

Trg = training, Val = validation

Performance of these three models has been summarized in Table 3. It can be observed that value of R during validation of the model WM-6 is much higher than the model WM-20 (0.3656) and the model WM-11(0.4782). However value R during training of the model WM-6 (0.8669) is slightly lower than the model WM-20 (0.8701). RMSE AND MAE for the model WM-6 are higher than the model WM-20 and WM-11, hence model WM-6 has been selected for further refinement.

4.1 Training Algorithms for Wheat Yield Model

The WM-6 model was tested with different training algorithms. For developing the ANN based wheat yield model, performance of 12 training algorithms were evaluated. The model WM-6 was developed using Levenverg Marquardt Algorithm (trainlm).The best training algorithm in the hidden layer of ANN model can be determined by trial and error, at which the model performs best.

Table 4 indicates that the training algorithm “traincgf” resulted in model with highest value of correlation coefficient of 0.871 and 0.511 during training and validation respectively. Model performance indicator; MSE with scaled estimate and the target are lowest at 0.216 and 0.747 during training and validation respectively. RMSE has been worked out as 276.099 and 474.092; and MAE as 204.618 and 357.194 during training and validation respectively.

Hence the ANN wheat model with 6 input variables “mapstd” method of normalization, “traincgf” training algorithm at 7 neurons performed best amongst all twelve algorithms used for training.

4.2 Selection of Optimum Number of Neurons in the Hidden Layer for the Wheat Yield

Increasing the number of neurons in the hidden layer, the network gets an over fit, that is the network have problem to generalize. Trial and error method is applied to determine the optimum number of neurons, at which the network performs best.

Table 4 Performance of different training algorithm methods for ANN based wheat yield modeling

S. No.	Algorithm	R		MSE		RMSE		MAE	
		Trg	Val	Trg	Val	Trg	Val	Trg	Val
1	Traingdx	0.66	0.45	0.55	0.72	421.17	582.19	304.02	428.29
2	Traingd	0.60	0.46	0.71	0.64	494.88	575.68	383.70	332.62
3	Trainscg	0.61	0.51	0.37	0.98	443.00	527.16	341.45	402.46
4	Trainrp	0.46	0.58	0.43	0.70	555.22	692.13	411.44	308.82
5	Trainoss	0.76	0.43	0.25	1.50	366.85	617.04	276.11	519.77
6	Trainlm	0.87	0.51	0.00	2.78	284.74	487.40	225.02	394.29
7	Traincgp	0.78	0.43	0.16	1.24	354.98	519.38	252.37	408.01
8	Traincgf	0.87	0.52	0.22	0.75	276.10	474.09	204.62	357.19
9	Traincgb	0.79	0.49	0.25	0.70	340.53	495.93	250.19	385.58
10	Trainbfg	0.85	0.41	0.12	1.12	293.24	544.38	230.70	389.90
11	Traingdm	0.72	0.37	0.54	0.86	400.63	579.94	311.35	450.09
12	Traingda	0.27	0.55	1.21	0.80	734.11	831.59	576.55	374.83

Trg = training, Val = validation

Table 5 Performance of neural network with different number of neurons of wheat yields ANN modeling

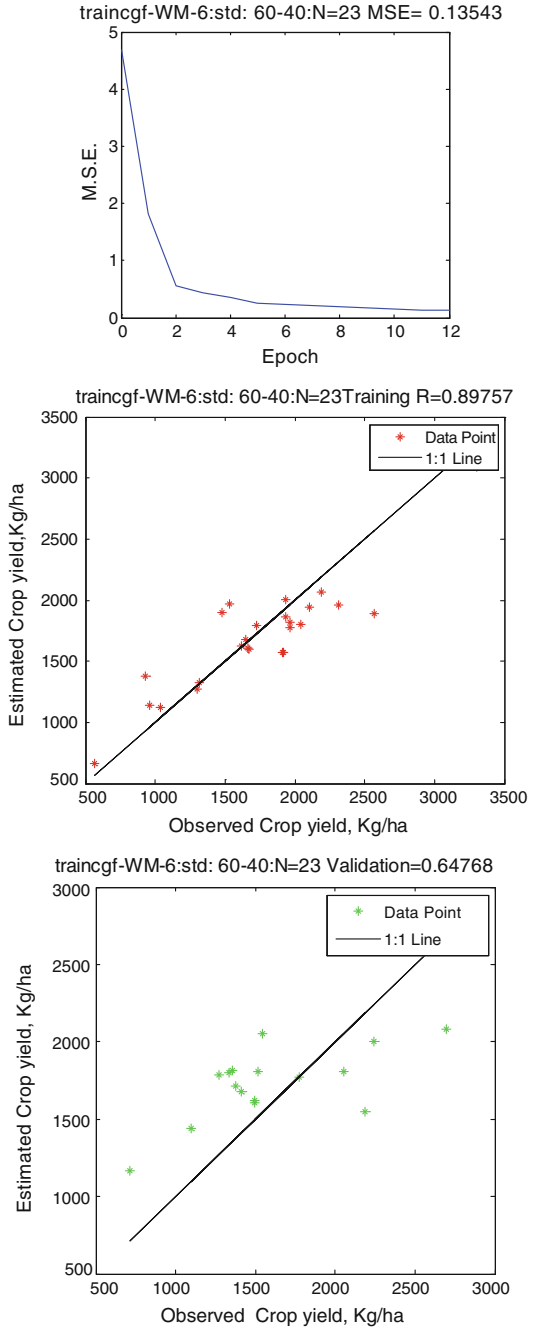
S.No.	Model	R		MSE		RMSE		MAE	
		Trg	Val	Trg	Val	Trg	Val	Trg	Val
1	N1	0.61	0.45	0.61	1.14	450.85	621.76	313.90	453.95
2	N2	0.58	0.48	0.47	1.47	502.79	633.69	371.39	380.97
3	N3	0.41	0.52	0.57	0.98	527.97	699.83	378.88	404.95
4	N5	0.61	0.38	0.29	1.36	486.05	685.67	389.70	402.42
5	N7	0.87	0.52	0.22	0.75	276.10	474.09	204.62	357.19
6	N9	0.80	0.49	0.22	0.76	338.78	490.62	265.58	372.12
7	N11	0.83	0.42	0.25	1.11	315.18	559.36	214.76	459.84
8	N15	0.61	0.25	0.26	1.40	450.49	643.62	357.09	485.02
9	N17	0.40	0.46	0.29	1.20	592.64	568.81	469.57	425.80
10	N19	0.47	0.47	0.29	0.90	537.86	645.50	478.41	373.93
11	N20	0.64	0.45	0.28	0.91	434.56	567.38	319.66	392.12
12	N23	0.90	0.65	0.14	0.71	259.07	438.68	196.09	352.53

Trg = training, Val = validation

Selection of optimum number of neurons is an essential part for wheat ANN model development. The model WM-6 with learning function “traincgf” and normalization function “mapstd” trained with 60 % of data has been evaluated for optimum number of neurons. Neurons in the hidden layer are varying from 1 to 25.

The comparison of performance parameters are presented in Table 5, it can be stated that the model W-6 is trained with “traincgf” algorithm, “mapstd” normalization function and 23 neurons have best performance (Fig. 2).

Fig. 2 Performance of WM-6 model with “traincgf” training algorithm at 23 neurons for wheat yield estimation



5 Conclusions

Artificial Neural Network (ANN) model for estimation of crop yield was developed in the present work. The model has one input layer, one hidden layer and one output layer. Method normalizes the data “std” which transforms the data such that the mean is zero and standard deviation is unity. The input dataset for crop yield modeling includes weekly rainfall, maximum and minimum temperature, relative humidity (morning, evening) for 41 years (1969–2010). ANN models were developed in Neural Network Module of MATLAB (7.6 version, 2008). Model performance has been evaluated in terms of R, MSE, RMSE and MAE. The basic ANN architecture was optimized in terms of training algorithm, number of neurons in the hidden layer, input variables for training of the model. Twelve algorithms for training the neural network have been evaluated. Performance of the model was evaluated with number of neurons varying from 1 to 25 in the hidden layer.

From this study following salient points emerged. Highest value of correlation coefficient between the estimated and observed wheat yield was 0.898 and 0.648 during training and validation by ANN model. The ANN wheat model with “traincgf” algorithm, 23 numbers of neurons, 60 and 40 % length of record for training and validation with 6 input variables is found to be the best model for wheat yield estimation.

References

1. <http://www.des.delhigovt.nic.in/>
2. IPCC: Summary for policymakers. In: Parry, M., Canziani, O., Palutik, J., vander linden, J., Handson, C.: Climate change 2007: Impacts adaptation and vulnerability. Contribution of working group to fourth Assessment Report of the Intergovernmental panel on Climate Change II, pp. 7–22. Cambridge University Press, Cambridge (2007)
3. Wang, X., Cai, J., Jiang, D., Liu, F., Dai, T.: Pre-anthesis high-temperature acclimation alleviates damage to the flag leaf caused by post-anthesis heat stress in wheat. *J. Plant Physiol.* **168**(6), 585–593 (2011)
4. You, M.W., Wood, S., Sun, D.: Impact of growing season temperature on wheat productivity in China. *Agric. Forest Meteorol.* **149**, 1009–1014 (2009)
5. Wheeler, T.R., Hong, T.D., Ellis, R.H., Batts, G.R., Morison, J., Hadley, P.: The duration and rate of grain growth, and harvest index of wheat (*Triticum aestivum* L.) in response to temperature and CO₂. *J. Exp. Bot.* **47**, 623–630 (1996)
6. Kitchen, N.R., Sudduth, K.A., Drummond, S.T.: Electrical conductivity as a crop productivity measure for claypan soils. *J. Prod. Agric.* **12**(4), 607–617 (1999)
7. Adams, M.L., Cook, S.E., Caccetta, P.A., Pringle, M.J.: Machine learning methods in site-specific management research: an Australian case study. In: Proceedings of 4th International Conference on Precision Agriculture, pp. 1321–1333 (1999)
8. Rumelhart, D.E., McClelland, J.L.: Parallel Distributed Processing, vol. 1. MIT Press, Boston, Mass (1986)

9. Burks, T.F., Shearer, S.A., Gates, R.S., Donohue, K.D.: Back propagation neural network design and evaluation for classifying weed species using color image texture. *Trans. ASAE* **43**(4), 1029–1037 (2000)
10. Mohanty, S., Jha, M.K., Kumar, A., Panda, D.K.: Comparative evaluation of numerical model and artificial neural network for simulating groundwater flow in Kathajodi-Surua Inter-Basin of Odisha, India. *J. Hydrology*. **495**, 38–51 (2013)
11. Sablani, S.S.: Artificial neural network modeling. In: *Handbook of Food and Bioprocess Modeling Techniques*, p. 375 (2010)
12. Parida, P.K.: Artificial neural network based numerical solution of ordinary differential equations (2012)
13. Maier, H.R., Dandy, G.C.: Neural networks for prediction and forecasting of water resources variables: a review of modeling issue and application. *Environ. Model. Softw.* **15**, 101–124 (2000)
14. ASCE: Artificial neural networks in hydrology. I: Preliminary concepts. *J. Hydrol. Eng. ASCE* **5**(2), 115–123 (2000)
15. Hardaha, M.K., Chouhan, S.S., Ambast, S.K.: Application of artificial neural network in predicting farmers' response to water management decisions on wheat yield. *J. Agric. Eng.* **49**(3), 32–40 (2013)

Computational Intelligence Methods for Military Target Recognition in IR Images

Jai Prakash Singh

Abstract Military target recognition in IR images is a complex area of computing where most of the existing computation methods do not provide satisfactory deployable solutions. Present paper brings out the different aspects of this problem and discusses how computational intelligence can be helpful in solving many of these complex issues. Computational Intelligence itself consists of different disciplines like Evolutionary Computation, Fuzzy Logic and Artificial Neural Networks which further have their own branches. These all combined can become a powerful tool to create an effective military target recognition solution.

Keywords Computational intelligence • Military target recognition • Neural nets • Evolutionary computation • Magician systems • Genetic algorithms • Fuzzy logic

1 Introduction

IR or thermal imaging is an important tool in military war fighting. It helps our forces not only to see enemy clearly in night covertly (without declaring our own position) but also to see the enemy in dust, smoke and hazy conditions. Military war fighting conditions are the most harsh and tough conditions and pose new challenges on both humans and machines. Friendly fire, collateral damages, fire on civilian vehicles, continuous stress on our soldiers are some of the serious problems which emerge due to inefficient target recognition due to humans or present day machines. Need exists to develop machine ability to go beyond human

J. P. Singh (✉)

Defence Technology Vision 2050 Team, Instruments R&D Establishment (DRDO),
Dehradun, Uttarakhand 248001, India
e-mail: jpsingh@irde.drdo.in; jpsingh1972@yahoo.co.in

intelligence and recognize military targets in the most difficult battle conditions and against enemy maneuvers.

This paper is about application of computational Intelligence (CI) in solving this problem. Which of these problems can be solved effectively through which tool of CI is a significant aspect to be discussed. Also, strengths and limitations of different computational intelligence methods have been brought out. These limitations also inspire us to go for hybrid solutions involving more than one computational intelligence method. Other techniques like Magician Systems and Neocognitrons also need mention here.

2 Problems Inherent in IR Imaging

IR radiation has its own characteristics and so the IR images. IR images as on date are monochrome (black and white) though pseudo coloring is being tried. The edges in IR images are not as sharp as in visible range images (though better than MMW or Microwave Images). Rain and fog are other limitations for IR imaging. Evolutionary computing may help in evolving best models for enemy targets so that the enemy weapons can be differentiated from our own ones and also from civilian equipments with high reliability.

3 Problems Due to Enemy Strategies

Enemy tries his best to safeguard his weapons through hiding, camouflaging and deception etc. Hiding behind bushes or hills make partial appearances sometimes and full picture has to be evolved through these partial appearances. Neural Networks can help in this aspect of evolving full picture from partial appearances. IR camouflaging paints and active electronic camouflaging are the technologies of today and near future. Optical cloaking may come in far future. Deception through decoys are presently being used though in future 3D holographic projections may be used for deception. All this brings another set of problems specific to the battlefield environment due to the war fighting strategies of our adversary. This falls in the class of fuzzy multi-valuedness. Suppose, we have improved our imaging system to the extent that we can make out very clearly that there is a clear enemy weapon at a certain place with very high probability. But at the same point, if we receive information from a possibly reliable source that enemy has made all his weapons appear as non-targets for us and decoys appear as targets. In such situations, our entire assessments change. As per this new information, non target like things become targets and target like things become non-targets. More the probability of target previously, more the probability of non-target now. Such situations are typical examples where computational intelligence like soft-computing methods can help especially fuzzy logic based computing.

4 Problems Due to Our Own Strategies

Sometimes our own soldiers' war fighting strategies pose significant problems. For example, our forces may decide to hide fully behind hills. In this situation, our imagers may get only limited opportunity to see the enemy. Only partial appearances of enemy targets may be received on the basis of which full picture has to be evolved.

5 Computational Intelligence

Computational Intelligence (CI) is an offshoot of artificial intelligence in which the emphasis is placed on heuristic algorithms such as fuzzy systems, neural networks and evolutionary computation. The IEEE Computational Intelligence Society uses the tag-line 'Mimicking Nature for Problem Solving' to describe Computational Intelligence, although mimicking nature is not a necessary element. In addition to the three main pillars of CI (fuzzy systems, neural networks and evolutionary computation), Computational Intelligence also encompasses elements of learning, adaptation, heuristic and meta-heuristic optimization, as well as any hybrid methods which use a combination of one or more of these techniques. More recently, emerging areas such as artificial immune systems, swarm intelligence, chaotic systems, and others, have been added to the range of Computational Intelligence techniques. The term 'Soft Computing' is sometimes used almost interchangeably with Computational Intelligence.

Computational Intelligence techniques have been successfully employed in a wide range of application areas, including decision support, generic clustering and classification, consumer electronic devices, stock market and other time-series prediction, combinatorial optimization, medical, biomedical and bioinformatics problems, and many others. Although CI techniques are often inspired by nature, or mimic nature in some way, CI applications are not restricted to solving problems from nature [1].

6 Tools of Computational Intelligence

So, Computational Intelligence can be best defined by the tools it has to solve the real world complex problems. We will discuss these tools one by one here.

7 Evolutionary Computation

This is the foundation of CI. These are machine learning optimization and classification paradigms based on mechanisms of evolution such as biological genetics and natural selection. This includes genetic algorithms, evolutionary programming,

genetic programming, evolution strategies and particle swarm optimization depending on different combinations of selection criteria, crossover and mutation [2]. Role of evolutionary computation is mainly in training the Artificial Neural Networks and forming the optimum rules for fuzzy systems.

8 Artificial Neural Networks (ANNs)

This is an analysis paradigm roughly modeled after the massively parallel structure of the brain. It simulates a highly interconnected parallel structure with many relatively simple processing elements (PEs). Different models of ANNs include Single and Multi layer Perceptrons, Back Propagation Networks, Bidirectional Associative Memories, Hopfield Nets, Kohonen Maps, Learning Vector Quantizers (LVQs), Artificial Resonance Theory (ART) based Networks, Necognitrons etc. Role of ANNs in computational intelligence are classification, pattern matching, pattern completion, noise removal, optimization, control and simulation etc.

9 Fuzzy Logic

While probability theory deals with statistical uncertainty, fuzzy logic deals with non-statistical uncertainty like linguistic uncertainty. Fuzzy logic is one of many logics of “approximate reasoning”. It comprises of different operations on fuzzy sets like equality, containment, complementation, intersection and union etc. Fuzzy sets are real-valued functions mapping a collection of objects, or the points within the dynamic range of a variable, to values between 0 and 1. The value associated with each object or point is the degree to which that object or point belongs to the fuzzy set. An important capability of fuzzy logic is that it can handle “X is non-X” type Cretian paradoxes and multivalued logics. Role of fuzzy logic is in fuzzy control systems, handling multivaluedness in target recognition systems, fuzzy image processing etc.

10 Soft Computing and its Relation with Computational Intelligence

Soft Computing could be defined as a computational approach to solve problems under circumstances of uncertainty and/or imprecision, being those circumstances either inherent to the problem (the available information is imprecise) or “added” as a way to cope with (to soften) complexity. It is quite important to realize that the presence of uncertainty or imprecision is not a target for SC. It derives from properties or circumstances of the problem under consideration, or is added as a

need to overcome the complexity of the problem. It is not easy to find (either from definitions or use) any significant difference among the characteristics, properties and components of Soft Computing and those of Computational Intelligence. So, we can say that both terms apparently refer to the same, or at least a quite related, concept [3].

11 Applying Computational Intelligence/Soft Computing Tools to Military Target Recognition in IR Images

After discussing different tools of computational intelligence, we can now discuss how these tools can be applied to military target recognition in IR images. Artificial Neural Networks (ANNs) can help in classifying objects appearing in IR images in different classes based on its training. ANNs can also help in handling partial occlusions, noisy images, new targets etc. Fuzzy logic can help in tackling tricky situations like deception and camouflage. Fuzzy control can also help in searching, tracking and pointing. Human Experts' knowledge can be coded better as Fuzzy rules. Evolutionary computation can make both neural and fuzzy systems more accurate and effective in handling target recognition in fog, rain, camouflaged and deception conditions. Neocognitrons are neural network models based on Hubel and Wiesel's studies on Biological Vision Systems. Modified forms of these models can be very helpful. Similarly, Magician Systems are more general forms of Genetic Algorithms and can be more effective in dealing Target Recognition problems.

12 Conclusions and Future Scope

This paper has brought out some aspects of application of CI/SC for military target recognition IR images and opened up the discussion. Practical application of CI in this area will bring several new results and deployable robust systems.

References

1. The 11th UK Workshop on Computational Intelligence—UKCI 2011, <http://ukci.cs.manchester.ac.uk/intro.html>
2. Eberhart, R., Shi, Y.: Computational Intelligence: Concepts to Implementations. Morgan Kaufmann Publishers, San Mateo (2007). (Elsevier)
3. Magdalena, L.: What is soft computing? Revisiting possible answers. *Int. J. Comput. Intell. Syst.* **3**(2), 148–159 (2010)
4. Goertzel, B.: From Complexity to Creativity: Explorations in Evolutionary, Autopoietic, and Cognitive Dynamics. Springer, Berlin (1997)

5. Perlovsky, L.: Computational intelligence: natural information processing, congress (and also course) on evolutionary computation. In: Proceedings of IEEE Computational Intelligence Society, July 2007
6. Kosko, B.: Neural Networks and Fuzzy Systems, PHI (2000)
7. Haykin, S.: The Neural Networks and Learning Machines, 3rd edn. Pearson Education, New York (2013)
8. Sumathi, S., Surekha, P.: Computational Intelligence Paradigms. CRC Press, Alberta (2010)

Flash Webpage Segmentation Based on Image Perception Using DWT and Morphological Operations

A. Krishna Murthy, K. S. Raghunandan and S. Suresha

Abstract Web page segmentation is an important step for many applications such as Information Retrieval, Noise Removal, Full Text Search, Information Extraction, and Automatic page adaptation and so on can benefit from this structure. Many segmentation methods have been proposed on HTML Web page segmentation whereas Flash Web pages have been omitted because of their less availability. But in recent days, we can see many Flash Web pages taking their appearance. In this paper, we are proposing segmentation method by using image processing techniques after processing Web pages as images, because of their unavailability of semantic structure. We perform the experimental analysis based on ground truth analysis (actual blocks in Web page as per human perception) and obtained the better performance level. We also measure the usefulness of Flash Web page blocks.

Keywords Web page segmentation · Flash web image segmentation · Web blocks · Haar wavelet

1 Introduction

In recent decades, research in the field of Web Data Mining is emerging. For example improving the quality of Web by Analyzing Usability Test, Web Information Extraction, Opinion Mining Browsing on Small Screen Device's (SSD's)

A. Krishna Murthy (✉) · K. S. Raghunandan · S. Suresha
Department of Studies in Computer Science, University of Mysore, Mysore, India
e-mail: krishnarjun.research@gmail.com

K. S. Raghunandan
e-mail: raghu0770@gmail.com

S. Suresha
e-mail: sureshabm@yahoo.com

[1] like mobile, Personal Digital Assistance (PDA) and many others. At present, experiencing the internet on Small Screen terminals like Mobile, PDA etc., is becoming very popular. But the current Web pages intended for Large Screen Device's (LSD's) are not suitable for SSD's. Web page segmentation plays major role to accomplish the above mentioned issue(s) and also it helps to many other applications such as, (1) To improve the accuracy of search results as well as search speed, (2) Web Mining tasks (such as Clustering, Classification etc.), (3) It helps to understand the layout structure of Webpage, (4) It helps to overcome the citation issues and so on.

From past, several attempts have been made on Web page segmentation specifically on HTML based Web pages [1]. Thus, there is a lot of scope on it. As W3 (WWW) consortium stated that, HTML has a lot of drawbacks such as (1) Limited defined tags, (2) Not case sensitive, (3) Semi-structured and (4) Designed for only to display data with limited options. To overcome these drawbacks few technologies have been introduced such as XML, Flash and so on [2].

2 Related Work

DOM tree construction for Web pages, tries to extract the structural information from HTML [3]. However because of flexibility of HTML syntax, DOM might cause mistakes in tree structure. Moreover, DOM tree is initially introduced for presentation in the browser rather than description of the semantic structure of the Web page. For example, even though two nodes in the DOM tree have the same parent, it might not be the case that the two nodes are more semantically related to each other than to other nodes [4]. In 2003, Vision Based Page Segmentation (VIPS) algorithm [4] proposed to extract the semantic structure of a Web page. Semantic structure is a hierarchical structure in which each node will correspond to a block and each node will be assigned a value to indicate degree of coherence based on visual perception. It may not work well and in many cases the weights of visual separators are inaccurately measured, as it does not take into account the document object model (DOM) tree information and when the blocks are not visibly different.

Gestalt Theory [5]: a psychological theory that can explain human's visual perceptive process. The four basic laws, Proximity, Similarity, Closure and Simplicity are drawn from Gestalt Theory and then implemented in a program to simulate how human understand the layout of Web pages. A graph-theoretic approach [6] is introduced based on formulating an appropriate optimization problem on weighted graphs, where the weights capture if two nodes in the DOM tree should be placed together. Liu et al. [7] proposed a novel Web page segmentation algorithm based on finding the Gomory-Hu tree in a planar graph. The algorithm initially distills vision and structure information from a Web page to construct a weighted undirected graph, whose vertices are the leaf nodes of the DOM tree and the edges represent the visible position relationship between

vertices. It then partitions the graph with the Gomory-Hu tree based clustering algorithm. Since the graph is a planar graph, the algorithm is very efficient.

From literature [1] it has been observed that no concrete work has done on Flash Web pages. Hence here we concentrated to work on Flash Web pages for its Segmentation.

3 Working on Existing Systems

In order to study the feasibility of the existing approaches on Flash Web pages, we have created a dataset for Flash URL's manually based on its semantic analysis, as its unavailability and it does not contain any identical/unique extension such as .html, .xml, .aspx and so on. Here we have collected nearly 50 various Flash Web domains containing approximately 200 URL Web pages for the experimentation. In order to study feasibility of existing approaches on SSD's, we have implemented a few existing approaches that predominantly deal with segmentation for displaying Web contents on SSD's. In our prior work [8], we have described the working of two such popular approaches namely VIPS and Boiler pipe systems on our Datasets. Experimental results clearly say that existing methods fail to segment Flash Web pages as blocks. We failed to extract the semantic structure of Flash Web pages as its semantic structure is in movie file. Therefore we have planned to work on Web page images; here we first process the Flash Web pages as images then apply the Image Processing concepts to identify blocks.

4 Proposed Segmentation Methodologies

4.1 Segmentation Based on Haar DWT

Web page image is converted into gray scale level and then it is resized into 512×512 to preserve the common size of input images as its dimension varies one Web page to another Web page. Then 2-D Discrete Wavelet Transformation (2-D DWT) processed on gray level images. Here Haar 2D-DWT 2-Level decomposition has been used to extract the content regions. The operation speed of Haar discrete wavelet transform (DWT) operates the fastest among all wavelets because its coefficients are either 1 or -1 . This wavelet technique is very simpler than any other wavelets. This is one of the reasons we employ Haar DWT to detect edges of candidate regions [9]. It decomposes image into four (1 average and 3 detailed) different components such as average, horizontal edge, vertical edge and diagonal edge. Figure 1 shows the decomposed components of input image. After this, strong edges on Horizontal Edge, Vertical Edge and Diagonal Edge are computed using Robert's edge detection (other standard edge detection methods

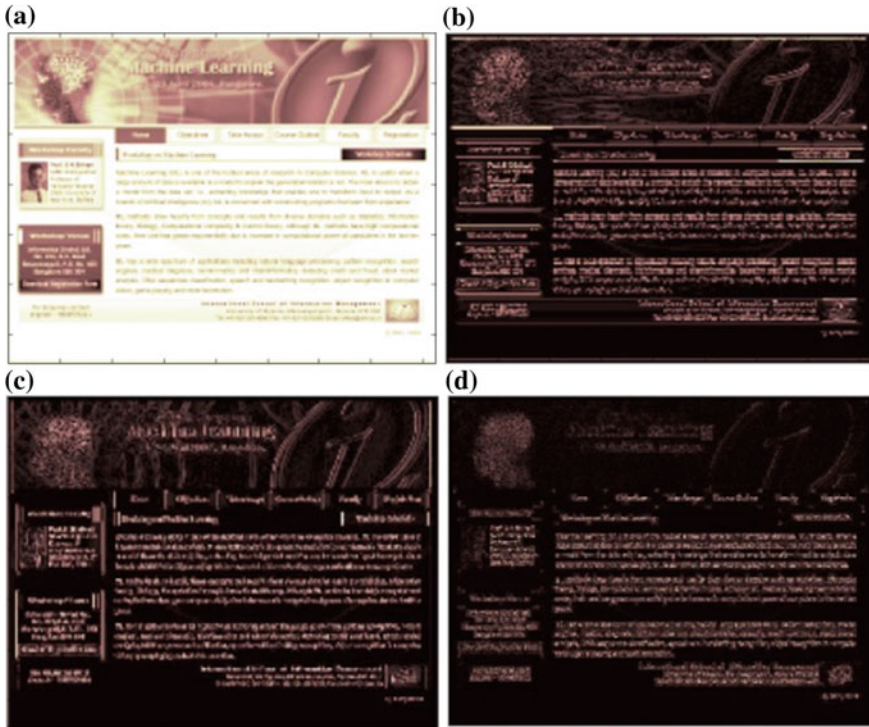


Fig. 1 a Average edge. b Horizontal edge. c Vertical edge. d Diagonal edge

are failed) and Fused these three edge detected images to obtain the one strong edge image. An average component looks very similar to input image, therefore we omitted it. Figure 2 gives edge detected images of Haar DWT components and Fused image (Fig. 2d) after Robert’s edge detection on each component.

Then a morphological dilation (structuring element 3×3) and close (structuring element 2×2) operations are performed on the processed edge map as per Eq. (1) (Fig. 3a). Closing perform the smooth section of contours, Fuses narrow break and long thin gulfs, Eliminates small holes and fill gaps in the contour.

$$(I \oplus S1) \cdot S2 \tag{1}$$

where, I is fused image, S1 is 3×3 structuring element and S2 is structuring element of size 2×2 . Followed by this, connected components are labelled. Features such as Area, Bounding Box are extracted for each labelled components to reduce the error and identify the segmented blocks. Based on Area, very small rectangle block(s) are removed which are considered as outliers. Detected blocks are highlighted (Fig. 3b) by using the bounding box values [10].

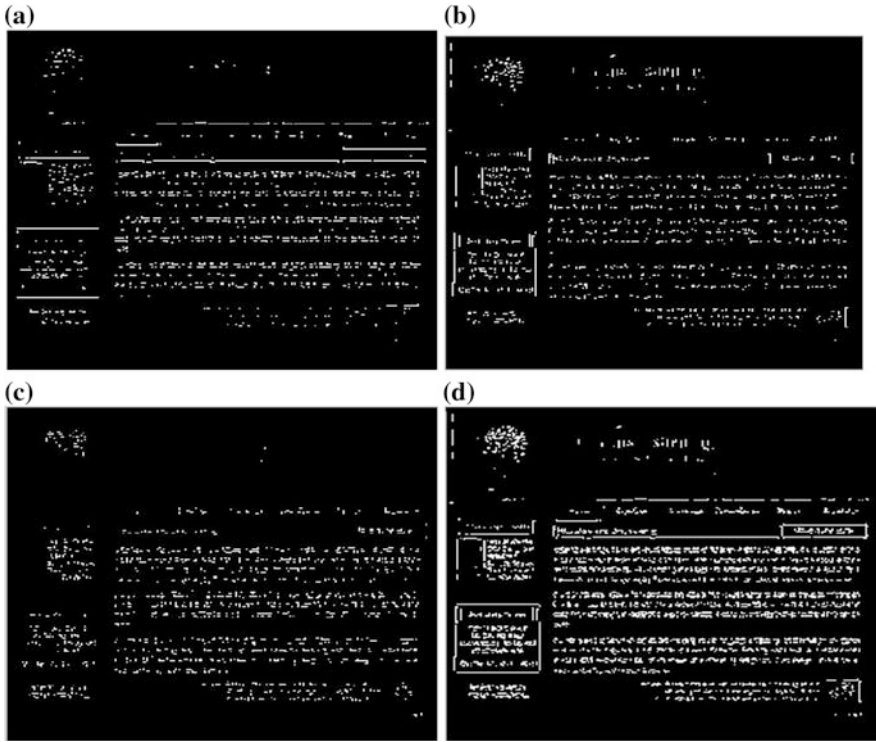


Fig. 2 a Edge detected horizontal sub image. b Edge detected vertical sub image. c Edge detected diagonal sub image. d Fused images



Fig. 3 a Morphological image. b Flash web page segmented as blocks

In Fig. 3b besides text lines superimposed on simple background are detected and images embedded in complex background are also detected with less error rate.



Fig. 4 a RGB image. b Threshold image

4.2 Segmentation Based on Connected Component

In this method before processing input Web page images into binary, we fix the threshold value based on the intensity value to retain the foreground information’s Fig. 4a. After retaining foreground information’s, images are converted into binary.

Then the morphological operations are performed on binary image (Fig. 5a) such as dilation (Fig. 5b) and closing to group the nearest neighbours (Fig. 5c) as stated in Eq. (2). Followed by this, connected components are labelled.

$$(I \oplus S1) \oplus S2) \cdot S3 \tag{2}$$

where, I is Binary image, S1 and S2 are structuring elements of size 4 × 4 and 1 × 3 respectively and S3 is structuring element of size 4. Followed by this, connected components are labelled. For each label features such as Area and Bounding Box are extracted.

Here by using Area feature, very small rectangle block(s) are removed to reduce the error rate. Detected blocks are highlighted (Fig. 3b) by using the bounding box values [10]. In Fig. 5d besides text lines superimposed on simple background are detected and images embedded in complex background are also detected with less error rate.

5 Experimental Results

5.1 Performance Analysis of Proposed Methods

To evaluate the performance of our algorithms, we selected 50 Flash Web pages as images from popular Flash Web sites. Proposed algorithms are run on these set of Flash Web images and results assessed by five individuals are tabulated as follows.

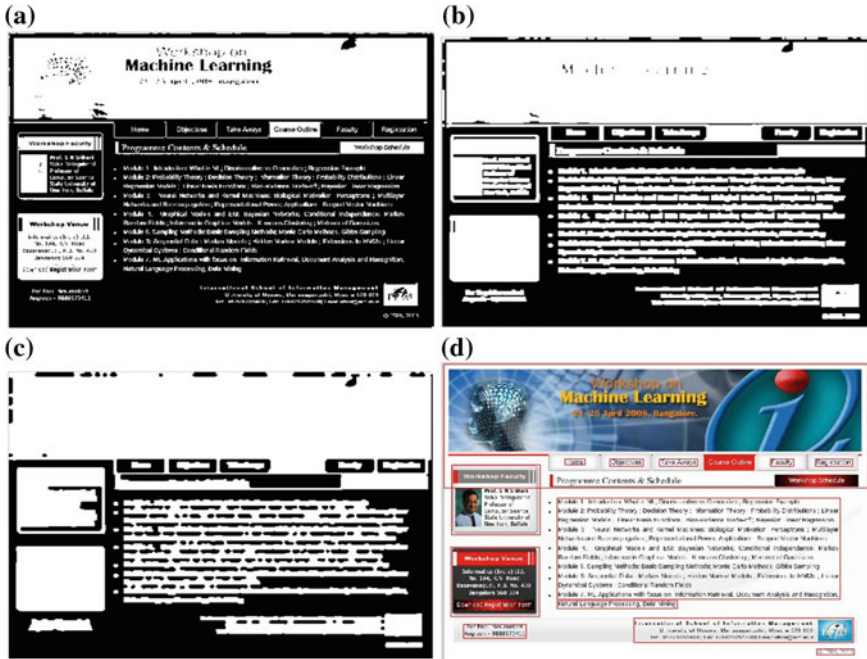


Fig. 5 a Binary image. b Dilated image using 4×4 and 1×3 structuring elements. c Closing image using disk 4 structuring element. d Segmented content block

Table 1 Performance analysis for Haar DWT

Human judgment	User1	User2	User3	User4	User5	All
Perfect	27	26	28	26	24	1311
Satisfactory	5	6	9	8	7	35
Fair	7	9	4	6	8	34
Bad	11	9	9	10	11	50
Total no of pages	50	50	50	50	50	2500
Overall performance						0.8000

Here, percentage segmented Web blocks are calculated based on ground truth analysis and updated the performance factors.

As can be seen in Table 1, $131 + 35 + 34 = 200(80 \%)$ pages have their content blocks correctly detected based on ground truth.

$$A = \frac{\text{Correctly detected block}}{\text{Correctly detected block} + \text{Missed detected blocks}} \tag{3}$$

Fig. 6 Performance analysis of Haar DWT

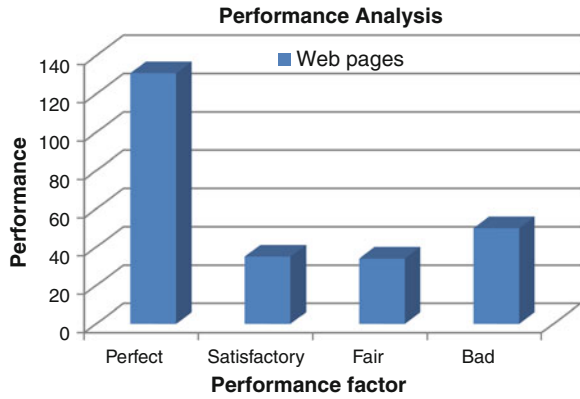
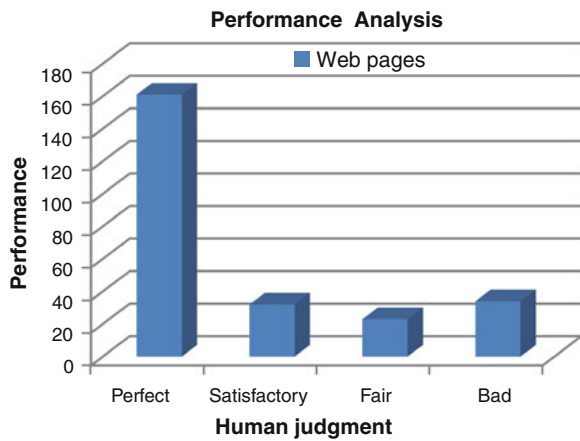


Fig. 7 Performance analysis of connected component with thresholding



$$A = \begin{cases} \text{Perfect} & A \geq 90 \% \\ \text{Satisfactory} & 90 \% < A \geq 80 \% \\ \text{Fair} & 80 \% < A \geq 70 \% \\ \text{Bad} & A \leq 70 \% \end{cases}$$

Proposed method Haar DWT achieves overall 80 % performance on static Flash Web pages. Similarly second method achieves overall 86.46 %. Graphs (Figs. 6, 7) are drawn based on obtained performance factors (Perfect, Satisfactory, Fair, Bad) from the performance Tables 1 and 2 respectively.

Table 2 Performance analysis for connected component with thresholding

Human judgment	User1	User2	User3	User4	User5	All
Perfect	33	31	34	32	31	161
Satisfactory	8	6	6	5	7	32
Fair	3	5	4	6	5	23
Bad	6	8	6	7	7	34
Total no of pages	50	50	50	50	50	250
Total performance						0.8640

Table 3 Accuracy measures for Haar DWT

WEBSITE	Pages	TDB	FDB	MDB	ADB	Recall	Precision	F-measure
Isim	1	10	4	8	18	0.5556	0.7143	0.6250
	2	10	4	5	15	0.6667	0.7143	0.6897
	3	11	4	3	14	0.7857	0.7333	0.7586
	4	12	4	3	15	0.8000	0.7500	0.7742
	5	12	3	4	16	0.7500	0.8000	0.7742
Swf address	1	3	0	0	3	1.0000	1.0000	1.0000
	2	3	1	0	3	1.0000	0.7500	0.8571
	3	3	1	1	4	0.7500	0.7500	0.7500
	4	3	1	1	4	0.7500	0.7500	0.7500
	5	2	0	1	3	0.6667	1.0000	0.8000
Studio	1	13	3	5	18	0.7222	0.8125	0.7647
	2	15	4	6	21	0.7143	0.7895	0.7500
	3	35	5	9	44	0.7955	0.8750	0.8333
	4	20	5	7	27	0.7407	0.8000	0.7692
	5	18	6	13	31	0.5806	0.7500	0.6545
Digaworks	1	19	15	11	30	0.6333	0.5588	0.5938
	2	15	1	5	20	0.7500	0.9375	0.8333
	3	11	7	10	21	0.5238	0.6111	0.5641
	4	15	13	16	31	0.4839	0.5357	0.5085
	5	16	8	15	31	0.5161	0.6667	0.5818
Average								0.7316

5.2 Accuracy Analysis of Proposed Methods

In this section, we conducted several experiments on Flash Web pages from various modalities. Subsequently, quantitative evaluation of the proposed Flash Web page segmentation methods in terms of Precision, Recall and F-measures are given based on *Truly Detected Block (TDB)*: A detected block that encloses more than 80 % of content based on ground truth. *Falsely Detected Block (FDB)*: A detected block that misses more than 20 % of the content block based on ground truth. *Missed Detected Block (MDB)*: Information blocks which are not identified. The percentage is chosen according to [10]. We count manually Actual Detected

Table 4 Accuracy measures for connected component with threshold

WEBSITE	Pages	TDB	FDB	MDB	ADB	Recall	Precision	F-measure
Isim	1	14	2	4	18	0.7778	0.8750	0.8235
	2	12	0	3	15	0.8000	1.0000	0.8889
	3	12	0	2	14	0.8571	1.0000	0.9231
	4	11	2	4	15	0.7333	0.8462	0.7857
	5	13	0	3	16	0.8125	1.0000	0.8966
Swf address	1	2	2	1	3	0.6667	0.5000	0.5714
	2	3	2	1	4	0.7500	0.6000	0.6667
	3	3	2	1	4	0.7500	0.6000	0.6667
	4	3	3	1	4	0.7500	0.5000	0.6000
	5	3	0	0	3	1.0000	1.0000	1.0000
Studio	1	18	1	0	18	1.0000	0.9474	0.9730
	2	20	1	1	21	0.9524	0.9524	0.9524
	3	44	2	0	44	1.0000	0.9565	0.9778
	4	27	1	0	27	1.0000	0.9643	0.9818
	5	29	1	3	32	0.9063	0.9667	0.9355
Digaworks	1	35	14	12	47	0.7447	0.7143	0.7292
	2	22	2	3	25	0.8800	0.9167	0.8980
	3	25	8	4	29	0.8621	0.7576	0.8065
	4	35	10	1	36	0.9722	0.7778	0.8642
	5	33	10	1	34	0.9706	0.7674	0.8571
Average								0.8399

Blocks (ADB) in the images and we presented the detailed analysis of the results obtained. The accuracy measures are defined as follows.

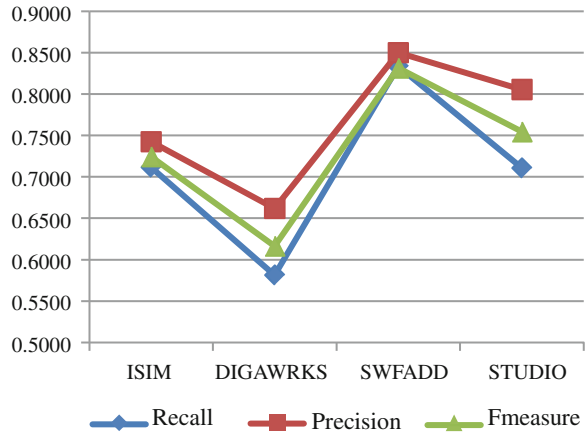
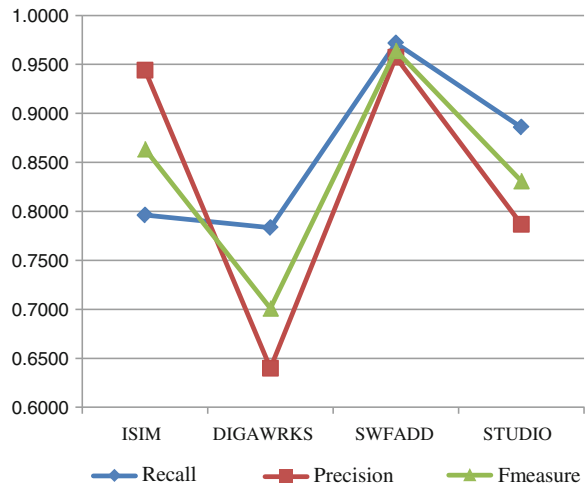
$$\text{Recall (R)} = \frac{TDB}{ADB} \quad (4)$$

$$\text{Precision (P)} = \frac{TDB}{(TDB + FDB)} \quad (5)$$

$$\text{F-measure (F)} = \frac{2PR}{P + R} \quad (6)$$

By using the above mentioned formulations, accuracy measures (Recall, Precision and F-measure) are calculated for proposed methods, which are shown in Tables 3 and 4. And also graphs are plotted by calculating the average of Recall, Precision and F-measure of each Web domains (Figs. 8, 9 respectively).

Compare to DWT, Threshold based Connected component method achieves better performance as well as average accuracy. Graph's clearly shows that accuracy of proposed system is inversely proportional to complexity of Flash Web page image. If complexity is less, then accuracy will be high and wise versa. From Figs. 8 and 9 incurs that, when the high complexity of Digaworks our methods achieve very less accuracy and also for less complexity of Swf Address our methods achieve high accuracy.

Fig. 8 Accuracy analysis for DWT**Fig. 9** Accuracy analysis for connected component with threshold

6 Conclusion

After creating data set of Flash Web images, to segment Web pages as individual blocks, we proposed two different methods such as Haar Discrete Wavelet Transformation and Connected Component with Threshold concepts. We have conducted experiments based on human judgment to analyze the performance of proposed systems. Here, we achieved the overall performance of 80, 86.46 % respectively for the considered dataset. And also experiments are conducted on accuracy measures such as Recall, Precision and F-Measure. Our proposed methods achieve average 73.16 and 83.99 % F-Measure respectively. Interesting

future directions of research are, to improve accuracy based on performing AND operations on matrices and to identify the noise blocks by analyzing the feature vectors of segmented blocks.

References

1. Krishna Murthy, A., Suresha, S.: Comparative study on browsing on small screen devices. *Int. J. Mach. Intell.* **3**(4), 354–358 (2011). ISSN: 0975–2927 and E-ISSN: 0975–9166
2. Book: Ed Tittel Complete Coverage of XML. Tata McGraw-Hill Edition
3. <http://www.w3c.org/DOM/>
4. Cai, D., Yu, S., Wen, J.-R., Ma, W.-Y.: VIPS: a vision based page segmentation algorithm. Technical Report MSR-TR-2003-79 (2003)
5. Xiang, P.F., Yang, X., Shi, Y.C.: Web page segmentation based on gestalt theory. In: Conference on Multimedia and Expo, pp. 2253–2256. IEEE (2007)
6. Chakrabarti, D., Kumar, R., Punera, K.: Small a graph-theoretic approach to webpage segmentation. In: 17th International Conference on WWW (2008)
7. Liu, X., Zhang, X., Tian, Y.: Webpage segmentation based on Gomory-Hu tree clustering in undirected planar graph. NSFC (2010)
8. Krishna Murthy, A., Suresha, S., Anil Kumar, K.M.: Analysis of issues in adapting web contents on mobile devices. In: International Conference on Data Mining and Warehousing. Elsevier Publications (2013)
9. Audithan, S., Chandrasekaran, R.M.: Document text extraction from document images using haar discrete wavelet transform. *Eur. J. Sci. Res.* **36**(4), 502–512 (2009). ISSN: 1450-216X
10. Kovacevic, M., Dilligenti, M.: Recognition of common areas in a web page using visual information: a possible application in a page classification. In: 2nd IEEE International Conference on Data Mining (ICDM'02), p. 250 (2002)

Model Order Reduction of Time Interval System: A Survey

Mahendra Kumar, Aman and Siyaram Yadav

Abstract The Complexity of higher order linear systems are large and order of matrices are higher. Matrices of higher order are difficult to deal with. The main objectives of order reduction is to design a controller of lower order which can effectively control the original high order system so that the overall system is of lower order and easy to understand. Analysis and design of reduced order model becomes simpler and economic. Parametric uncertainties exist in practical systems for entire range of the operating conditions. To overcome this, time interval system is employed. This paper presents a survey on design of reduced order model for large scale time interval systems.

Keywords Interval system · Model order reduction · Reduced order model · Stability

1 Introduction

In recent decades the systems are becoming large and complex and their mathematical modeling are complicated and computationally uneconomic. The accurate analysis of such higher order systems is complex to be used in real problems therefore, it's reasonable to sacrifice with model accuracy in order to obtain a

M. Kumar (✉) · S. Yadav

Department of Electronics Engineering, Mewar University, Gangrar,
Chittorgarh, India
e-mail: miresearchlab@gmail.com

S. Yadav

e-mail: er.siyaramyadav@gmail.com

Aman

Department of Electronics Engineering, Amity University, Noida, India
e-mail: aman81kathuria@gmail.com

simpler model. There are so many methods to reduce the order of the system model while keeping the dominant behavior of the original system. This lower order model helps to better understanding of physical system, reduce computational complexity, reduce hardware complexity, simplify controller design and finally simulation becomes computationally cheaper which saves the time and resources [1–4].

In field of model reduction of linear and nonlinear dynamic system, several methods like, Moment Matching, Aggregation Method, Pade Approximation, Routh Stability Techniques, L^∞ optimization technique have been proposed among all these routh stability techniques is simplest and powerful method because of its ability to yield stable reduced models for stable high-order systems. Further, numerous methods of order reduction are also available in the literature, which are based on minimization of the ISE criterion [5–9].

Many practical systems have parametric uncertainties for the entire range of operating conditions, the interval system analysis and design has received considerable attention [10]. The γ - δ Routh approximation for interval systems was proposed by [11]. In the γ - δ Routh approximation of interval systems, a r th order model was obtained by retaining first r , γ 's and δ 's of the higher order interval system [12], shows that stability preservation of non-interval Routh approximants cannot be claimed for the interval Routh approximation, even if they are obtained by using the optimization techniques [12]. The probability of the interval Routh approximant yielding unstable models for stable original systems increases with the increase in the model order. This is due to the fact that the interval operations are implicitly over bounded. In γ - δ Routh approximation of the interval system to obtain r th order model, it is required to perform r interval γ multiplications. Thus, increase in the approximation order r results in growing over bounds on interval coefficients of the reduced interval denominator polynomial. An improvement is proposed to the γ - δ Routh approximation technique for interval systems using the Kharitonov polynomials in [13]. An advantage of using the Kharitonov polynomials is that the resulting reduced order interval model is always robustly stable.

New mixed method was suggested to overcome the instability of the system which proved wrong assumption in further investigation. Mixed methods proposed recently may be considered powerful techniques to reduce the complexity and to increase the accuracy of the system.

The organization of paper ahead is as follows: basic theory of the Kharitonov polynomials and background of interval system for reduced order model is discussed in Sect. 2. Section 3 gives a review of the model order reduction of interval systems. In Sect. 4 the experimental results from reviews of different papers from Sect. 3, followed by conclusions.

2 Background

2.1 Problem Statement Without Time Interval

Let n th order system and its reduced model ($r < n$) be given by the transfer function:

$$G(s) = \frac{\sum_{i=0}^{n-1} a_i s^i}{\sum_{j=0}^n e_j s^j} \quad (1)$$

$$R(s) = \frac{\sum_{i=0}^{r-1} a_i s^i}{\sum_{j=0}^r b_j s^j} \quad (2)$$

a_i, b_j, e_j are scalar constants. The objective is to find a reduced r th order model $R(s)$ such that it retains the important properties of $G(s)$ for the same type of input.

2.2 Problem Statement with Time Interval

Consider high order linear SISO interval system represented by the transfer function as:

$$G(s) = \frac{[c_0^-, c_0^+] + [c_1^-, c_1^+]s + [c_2^-, c_2^+]s^2 + \dots + [c_{n-1}^-, c_{n-1}^+]s^{n-1}}{[d_0^-, d_0^+] + [d_1^-, d_1^+]s + [d_2^-, d_2^+]s^2 + \dots + [d_{n-1}^-, d_{n-1}^+]s^{n-1}} \quad (3)$$

where $[c_i^-, c_i^+], i = 0, 1, 2, \dots, n - 1$ and $[d_i^-, d_i^+]; i = 0, 1, 2, \dots, n$. are interval coefficients of higher order numerator and denominator polynomials respectively. The objective is to find an r th order reduced interval system. Let corresponding r th order reduced model is

$$R(s) = \frac{[a_i^-, a_i^+] + [a_1^-, a_1^+]s + [a_2^-, a_2^+]s^2 + \dots + [a_r^-, a_r^+]s^{r-1}}{[b_0^-, b_0^+] + [b_1^-, b_1^+]s + [b_2^-, b_2^+]s^2 + \dots + [b_r^-, b_r^+]s^r} \quad (4)$$

where $[a_i^-, a_i^+], i = 0, 1, 2, \dots, r - 1$ and $[b_i^-, b_i^+], i = 0, 1, 2, \dots, r$ are the interval coefficients of lower order numerator and denominator polynomials respectively.

2.3 Interval Arithmetic

The rule for interval arithmetic has been defined as follows. Let $[a, b]$ and $[c, d]$ be two intervals.

- **Addition:** $[a, b] + [c, d] = [a + c, b + d]$
- **Subtraction:** $[a, b] - [c, d] = [a - d, b - c]$
- **Multiplication:** $[a, b][c, d] = [\text{Min}(ac, ad, bc, bd), \text{Max}(ac, ad, bc, bd)]$
- **Division:** $\frac{[a, b]}{[c, d]} = [a, b] \left[\frac{1}{d}, \frac{1}{c} \right]$ provided $[c, d] \neq 0$.

2.4 Kharitonov Robust Stability Criteria

An interval polynomial $P(s) = \sum_{i=0}^n [a_i^-, a_i^+]s^i$, is associated with four following polynomials, called Kharitonov Polynomials:

$$\begin{aligned}
 K_1 &= a_0^- + a_1^-s + a_2^+s^2 + a_3^+s^3 + a_4^-s^4 + a_5^-s^5 + a_6^+s^6 + \dots \\
 K_2 &= a_0^+ + a_1^+s + a_2^-s^2 + a_3^-s^3 + a_4^+s^4 + a_5^+s^5 + a_6^-s^6 + \dots \\
 K_3 &= a_0^+ + a_1^-s + a_2^-s^2 + a_3^+s^3 + a_4^+s^4 + a_5^-s^5 + a_6^-s^6 + \dots \\
 K_4 &= a_0^- + a_1^+s + a_2^+s^2 + a_3^-s^3 + a_4^-s^4 + a_5^+s^5 + a_6^+s^6 + \dots
 \end{aligned}
 \tag{5}$$

It can be seen that these polynomials are constructed from two different even and odd parts as follows:

$$\begin{aligned}
 K_1 &= K_{\min}^{\text{even}} + K_{\min}^{\text{odd}} \\
 K_2 &= K_{\min}^{\text{even}} + K_{\max}^{\text{odd}} \\
 K_3 &= K_{\max}^{\text{even}} + K_{\min}^{\text{odd}} \\
 K_4 &= K_{\max}^{\text{even}} + K_{\max}^{\text{odd}}
 \end{aligned}
 \tag{6}$$

Theorem 1 (Robust Stability of interval Systems): An Integral Polynomial family $P(s) = \sum_{i=0}^n [a_i^-, a_i^+]s^i$ with invariant degree is stable if it's four Kharitonov polynomials are stable.

3 Review on Model Order Reduction Techniques for Linear and Non-linear Interval System

Depending on the properties of the original system that are retained in the reduced model, there are different model reduction methodologies. Many of the model reduction methods can be classified as Routh Approximation, such as Routh Pade, Revised Routh Approximation, Improved Routh Pade methods and Mihailov Criterion with Factor division, Cauer II form, and differential methods with Factor division, Cauer II form. These methods construct a reduced order model via the $\gamma-\delta$ Routh Approximation. These also determine which properties are preserved after reduction.

3.1 Order Reduction of Interval Systems Using Alpha Table and Factor Division Method and Cauer II Form

Determination of the denominator polynomial of the kth reduced model as given in Eq. 4 by using α table:

Reciprocal of higher order denominator gives:

$$D(s) = \frac{1}{s} \hat{D}\left(\frac{1}{s}\right) \tag{7}$$

Step by step Algorithm:

1. Determine the reciprocal $\hat{D}(s)$ of the full model $D(s)$.
2. Construct the alpha table corresponding to $\hat{D}(s)$.
3. Determine kth denominator by Routh convergent $\hat{D}(s) = A_k(s)$
4. Reciprocate $\hat{D}(s)$ to determine denominator by Routh convergent $D_k(s)$.

Determine the numerator coefficients of the kth order reduced model by using factor division method and cauer second form method.

3.2 Denominator Polynomial of the kth Order Reduced Model by Using Differentiation Method

Determination of the denominator polynomial of the kth order reduced model as given in Eq. 4 by using differentiation method: $D_k(s) = \frac{1}{s} D\left(\frac{1}{s}\right)$

Differentiate the denominator of Eq. 4 into $(n - k)$ times:

$$D_k(s) = [d_{11}^-, d_{11}^+] + [d_{12}^-, d_{12}^+]s + \dots + [d_{1,k+1}^-, d_{1,k+1}^+]s^k \tag{8}$$

Case 1: Determine of the numerator coefficient of the kth order reduced model by using differentiation method: $N_k(s) = \frac{1}{s}N(\frac{1}{s})$

Differentiate the numerator of Eq. 5 into $(n - k)$ times:

$$N_k(s) = [d_{21}^-, d_{21}^+] + [d_{22}^-, d_{22}^+]s + \dots + [d_{2,k+1}^-, d_{2,k+1}^+]s^{k-1} \tag{9}$$

Case 2: Determination of the numerator polynomial of the kth order reduced model by using factor division method.

Any method of reduction which relies upon calculating the reduced denominator, where $D_k(s)$ has already been calculated:

$$G_n(s) = \frac{N(s)D_k(s)/D(s)}{D(s)} \tag{10}$$

The reduced transfer function is given by

$$R_k(s) = \frac{[\alpha_{11}^-, \alpha_{11}^+] + [\alpha_{12}^-, \alpha_{12}^+]s + \dots + [\alpha_{1,k-1}^-, \alpha_{1,k-1}^+]s^{k-1}}{D_k(s)} \tag{11}$$

Case 3: Determination of the numerator polynomial of the kth order reduced model by using Caer second form:

Coefficient value from Caer II form $[h_i^-, h_i^+]$ ($i = 1, 2, 3, \dots, k$) are evaluated by forming Routh array as

$$[h_1^-, h_1^+] = \frac{[c_{11}^-, c_{11}^+]}{[c_{21}^-, c_{21}^+]} \left\{ \begin{matrix} [c_{11}^-, c_{11}^+] & [c_{12}^-, c_{12}^+] \dots \\ [c_{21}^-, c_{21}^+] & [c_{22}^-, c_{22}^+] \dots \end{matrix} \right\} \tag{12}$$

The coefficient values of $[d_{i,j}^-, d_{i,j}^+]$ ($j = 1, 2, \dots (k + 1)$) and Caer quotients $[h_i^-, h_i^+]$ ($i = 1, 2, \dots, k$) are matched for finding the coefficients of numerator of the reduced model $R_k(s)$. The inverse Routh array is constructed as:

$$[d_{i+1,1}^-, d_{i+1,1}^+] = \frac{[d_{i,1}^-, d_{i,1}^+]}{[h_i^-, h_i^+]} \tag{13}$$

where $i = 1, 2, \dots, k$ and $k \leq n$, also

$$[d_{i+1,j+1}^-, d_{i+1,j+1}^+] = \frac{[d_{i,j+1}^-, d_{i,j+1}^+] - [d_{i+2,j}^-, d_{i+2,j}^+]}{[h_i^-, h_i^+]} \tag{14}$$

where $i = 1, 2, \dots, (k - j)$ and $j = 1, 2, \dots, (k - 1)$.

Using the above equations, the numerator coefficients of the reduced model will be obtained.

Case 4: Determination of the numerator polynomial of the $(k - 1)$ th order reduced model by using Moment matching method: the Eq. 4 is modified and written as follows:

$$G(s) = \frac{[b_n^-, b_n^+] + [b_{n-1}^-, b_{n-1}^+]s + \dots + [b_1^-, b_1^+]s^{n-1}}{[a_n^-, a_n^+] + [a_{n-1}^-, a_{n-1}^+]s + \dots + [a_0^-, a_0^+]s^n} = \frac{N(s)}{D(s)} \quad (15)$$

Therefore, the reduced transfer function in general form is:

$$R_k(s) = \frac{[B_{21}^-, B_{21}^+] + [B_{22}^-, B_{22}^+]s + \dots}{[1, 1] + [B_{12}^-, B_{12}^+]s + [B_{12}^-, B_{12}^+]s^2 + \dots} \quad (16)$$

3.3 The Integral Square Error

The Integral Square Error (ISE) between the transient response of higher order system (HOS) and lower order system (LOS) is determined to compare different approaches of model reduction. This is given by

$$ISE = \int_0^{\infty} [y(t) - y_r(t)]^2 \quad (17)$$

Where, $y(t)$ and $y_r(t)$ are unit step responses of original system $G_n(s)$ and reduced order system $R_k(s)$.

Routh–Pade’ approximation for interval system [11]: Author’s presents a method for the reduction of the order of interval system. The denominator of the reduced model is obtained by a direct truncation of the Routh table of the interval system. The numerator is obtained by matching the coefficients of power series expansions of the interval system and its reduced model.

γ – δ Routh Approximation of interval System [14]: Author’s presents the γ – δ Routh approximation for interval systems. The interval γ ’s and δ ’s are evaluated for the higher order interval systems, and then an r th-order approximant is obtained by retaining the first r , interval δ ’s, and γ ’s. In Bandyopadhyay et al. [11], the Routh approximation method was extended to derive reduced-order interval models for linear interval systems. The authors show that: (1) interval Routh approximants to a high-order interval transfer function depend on the implementation of interval Routh expansion and inversion algorithms; (2) interval Routh expansion algorithms cannot guarantee the success in generating a full interval Routh array; (3) some interval Routh approximants may not be robustly stable even if the original interval system is robustly stable; and (4) an interval Routh approximant is in general not useful for robust controller design because its

dynamic uncertainties (in terms of robust frequency responses) do not cover those of the original interval system.

Stable γ - δ Routh Approximation of Interval Systems Using Kharitonov Polynomials [15]: An improvement is proposed to the γ - δ Routh approximation for interval systems using the Kharitonov polynomials such that the resulting interval Routh approximant is robustly stable and overcome the problem of stable 5th order reduced order model derived from 7th order original interval system [14].

A Mixed Method for Order Reduction of Interval Systems [15]: An effective procedure to determine the reduced order model of system using mixed methods of Eigen spectrum and Pade approximation is discussed. The eigen values of reduced order interval system are obtained using Eigen spectrum by preserving some of the characteristics such as centroid and stiffness of the original high order interval system. The numerator polynomial is obtained using Pade approximation by preserving some of the time moments and Markov parameters.

Model Order Reduction of Interval Systems using Mihailov Criterion and Factor Division Method [16]: Author presents a mixed method for reducing order of the large scale interval systems using the Mihailov Criterion and factor division method. The denominator coefficients of reduced order model is determined by using Mihailov Criterion and numerator coefficients are obtained by using Factor division method. The mixed methods are simple and guarantee the stability of the reduced model if the original system is stable.

Order Reduction of Linear Interval Systems Using Genetic Algorithm [17]: Author proposed, an evolutionary method using Genetic algorithm for reducing a high order large scale linear interval system into a lower order interval system. In this algorithm the numerator and denominator polynomials are determined by minimizing the Integral square error (ISE) using genetic algorithm (GA). The algorithm is simple, rugged and computer oriented. It is shown that the algorithm has several advantages, e.g. the reduced order models retain the steady-state value and stability of the original system. The proposed algorithm guarantees stability for a stable higher order linear Interval system and thus any lower order Interval model can be derived with good accuracy.

Mixed Evolutionary Techniques to Reduce Order Of Linear Interval Systems Using Generalized Routh Array [5]: In this work, two evolutionary methods for reducing a high order large scale linear interval system into a lower order interval system have been proposed. Particle swarm optimization and genetic algorithm methods based evolutionary optimization techniques are employed for the order reduction where the numerator polynomials are determined by generalized Routh array truncation and denominator coefficients of reduced order model are obtained by minimizing an Integral Squared Error (ISE) criterion. The proposed method is mathematically simple and gives all possible stable lower order models. The proposed method guarantees stability for a stable higher order system and thus any lower order model can be derived with good accuracy. Both techniques are simple rugged and computer oriented.

A New Algorithm for Model Order Reduction of Interval Systems [6]: Mixed method of interval systems is a combination of classical reduction methods and stability preserving methods of interval systems. The bounds on the uncertain parameters are known a priori. The numerator parameters are obtained by either of these methods such as differentiation method, factor division method, Cauey II form, and moment matching method or Pade approximation method. The denominator is obtained by the differentiation method in all the cases. The Moment matching method is not guaranteed to be good method for reduction but is worth a new methodology to obtain a reduced order system. Being not useful here is not enough for its suitability, as when used along with some other methods generates useful and brilliant results. From the above mixed methods, differentiation method and Cauey second form as resulted in better approximation when compared with other methods.

Design of Robust PID controller for reduced order interval system [18]: A system with parameter variation within bound creates interval in coefficients in the system polynomial and hence it is called as interval system. This model reduction method is based on convergence technique, generates stable reduced order model for original higher order stable model by retaining both initial time moments and Markov parameters. Reduced order model performance match the higher order model performance for both steady state and transient state of the time response. Also in this study a novel technique using Particle Swarm Optimization–Bacterial Forging (PSO–BFO) based hybrid algorithm is proposed to search the PID controller parameter such as K_p , K_i and K_d . The algorithm is to obtain the best possible PID parameters with ISE criterion minimization is as the objective function. A controller parameter is identified for the reduced order model; using Bacterial Forging based Particle Swarm Optimization (PSO) which controls the higher order model effectively. Step and impulse response are obtained are compared. But the time taken by the algorithm is large compared to other methods.

4 Experimental Results

Case study 1: Consider the third order system described by the transfer function:

$$G_3(s) = \frac{[2, 3]s^2 + [17.5, 18.5]s + [15, 16]}{[2, 3]s^3 + [17, 18]s^2 + [35, 36]s + [20.5, 21.5]} \tag{18}$$

Case study 2: Consider the 7th order system described by the transfer function (Tables 1, 2):

Table 1 Comparison of reduced order model for case study 1

Method of order reduction	Reduced order system	ISE for lower limit	ISE for upper limit
Differentiation [6]	$R_2(s) = \frac{[5.8333,6.1667]S+[10.0005,10.6672]}{[4.25,4.5]S^2+[17,18]S+[15.375,16.125]}$	0.0531	0.0617
Differentiation and factor division [6, 7]	$R_2(s) = \frac{[3.4809,10.7884]S+[10.7267,12.5854]}{[4.25,4.5]S^2+[17,18]S+[15.375,16.125]}$	0.0094	0.0074
Differentiation and cauer II form [6]	$R_2(s) = \frac{[4.2604,9.6099]S+[10.7271,12.5856]}{[4.25,4.5]S^2+[17,18]S+[15.375,16.125]}$	0.0094	0.0073
Differentiation and moment matching [6]	$R_2(s) = \frac{[0.0111,0.0449]S+[0.0433,0.0508]}{[0.2636,0.2927]S^2+[1.0543,1.1707]S+[1,1]}$	4.1118	4.219
Differentiation and pade approximation [6]	$R_2(s) = \frac{[3.30187,10.2867]S+[10.7267,12.5854]}{[4.25,4.5]S^2+[17,18]S+[15.375,16.125]}$	0.0094	0.0105
Sastry [19]	$R_2(s) = \frac{[0.94,1.35]S+[0.8409,1.168]}{s^2+[2.0181,2.4430]S+[1.492,1.5007]}$	0.2256	0.0095
Mihailov criterion and factor division [9]	$R_2(s) = \frac{[35.6065,49.4454]S+[14.0331,17.1024]}{[17.0011,18.0007]S^2+[31.3826,2.4430]S+[20.3061,21.7052]}$	0.0124	0.0169
Mihailov criterion and cauer ii form [16]	$R_2(s) = \frac{[11.1949,20.3706]S+[14.1674,16.9413]}{[17.0011,18.0007]S^2+[31.3826,33.6111]S+[20.3061,21.7052]}$	0.0264	1.0314
Alpha and factor division	$R_2(s) = \frac{[0.4838,1.826]S+[0.8241,1.1307]}{s^2+[2.0758,2.3751]S+[1.1896,1.459]}$	0.0125	0.0082
Alpha and cauer II form	$R_2(s) = \frac{[0.7236,1.5]S+[0.8238,1.1309]}{s^2+[2.0758,2.3751]S+[1.1896,1.459]}$	0.0094	0.0073

Table 2 Comparison of reduced order model for case study 2

Method of order reduction	2nd reduced order model	ISE for lower limit	ISE for upper limit
Babu and Natarajan Proposed Method [18]	$R_2(s) = \frac{[49.32,66.6]S+[169.24,229.64]}{[27.26,30.13]S^2+[92.937,102.72]S+[57.352,63.789]}$	0.808	0.8215
Saini and Prasad GA Algorithm [17]	$R_2(s) = \frac{[553.9,568.3]S+[181.9,205]}{[319.49,406.63]S^2+[259.89,300.36]S+[57.352,63.389]}$	0.825	5.562
Saini and Prasad PSO Algorithm [20]	$R_2(s) = \frac{[562.4,555.6]S+[181.6,205.4]}{[319.49,406.63]S^2+[259.89,300.36]S+[57.352,63.389]}$	0.822	5.558
Bandyopadhyay [15]	$R_2(s) = \frac{[1.16,1.84]S+[0.27,0.53]}{s^2+[0.52,.83]S+[0.08,0.16]}$	2.2599	5.954

$$G_7(s) = \frac{[1.9, 2.1]s^6 + [24.7, 27.3]s^5 + [157.7, 174.3]s^4 + [541.975, 599.025]s^3 + [929.955, 1027.845]s^2 + [721.81, 797.79]s + [187.055, 206.745]}{[0.95, 1.05]s^7 + [8.779, 9.703]s^6 + [52.231, 57.729]s^5 + [182.875, 202.125]s^4 + [429.02, 474.18]s^3 + [572.47, 632.73]s^2 + [325.28, 359.52]s + [57.352, 63.389]}$$

(19)

5 Conclusion and Future Aspects

Review on all the above techniques and methods leads to the conclusion that mixed Differentiation and Caue second form Techniques [6] are better compared to other MOR techniques in case 1st where the original 3rd order system is reduced in second order system using different techniques. For higher order system, in case 2nd the original 7th order system is reduced in second order system using different techniques and found Babu and Natarajan Proposed Method [18] for model order reduction of time interval system to be better compared to other MOR techniques. All the above techniques preserve the stability of original time interval system in reduced order time interval system.

Future systems will be more complex than the problems in hand now. Therefore recognizing the versatility of “mixed techniques with evolutionary techniques for model order reduction of time interval system” propose to take into account these techniques for design and development of controller to be applied on complex systems in times to come and robustly stable reduced order model for sufficiently complex higher order systems.

Acknowledgments Siyaram Yadav sincerely acknowledge Mr. Mahendra Kumar, Assistant professor, Mewar University, Chittorgarh (Rajasthan, India) for his kind cooperation and consistent guidance during M.Tech Program.

References

1. Kumar, M., Gupta, R.: Design of decentralized PSSs for multi machine power system via reduced order model. In: 4th IEEE international conference on computational intelligence and communication networks (CICN), pp. 617–621 (2012)
2. Singh, V.P., Chandra, D.: Routh approximation based model reduction using series expansion of interval systems. In: IEEE International conference on power, control & embedded systems (ICPCES), vol 1, pp. 1–4 (2010)
3. Hansen, E.: Interval arithmetic in matrix computations. Part I, SIAM J. Num.Anal. **2**, 308–320 (1965)
4. Gutman, P.O., Mannerfelt, C.F., Molander, P.: Contributions to the model reduction problem. IEEE Trans. Autom. Control **27**(2), 454–455 (1982)
5. Saini, D.K., Prasad, R.: Mixed evolutionary techniques to reduce order of linear interval systems using generalized routh array. Int. J. Eng. Sci. Technol. **2**(10), 51976–55205 (2010)
6. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: A new algorithm for model order reduction of interval systems. Bonfring Int. J. Data Min. **3**(1), 6 (2013)

7. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: Model order reduction of interval systems using modified Routh approximation and factor division method. In: 35th national system conference (NSC), IIT Bhubaneswar, India (2011)
8. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: Model order reduction of interval systems using mihailov criterion and routh approximations. *Int. J. Eng. Sci. Technol. (IJEST)* **3**(7), 5593–5598 (2011)
9. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: Model order reduction of interval systems using mihailov criterion and factor division method. *Int. J. Comput. Appl. (IJCA)* **28**(11), 4–8 (2011)
10. Bhattacharya, S.P.: *Robust Stabilization Against Structured Perturbations*. Lecture Notes in Control and Information Sciences. Springer, New York (1987)
11. Bandyopadhyay, B., Upadhye, A., Ismail, O.: γ - δ Routh approximation for interval systems. *IEEE Trans. Autom. Control* **42**(8), 1127–1130 (1997)
12. Hwang, C., Yang, S.F.: Comments on the computations of interval Routh approximants. *IEEE Trans. Autom. Control* **44**(9), 1782–1787 (1999)
13. Ismail, O., Bandyopadhyay, B.: Model order reduction of linear interval systems using pade approximation. In: *IEEE International Symposium on Circuit and Systems* (1995)
14. Dolgin, Y., Zeheb, E.: On Routh pade model reduction of interval systems. *IEEE Trans. Autom. Control* **48**(9), 1610–1612 (2003)
15. Bandyopadhyay, B., Sreeram, V., Shingare, P.: Stable γ - δ Routh approximation for interval systems using Kharitonov polynomials. *Int. J. Inf. Syst. Sci.* **44**(3), 348–361 (2008)
16. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: Model order reduction of interval systems using mihailov criterion and cauer second form. *Int. J. Comput. Appl.* **32**(6), 17–21 (2011)
17. Saini, D.K., Prasad, R.: Order reduction of linear interval systems using genetic algorithm. *Int. J. Eng. Technol.* **2**(5), 316–319 (2010)
18. Babu, T., Natarajan, P.: Design of robust pid controller using hybrid algorithm for reduced order interval system. *Asian J. Sci. Res.* **5**(3), 108–120 (2012)
19. Sastry, G.V.K., Raja Rao, G.R., Rao, P.M.: Large scale interval system modeling using Routh approximants. *Electron. Lett.* **36**(8), 768 (2000)
20. Saini, D.K., Prasad, R.: Order reduction of linear interval systems using particle swarm optimization. *MIT Int. J. Elect. Instrum. Eng.* **1**(1), 16–19 (2011)
21. Kharitonov, V.L.: Asymptotic stability of an equilibrium position of a family of systems of linear differential equations. *Differentsial'nye Uravneniya* **14**, 2086–2088 (1978)
22. Bandyopadhyay, B., Ismail, O., Gorez, R.: Routh pade approximation for interval systems. *IEEE Trans. Autom. Control* **39**, 2454–2456 (1994)
23. Saraswathi, G.: A mixed method for order reduction of interval systems. In: *International Conference on Intelligent and Advanced Systems*, pp. 1042–1046 (2007)
24. Dolgin, Y.: Author's reply. *IEEE Trans. Autom. Control* **50**(2), 274–275 (2007)
25. Singh, V.P., Chandra, D.: Model reduction of discrete interval system using dominant poles retention and direct series expansion method. In: *IEEE 5th international power engineering and optimization conference (PEOCO)*, vol 1, pp. 27–30 (2011)
26. Choo, Y.: A note on discrete interval system reduction via retention of dominant poles. *Int. J. Control Autom. Syst.* **5**(2), 208–211 (2007)
27. Kumar, D.K., Nagar, S.K., Tiwari, J.P.: Model order reduction of interval systems using Routh approximations and cauer second form. *Int. J. Adv. Sci. Technol. (IJAST)* (2011)

Author Biography

Mahendra Kumar has obtained M.Tech (Control and Instrumentation Engg.) and B.Tech (Electronics & Communication Engg.) from Rajasthan Technical University, Kota in 2012 and 2010 respectively. He is working as Assistant Professor in Electronics Engineering Department at Mewar University, Gangrar. His research interests are power system stabilizer, Multi-rate output feedback (fast output sampling feedback) techniques and model order reduction method, Evolutionary optimization techniques, Optimal control system, time interval (uncertain) system analysis, signal processing, image processing, Non-linear control system.

Design and Implementation of Fuzzy Logic Rule Based System for Multimodal Travelling Network

Madhavi Sharma, Jitendra Kumar Gupta and Archana Lala

Abstract In a country like India, which has large number of routes between two given places, it is difficult to identify the most convenient and best route for travelling. Also, deciding the mode of transport according to the comfort of the person will require deep knowledge of the routes and the variables which affect the travel. To solve these problems, many algorithms have been suggested, but there are very few models that incorporate all the factors affecting the convenience of the routes and the modes of transport. In the developed Fuzzy Rule Based System, multiple modes are evaluated for all the existing routes incorporating all the factors and the favorable outcome is determined. I observed that finding the optimal route and the best transport mode becomes very easy with this system; also it emphasizes the convenience factor that is of utmost importance in any kind of transport applications. This system has additional advantage of economy of the time consumed for finding the results.

Keywords Route choice problem · Travelling network · Multimodal travelling network · Fuzzy rule based system for route choice

1 Introduction

India is a large subcontinent offering different routes and modes of travel between two given places. Therefore choice of the most appropriate route needs complex analysis of each and every route depending on a number of factors that may

M. Sharma (✉) · J. K. Gupta · A. Lala
SR Group of Institutions, CSE, Jhansi, India
e-mail: madhavi.dhingra@gmail.com

J. K. Gupta
e-mail: jitendra1503@gmail.com

A. Lala
e-mail: archanalala2006@yahoo.com

influence the efficiency of the route. For the last few years, the route choice problem has been a major research area in the field of transport planning. Route choice along with the choice of mode plays an important role in decision making of the better transport plan. From the background, it is observed that there are many techniques for solving the multimodal route choice problem. Many traffic and transportation problems and their deciding parameters are based on subjectivity [1]. The problems that deal with choice of route, mode of transportation, traveler's perception, and his preferences, all are based on subjective judgments. Many models were developed based on deterministic and stochastic methods using binary logic. But they cannot deal with the uncertainty, vagueness and ambiguity of the traveler's decision. Since the fuzzy set theory recognizes the vague boundary that exists in some sets, different fuzzy set theory techniques need to be used in order to properly model traffic and transportation problems characterized by ambiguity, subjectivity and uncertainty.

Fuzzy logic is a convenient way to map an input space to an output space. In terms of transport and routing application, the input space will be the information about all the routes between a pair of origin and destination, the output space will be the best possible route.

2 Description of the Multimodal Route Selection Fuzzy Rule Based System

In the proposed system, I have considered two approaches for finding the best route and mode. I have assumed that there will be a direct route between source and destination. Also, if there is no direct mode of transport between source and destination then indirect route will be considered by selecting the most appropriate intermediates. Distance is taken as an important factor in all Fuzzy Rule Based Systems. As travel time and cost of travelling are dependent on the distance, so they are not considered separately. I have assumed that all the modes are available at the time of travelling.

2.1 Direct Routing Fuzzy Rule Based System

Considering the three common modes of transport, I have selected some major factors, input variables that affect the effectiveness of the route. The preference of mode is determined by the respective values of the input values (Fig. 1).

Basic Assumptions for the Direct Approach

- I have not considered cost factor in our analysis of the best route. The whole stress is on comfort and convenience.
- I have assumed that all modes of travel would be available at time of journey.

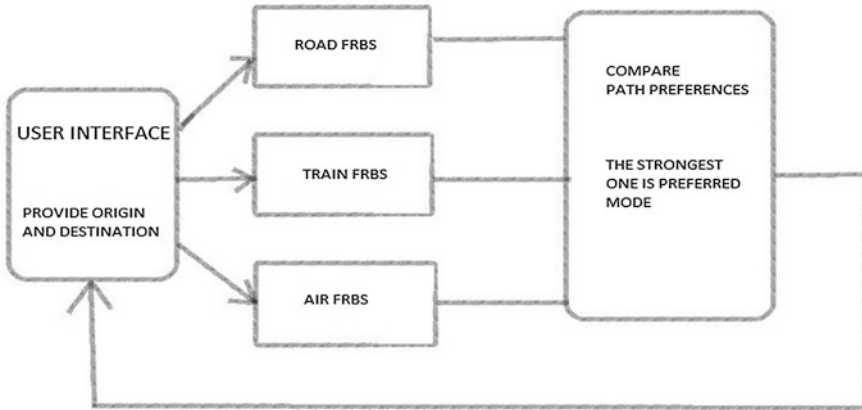


Fig. 1 Direct routing approach of route selection

- I have not taken any inconvenience due to odd timings of travel or odd timings of break in journey.
- Inconvenience of airport check in and check out and reaching the airport is taken to be same for all airports.

(1) Road Fuzzy Rule Based System:

Input Variables:

1. Distance—small, medium, large
2. Road condition—bad, medium, good
3. Traffic condition—bad, medium, good

Output Variable: Road Preference—weak, medium, strong

(2) Railway Fuzzy Rule Based System:

Input Variables:

1. Distance—small, medium, large
2. Comfort Factor—less comfortable, average comfortable, luxury

Output Variable: Railway Preference—weak, medium, strong

(3) Airway Fuzzy Rule Based System:

Input Variables:

1. Distance—small, medium, large
2. Comfort Factor—less fare, economy class, business class

Output: Airway Preference—weak, medium, strong

The output of roadway, railway and airway fuzzy rule based system is compared and then the best path preference is determined.

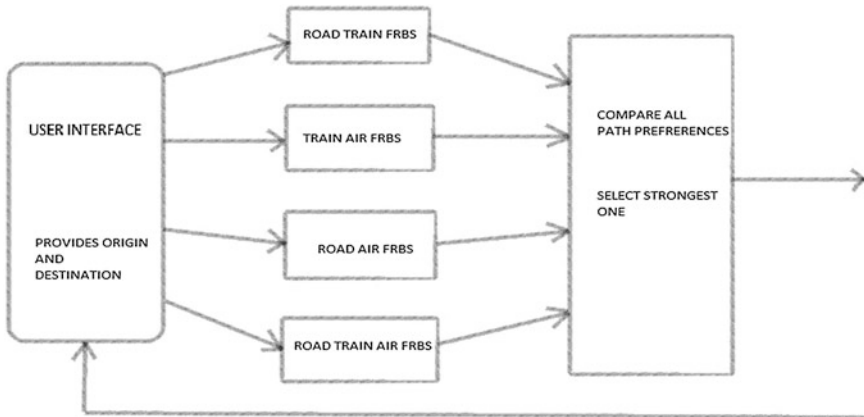


Fig. 2 Indirect approach for route selection

2.2 Indirect Routing Fuzzy Rule Based System

When direct mode of transport does not exist between two places, then two or more than two modes need to be considered. In this process, one more important factor affects the route selection, waiting period. Depending on several factors of direct routing approach and waiting period, the best route and mode is decided. Four fuzzy rule based systems are considered (Fig. 2).

Basic Assumptions for the Indirect Approach

- I have not considered cost factor in our analysis of the best route. The whole stress is on comfort and convenience.
- If there is no direct train or flight between origin and destination, then I have considered case where a switch over from one mode to other is required.
- I have assumed that all modes of travel would be available at time of journey.
- I have not taken any inconvenience due to odd timings of travel or odd timings of break in journey.
- Inconvenience of airport check in and check out and reaching the airport is taken to be same for all airports.

(1) Roadway Railway Fuzzy Rule Based System—This system considers route involving roadway and railway in no particular order. The inputs will be distance, comfort factor along with waiting period. Comfort factor depends on the road condition and traffic condition. The waiting period will include the time for catching two or more different mode of transport. The values for inputs are as follows:

1. Distance—small, medium, large
2. Comfort Factor—less comfortable, average comfortable, luxury
3. Waiting period—less than 2 h, between 2 and 4 h, more than 4 h

Output: Roadway Railway Preference—weak, medium, strong

(2) Railway Airway Fuzzy Rule Based System—This system considers route involving railway and airway in no particular order. The inputs will be distance, comfort factor along with waiting period. The values for inputs are as follows:

1. Distance—small, medium, large
2. Comfort Factor—less comfortable, average comfortable, luxury
3. Waiting period—less than 2 h, between 2 and 4 h, more than 4 h

Output: Railway Airway Preference—weak, medium, strong

(3) Roadway Airway Fuzzy Rule Based System—This system considers route involving roadway and airway in no particular order. The inputs will be distance, comfort factor along with waiting period. This waiting period will include the time for catching two or more different mode of transport. The values for inputs are as follows:

1. Distance—small, medium, large
2. Comfort Factor—less comfortable, average comfortable, luxury
3. Waiting period—less than 2 h, between 2 and 4 h, more than 4 h

Output: Roadway Airway Preference—weak, medium, strong

(4) Roadway Railway Airway Fuzzy Rule Based System—This system considers route involving roadway, railway and airway in no particular order. The inputs will be distance, comfort factor and waiting period. This waiting period will include the time for catching two or more different mode of transport. The values for inputs are as follows:

1. Distance—small, medium, large
2. Comfort Factor—less comfortable, average comfortable, luxury
3. Waiting period—less than 2 h, between 2 and 4 h, more than 4 h

Output: Roadway Railway Airway Preference—weak, medium, strong

I have implemented all these Fuzzy Rule Based System on MATLAB. Between the given source and destination, the mode of transport and the route is chosen by selecting either the direct or indirect approach. After determining the approach, the path preference is determined by processing all the Fuzzy Rule Based System.

3 MATLAB Implementation of Proposed System

For the implementation of the route selection Fuzzy Rule Based System, the following steps need to be done:

1. Defining the number of modes of transport.
2. Defining certain variables for each mode of transport.
3. Defining the source and destination.

4. Defining the existing routes between source and destination including direct routes. If no direct route exists, then all the indirect routes should be defined.
5. Defining the rule base that contains all the rules built by considering all the factors of particular mode.
6. Drawing the inferences from the rule viewer by considering different conditions.

I have implemented the Fuzzy Rule Based System on MATLAB which will work with a user interface that will supply input of origin, destination, and other necessary details like time of travel in different modes of travel for reaching destination, waiting period (if no direct route is available), and level of comfort that may be experienced during the journey. Also, a database including all the existing modes and routes works in the background. The proposed Fuzzy Rule Based System identifies all the routes in different modes and computes the path preference.

3.1 Direct Routing Fuzzy Rule Based System

3.1.1 Roadway Fuzzy Inference System

Roadway Fuzzy inference system is implemented as Road_Fuzzy Rule Based System using FIS editor. The input variables also called as FIS variables (Fuzzy Inference System variables) are the distance, road_condition and traffic_condition. The output is determined by using Mamdani method. This output will specify the path preference. The input variables are specified by using Membership Function editor in which the membership functions are chosen and the range of input is specified between [0,10]. Each input variable is defined by some parameters and each parameter has some value associated with it. The input variables are plotted using gaussmf function. The values for the input parameters for the corresponding FIS variables are as follows (Fig. 3):

Road condition—Bad—[1.5 0], Medium—[1.5 5], Good—[1.5 10]

Distance—Small—[1.5 0], Medium—[1.5 5], Large—[1.5 10]

Traffic condition—Bad—[1.5 0], Medium—[1.5 5], Good—[1.5 10] (Figs. 4, 5).

The output variable pref is plotted as trimf function taking collection of three points. The values of pref are taken as follows:

Weak—[0.0216 1.69 3.35]

Medium—[3.333 6.667]

Strong—[6.69 8.36 10].

After defining all the values for input and output variables, Rules are derived using the variables and their values by using Rule Editor. These rules form the fuzzy rule base. Different conditions are considered by taking the parameter values and accordingly, the pref output is defined. The ‘and’ operator is taken between the

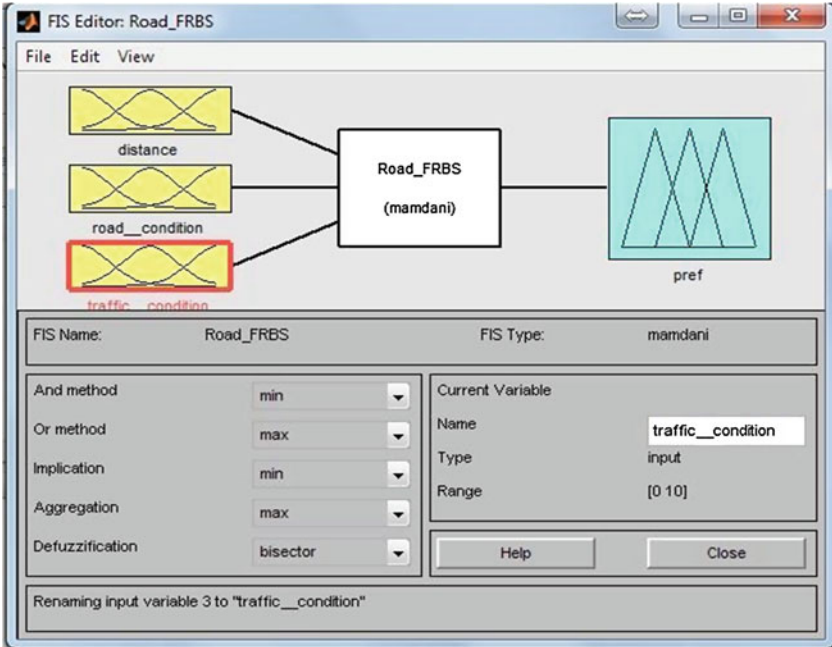


Fig. 3 Road fuzzy rule based system, FIS editor

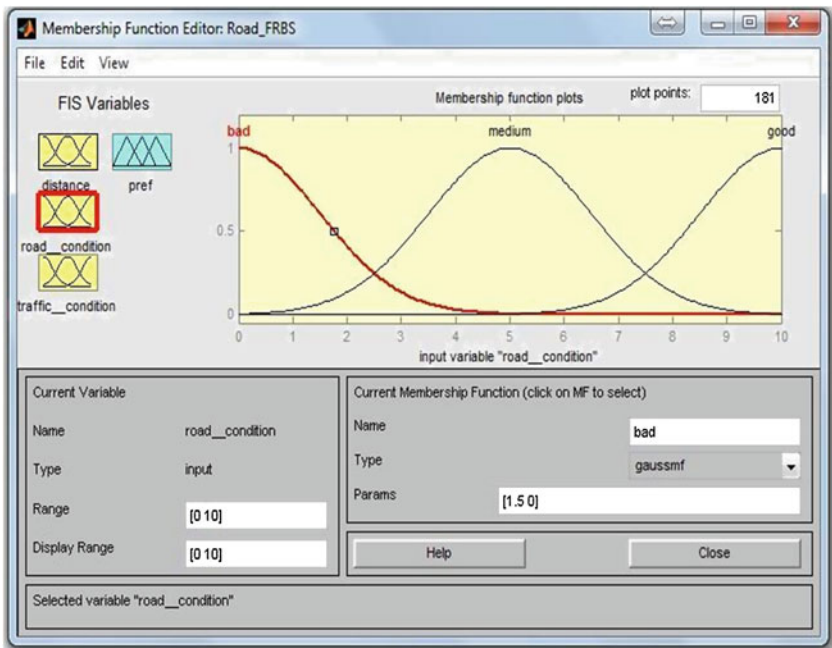


Fig. 4 Membership function editor of road fuzzy rule based system describing 'road_condition'

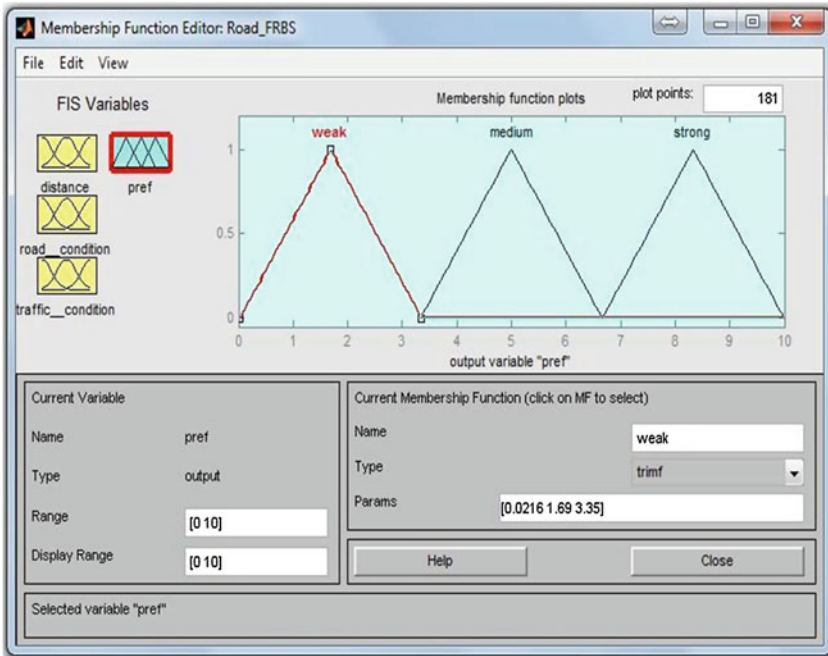


Fig. 5 Membership function editor of road fuzzy rule based system describing ‘pref’

input parameters as this will serve as maximum occurring condition which effect the resulting output (Fig. 6).

After the rules are done, the rule viewer will determine the path preference according to different values of the input variables and the defined rules. By changing the values in the rule viewer, the results are found at the same time. The values of input and output variables lie between 0 and 10. Here when distance is small i.e. 2.35, road condition is 6.81, which means more than medium and traffic condition is 7.77 i.e. more than medium, then road preference will be near to strong i.e. 6.6.

Railway and Airway Fuzzy Rule Based System are implemented in the same manner considering distance and comfort factor as input variable and path preference as output. The membership functions are designed for them in the same way as above but the rules are different.

3.2 Indirect Routing Fuzzy Rule-Based System

In Indirect Fuzzy Rule Based System, the rules are formed by combining the modes of transport. Distance, comfort factor and waiting period are taken as input variables, on the basis of which the path preference is determined. Membership

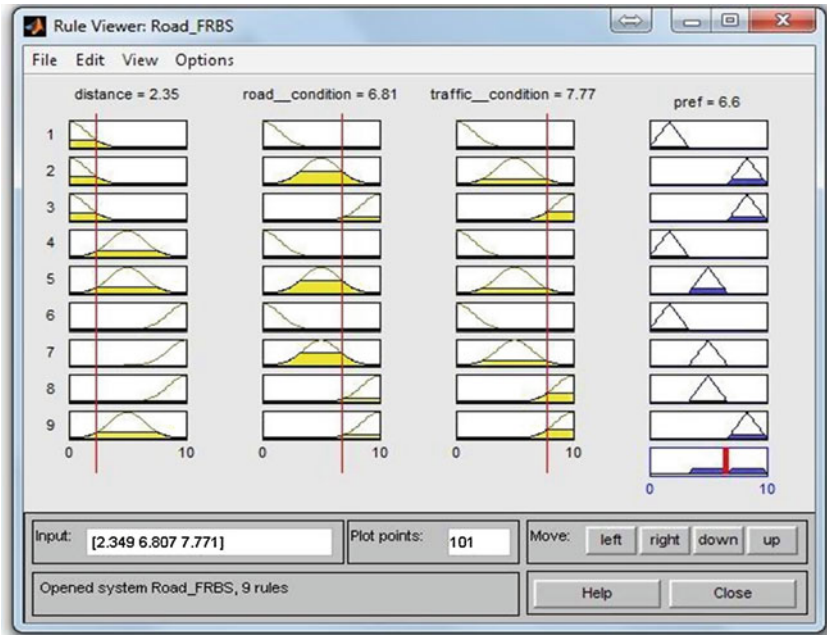


Fig. 6 Rule viewer of road fuzzy rule based system

functions are formed for each input variable using the Membership Function Editor and rules are formed in rule editor.

According to the rule base, rule viewer will determine the percentage of path preference. The percentage of path preference is compared to all other Fuzzy Rule Based System outputs and highest percentage of path preference is treated as the strongest path preference, which will be given as output of the system.

4 Conclusion

Fuzzy logic could be used successfully to model situations in which people have to make choice amongst multiples of decisions depending on underlying conditions. In present work I have developed a Route selection fuzzy rule based system which considers all the perspectives related to the selection of optimal route and mode of transport. The approach that is used is based on the availability of the routes between origin and the destination. The system identifies all the variables affecting the preference of route and mode. Rules are made according to the preference of the decision maker. Fuzzy inference system built from the determined rules compute the path preferences in all the modes of transport considering all the existing routes.

The implemented Fuzzy Rule Based System can be used in various transport applications like Trip generation, Route choice, etc. With increasing traffic in tourist and business activities the fuzzy logic applications are bound to be used for solving practical problems of mode and route selection.

References

1. Teodorovic, D.: Fuzzy logic systems for transportation engineering: the state of the art. *Transp. Res.: Part A* **33**, 337–364 (1999)
2. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: Introduction to fuzzy logic using MATLAB
3. Kammoun, H.M., Kallel, I., Alimi, A.M: Hierarchical fuzzy rule-base system for multiagent route choice: new developments in robotics automation and control. In: Lazinica, A. (ed.) ISBN: 978-953-7619-20-6. InTech, Winchester (2008)
4. Tao, W., Gang, W: A combined optimization model for transportation modes of multimodal transport. *Eng. Sci.* 46–50 (2005)
5. Jiang, J., Lu, J.: Research on optimal transport mode combination in a multimodal transportation system of containers, pp. 127–134 (2008)
6. Henn, V.: Fuzzy route choice model for traffic assignment. *Fuzzy Sets Syst.* **116**(1), 77–101, ISSN: 0165-0114 (2000)
7. Ridwan, M.: Fuzzy preference based traffic assignment problem. *Transp. Res. Part C: Emerg. Technol.* **12**(3–4), 209–233, ISSN: 0968-090X (2004)
8. Peeta, S., Yu, J.W.: Adaptability of hybrid route choice model to incorporating driver behaviour dynamics under information provision. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **34**(2), 243–256, ISSN: 1083-4427 (2004)
9. Qingbin, W., Zengxia, H: The optimal routes and modes selection in container multimodal transportation networks. In: IEEE International Conference on Optoelectronics and Image Processing, pp. 573–576 (2010)
10. Teodorovic, D., Kikuchi, S.: Transportation route choice model using fuzzy inference technique. In: Ayyub, B.M. (ed.) *Proceedings of ISUMA '90, The First International Symposium on Uncertainty Modeling and Analysis*. IEEE Computer Press, College Park, Maryland, pp. 140–145 (1990)
11. Fuzzy logic system: <http://www.mathworks.in/products/fuzzy-logic/description5.html>
12. Mendel, J.: Fuzzy logic systems for engineering: a tutorial. *Proc. IEEE* **83**, 345–377 (1995)

An Integrated Production Model in Fuzzy Environment Under Inflation

S. R. Singh and Swati Sharma

Abstract In this research article, an integrated system with variable production is developed in fuzzy environment. The demand rate is regarded as an increasing function of time. In developing the proposed model, it is assumed that the manufacturer takes the raw materials in multiple batches from the supplier, process them to produce finished products and deliver to the buyer in multiple shipments. The effect of inflation and time value of money is also taken into consideration. All the cost parameters are considered as triangular fuzzy numbers and signed distance method is used for defuzzification. The model is illustrated with numerical example and sensitivity analysis with respect to the system parameters is also executed.

Keywords Integrated · Production · Signed distance · Imprecise costs · Inflation

1 Introduction

Many traditional inventory models are implemented in static surroundings, where different inventory parameters are presumed to be known precisely, for the sake of easiness. Due to, the elevated competitiveness present in the markets it is complicated to define different inventory cost parameters precisely. To overcome this obscurity, academician and practitioners used fuzzy set theory. Initially, Zadeh [1], introduced the fuzzy set theory, after that Park [2] extended the traditional economic order quantity (EOQ) model by considering the fuzziness of ordering cost

S. R. Singh (✉) · S. Sharma
Department of Mathematics, D. N. College, Meerut 250001, India
e-mail: shivrajpundir@gmail.com

S. Sharma
e-mail: jmlashi0@gmail.com

and holding cost. After that, a lot of research work has been done on the defuzzification techniques. Later on, Chang [3] put forwarded an EOQ model by using fuzzy set theory. Some modern work in this direction is done by Singh et al. [4], Yadav et al. [5].

In many business practices it has been recognized that a coordinated supply chain, can improve service quality through capable information sharing and be able to minimize the allied costs of the supply chain system. Goyal [6] introduced the concept of integrated optimization for buyer and vendor. After that, Banerjee [7] developed a model with lot-for-lot policy and finite production rate. Later on, Banerjee and Kim [8] proposed an integrated just in time model with constant and deterministic production rate, demand rate and delivery time. Recently, Kumar et al. [9] and Omar et al. [10] discussed some motivating supply chain models with different hypotheses.

Moreover, inflation and time value of money affect the purchasing power of money; so these should not be overlooked. Several inventory models have been examined under the effect of inflation. Buzacott [11] developed an economic order quantity (EOQ) model by considering inflation for all associated costs. After that, Bierman and Thomas [12] proposed an EOQ model with inflation also incorporating discount rate. Later on, Yang et al. [13] discussed inventory models with fluctuating demand and inflation. Recently, Singh et al. [14], Singh and Swati [15], Singh et al. [16] established some remarkable inventory models considering the effect of inflation.

Most of the previous supply chain inventory models are developed in crisp environment with deterministic production rate and not including the effect of inflation. However, due to the high inflationary surroundings and fluctuating market conditions these realistic circumstances should not be ignored. In this study, an integrated model is conferred in fuzzy environment by considering variable demand, demand dependent production rate and incorporating the effect of inflation and time value of money. This paper extends the work of Omar et al. [10] by considering linearly increasing demand rate, variable production and inflation in fuzzy environment. A numerical example is provided to demonstrate the model and sensitivity analysis is also performed. Finally, the paper is concluded and further researches are suggested.

2 Assumptions and Notations

In this model, we have considered a manufacturer who purchases raw material from a supplier and produces the finished goods with a variable production rate and delivers these items to a buyer either at equal replenishment interval or at equal consignment bulk.

The following assumptions are made to develop the mathematical model:

1. The production rate is $P = \lambda f(t)$ where $\lambda > 1$ for all t .
2. During the production up-time, the finished goods becomes immediately exists to meet up the demand.
3. The consumption rate of finished goods at any time t during $(0, T)$ is $f(t)$ and assumed to be linearly increasing with time.
4. Shortages are not allowed.
5. A single product inventory system is considered over a finite planning horizon.
6. Only single raw material is considered.

The following notations are used in developing the model:

- C_p is the manufacturing set-up cost
- H_p is the inventory carrying cost per unit per unit time for finished item/product at the manufacturer side
- C_1 is the ordering cost for raw material
- H_1 is the inventory carrying cost per unit per unit time for raw material
- C_b is the shipment cost
- H_b is the inventory carrying cost per unit per unit time at the buyer's side
- k is the discount rate
- i is the inflation rate
- r ($= k - i$) is the discount rate minus the inflation rate
- m is the number of raw material batches from supplier to the manufacturer
- n is the number of shipment from manufacturer to the buyer.

3 Mathematical Formulation and Solution

3.1 Buyer's Inventory System

The buyer's inventory system is depicted in Fig. 1. During the time period $[0, T]$ there are 'n' cycles at the buyer's end and two policies: (1) when time for each shipment is same and (2) when lot size for each shipment is same. X is the amount of inventory which is needed by the buyer at the beginning of a production cycle to meet the customer demand that also provides the manufacturer adequate time to build up inventory for future deliveries. In every lot, due to the customer's demand buyer's inventory reduces up to zero. The stock level during any period $[t_i, t_{i+1}]$, where $i = 0, 1, 2, \dots, n - 1$ is given by the differential equation:

$$I'(t) = -f(t), \quad t_i \leq t \leq t_{i+1} \quad (1)$$

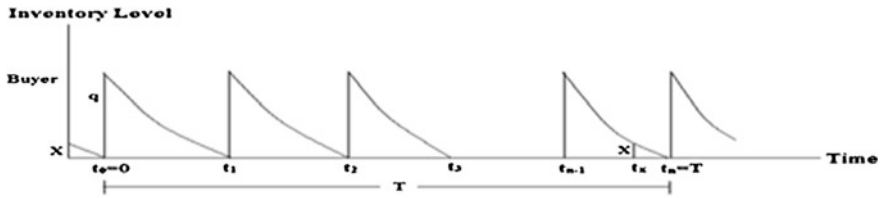


Fig. 1 Buyer’s inventory system for equal lots shipments size q

With boundary condition $I(t_i) = q$. Where q is given by

$$q = \int_{t_i}^{t_{i+1}} f(t)dt \tag{2}$$

Integrating Eq. (1) from t_i to t and using Eq. (2), we get

$$I(t) = q - \int_{t_i}^t f(u)du = \int_t^{t_{i+1}} f(u)du \tag{3}$$

The present worth of the shipment cost (OC_b) for the buyer for n -shipments is

$$OC_b = \sum_{i=0}^{n-1} C_b e^{-rt_i} \tag{4}$$

The present worth of the holding cost (HC_b) for n -shipments for the buyer is

$$HC_b = H_b \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} I(t)e^{-rt}dt = \frac{H_b}{r} \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} f(t)(e^{-rt_i} - e^{-rt})dt$$

When demand rate $f(t)$ is $(a + bt)$ then present worth of the holding cost becomes

$$HC_b = \frac{H_b}{r} \sum_{i=0}^{n-1} \left[e^{-rt_i} \left(at_{i+1} + \frac{bt_{i+1}^2}{2} - at_i - \frac{bt_i^2}{2} \right) + \frac{e^{-rt_{i+1}}}{r^2} \{ (a + bt_{i+1})r + b \} - \frac{e^{-rt_i}}{r^2} \{ (a + bt_i)r + b \} \right] \tag{5}$$

When we consider second policy, where for n -shipments we have

$$q = \frac{1}{n} \int_0^T f(t)dt \tag{6}$$

It follows that for any shipment cycle

$$q = \int_{t_i}^{t_{i+1}} f(t)dt \tag{7}$$

When demand rate is $(a + bt)$, then from Eqs. (6), (7), we have

$$t_i = \left[-a + \sqrt{(a^2n + 2abiT + ib^2T^2)/n} \right] / b \tag{8}$$

Now, we compute present worth of the holding cost (HC_{bt}) for policy 1. In this case, $t_i = iT/n$, so from Eq. (5), we get

$$HC_{bt} = \frac{H_b}{r} \left[\left\{ \frac{aT}{n} + \frac{bT^2}{2n^2} - \frac{(ar + b)}{r^2} + \frac{e^{-rT/n}}{r^2} \left(\left(a + \frac{bT}{n} \right) r + b \right) \right\} + \left\{ \frac{bT^2}{n^2} + \left(e^{-rT/n} - 1 \right) \frac{bT}{nr} \right\} \frac{e^{-rT/n}}{(1 - e^{-rT/n})} \right] \frac{(1 - e^{-rT})}{(1 - e^{-rT/n})} \tag{9}$$

Similarly, from Eq. (5) the present worth of the holding cost (HC_{bq}) for policy 2 is

$$HC_{bq} = \frac{H_b}{r} \sum_{i=0}^{n-1} \left[e^{-rt_i} \left(at_{i+1} + \frac{bt_{i+1}^2}{2} - at_i - \frac{bt_i^2}{2} \right) + \frac{e^{-rt_{i+1}}}{r^2} \{ (a + bt_{i+1})r + b \} - \frac{e^{-rt_i}}{r^2} \{ (a + bt_i)r + b \} \right] \tag{10}$$

where, t_i is given by Eq. (8).

So, present worth of the total average cost at the buyer’s end (TCB_t) for policy 1 is

$$TCB_t = \left[\{ C_b(1 - e^{-rT}) / (1 - e^{-rT/n}) \} + HC_{bt} \right] / T \tag{11}$$

Now, the present worth of the total average cost for the second policy is given by

$$TCB_q = \left[\sum_{i=0}^{n-1} C_b e^{-rt_i} + HC_{bq} \right] / T \tag{12}$$

3.2 Manufacturer’s Inventory System

Manufacturer’s Inventory System for finished Goods.

Figure 2 shows the manufacturing inventory level during the time period $[0, T]$. Initially, at $t = 0$ production starts and during the production period $[0, t_p]$ inventory level increases due to the combined effect of the production and demand. At time t_p production stops and after that the inventory level decreases during the

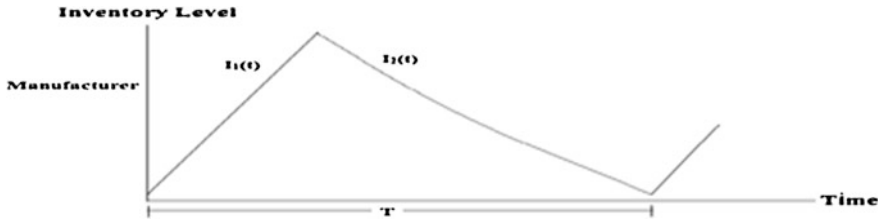


Fig. 2 Manufacturer’s inventory system (for finished goods) for equal lots shipments size q

period $[t_p, T]$ due to the demand of the products. The inventory level at any time for finished goods for the manufacturer is given by the following differential equations:

$$I_1'(t) = P - f(t) = (\lambda - 1)f(t), \quad 0 \leq t \leq t_p \tag{13}$$

$$I_2'(t) = -f(t), \quad t_p \leq t \leq T \tag{14}$$

With boundary conditions $I_1(0) = 0$ and $I_2(T) = 0$. After solving Eqs. (13), (14) we get

$$I_1(t) = (\lambda - 1) \int_0^t f(u)du, \quad 0 \leq t \leq t_p \tag{15}$$

$$I_2(t) = \int_t^T f(u)du, \quad t_p \leq t \leq T \tag{16}$$

Now, the total production quantity is equal to the total demand fulfilled during the period $[0, T]$, so we have

$$\int_0^{t_p} \lambda f(t)dt = \int_0^T f(t)dt \tag{17}$$

After solving Eq. (17) when $f(t) = a + bt$, we get

$$t_p = \left[-a + \sqrt{(a^2\lambda + bT(2a + bT))/\lambda} \right] / b \tag{18}$$

The quantity X is given by

$$X = \int_{t_x}^T f(t)dt \tag{19}$$

Now, for the first policy, we have

$$\int_{t_x}^T \lambda f(t)dt = \int_0^{T/n} f(t)dt \tag{20}$$

When $f(t) = a + bt$ then, from Eq. (20), we get

$$t_x = \left[-a + \sqrt{\{a^2n^2\lambda + 2abnT(n\lambda - 1) + b^2T^2(n^2\lambda - 1)\}/n^2\lambda} \right] / b \tag{21}$$

For the second policy, we have

$$\int_{t_x}^T \lambda f(t)dt = \frac{1}{n} \int_0^T f(t)dt \tag{22}$$

When $f(t) = a + bt$ then, from Eq. (22), we obtain

$$t_x = \left[-a + \sqrt{\{a^2n\lambda + (b^2T^2 + 2abT)(n\lambda - 1)\}/n\lambda} \right] / b \tag{23}$$

The present worth of the set-up cost (SC_m) for the manufacturer is

$$SC_m = C_p \tag{24}$$

Therefore, the present worth of the holding cost (HC_m) for the manufacturer is

$$HC_m = H_p \left[\int_0^T X e^{-rt} dt + \int_0^{t_p} I_1(t) e^{-rt} dt + \int_{t_p}^T I_2(t) e^{-rt} dt \right] \tag{25}$$

When $f(t) = a + bt$, then from Eq. (25) present worth of the holding cost (HC_{mt}) for policy 1 is

$$\begin{aligned} HC_{mt} = H_p & \left[\frac{(1 - e^{-rT})}{r} \left\{ a(T - t_x) + \frac{b}{2}(T^2 - t_x^2) \right\} \right. \\ & + \frac{(\lambda - 1)}{r} \left\{ \frac{a}{r} + \frac{b}{r^2} - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} + at_p + \frac{bt_p^2}{2} \right) \right\} \\ & \left. + \frac{1}{r} \left\{ e^{-rt} \left(aT + \frac{bT^2}{2} - at_p - \frac{bt_p^2}{2} \right) + e^{-rT} \left(\frac{(a + bT)}{r} + \frac{1}{r^2} \right) - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} \right) \right\} \right] \tag{26} \end{aligned}$$

where, t_p and t_x for the first policy are given by the Eqs. (18) and (21) respectively. Similarly, the present worth of the holding cost (HC_{mq}) for second policy is

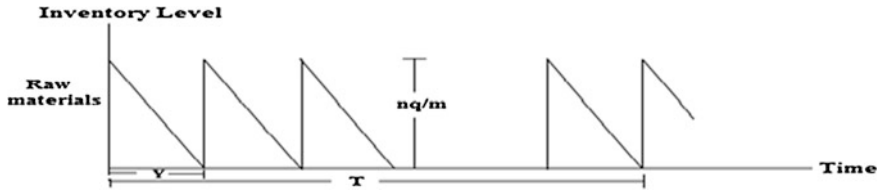


Fig. 3 Manufacturer’s inventory system for raw materials for equal lots shipments size q

$$\begin{aligned}
 HC_{mq} = H_p & \left[\frac{(1 - e^{-rT})}{r} \left\{ a(T - t_x) + \frac{b}{2} (T^2 - t_x^2) \right\} + \frac{(\lambda - 1)}{r} \left\{ \frac{a}{r} + \frac{b}{r^2} \right. \right. \\
 & \left. \left. - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} + at_p + \frac{bt_p^2}{2} \right) \right\} + \frac{1}{r} \left\{ e^{-rt_p} \left(aT + \frac{bT^2}{2} - at_p - \frac{bt_p^2}{2} \right) \right. \right. \\
 & \left. \left. + e^{-rT} \left(\frac{(a + bT)}{r} + \frac{1}{r^2} \right) - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} \right) \right\} \right] \tag{27}
 \end{aligned}$$

where, t_p and t_x are given by the Eqs. (18) and (23) respectively.

Manufacturer’s Inventory System for Raw Materials

Manufacturer’s inventory system for raw material is depicted in Fig. 3. There are ‘m’ shipments during the period [0, T] of equal lot size with equal time interval Y. The shipment lot size diminishes owing to the production. Now, the inventory level at any time for the raw material is given by the differential equation:

$$I'(t) = -\lambda f(t), \quad 0 \leq t \leq Y \tag{28}$$

$$I(0) = \frac{nq}{m} = \frac{1}{m} \int_0^T f(t)dt \tag{29}$$

After, solving Eq. (28), with the help of Eq. (29), we get

$$I(t) = (nq/m) - \int_0^t \lambda f(u)du, \quad 0 \leq t \leq Y \tag{30}$$

Since $I(Y) = 0$, so we have

$$Y = \left[-a + \sqrt{\{a^2m\lambda + bT(2a + bT)\}/m\lambda} \right] / b \tag{31}$$

The present worth of the holding cost for the raw material (HCR_m) is

$$\begin{aligned}
 HCR_m = H_1 \sum_{i=0}^{m-1} e^{-niY} \left[\int_0^Y I(t)e^{-rt}dt \right] & = H_1 \sum_{i=0}^{m-1} e^{-niY} \left[\frac{(1 - e^{-rY})}{mr} \int_0^T f(t)dt + \frac{\lambda}{r} \int_0^Y (e^{-rY} - e^{-rt})f(t)dt \right] \\
 & \tag{32}
 \end{aligned}$$

$$\begin{aligned}
 HCR_m = H_1 & \left[\frac{(1 - e^{-rY})}{mr} \left(aT + \frac{bT^2}{2} \right) + \frac{\lambda}{r} \left\{ \left(aT + \frac{bT^2}{2} \right) e^{-rY} \right. \right. \\
 & \left. \left. + \left(\frac{(a + bY)}{r} + \frac{b}{r^2} \right) e^{-rY} - \frac{a}{r} - \frac{b}{r^2} \right\} \right] \frac{(1 - e^{-rmY})}{(1 - e^{-rY})} \tag{33}
 \end{aligned}$$

The present worth of the ordering cost for raw material (OCR_m) is

$$OCR_m = \sum_{i=0}^{m-1} C_1 e^{-riY} = [C_1(1 - e^{-mY})]/(1 - e^{-rY}) \tag{34}$$

The present worth of the total cost per unit time for the raw material (TCR_m) is

$$TCR_m = [OCR_m + HCR_m]/T \tag{35}$$

Now, the present worth of the total cost per unit time (TCM_t) for the manufacturer when the first policy is considered, is

$$TCM_t = [(SC_m + HC_{mt})/T] + TCR_m \tag{36}$$

Similarly, the present worth of the total cost per unit time (TCM_q) for the manufacturer when the second policy is considered, is

$$TCM_q = [(SC_m + HC_{mq})/T] + TCR_m \tag{37}$$

The present worth of the total average cost (TC_t) for the system, for policy 1 is

$$TC_t = TCM_t + TCB_t = [C_p + H_p F_1 + (C_1 + H_1 F_2)G_1 + (C_b + H_b F_3)G_2]/T \tag{38}$$

$$\begin{aligned}
 F_1 &= \left[\frac{(1 - e^{-rT})}{r} \left\{ a(T - t_x) + \frac{b}{2} (T^2 - t_x^2) \right\} + \frac{(\lambda - 1)}{r} \left\{ \frac{a}{r} + \frac{b}{r^2} - e^{-rT} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} + at_p + \frac{bt_p^2}{2} \right) \right\} \right. \\
 & \left. + \frac{1}{r} \left\{ e^{-rt_p} \left(aT + \frac{bT^2}{2} - at_p - \frac{bt_p^2}{2} \right) + e^{-rT} \left(\frac{(a + bT)}{r} + \frac{1}{r^2} \right) - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} \right) \right\} \right] \\
 F_2 &= \left[\frac{(1 - e^{-rY})}{mr} \left(aT + \frac{bT^2}{2} \right) + \frac{\lambda}{r} \left\{ \left(aT + \frac{bT^2}{2} \right) e^{-rY} + \left(\frac{(a + bY)}{r} + \frac{b}{r^2} \right) e^{-rY} - \frac{a}{r} - \frac{b}{r^2} \right\} \right] \\
 F_3 &= \frac{H_b}{r} \left[\left\{ \frac{aT}{n} + \frac{bT^2}{2n^2} - \frac{(ar + b)}{r^2} + \frac{e^{-rT/n}}{r^2} \left(\left(a + \frac{bT}{n} \right) r + b \right) \right\} + \left\{ \frac{bT^2}{n^2} + (e^{-rT/n} - 1) \frac{bT}{nr} \right\} \frac{e^{-rT/n}}{(1 - e^{-rT/n})} \right] \\
 G_1 &= (1 - e^{-mY})/(1 - e^{-rY}), \quad G_2 = (1 - e^{-rT})/(1 - e^{-rT/n})
 \end{aligned}$$

Again, the present worth of the total cost per unit time (TC_q) for the system when the second policy is deliberated, is

$$TC_q = TCM_q + TCB_q = [C_p + H_p K_1 + (C_1 + H_1 F_2)G_1 + (C_b K_2 + H_b K_3)]/T \tag{39}$$

$$\begin{aligned}
 K_1 &= \left[\frac{(1 - e^{-rT})}{r} \left\{ a(T - t_x) + \frac{b}{2} (T^2 - t_x^2) \right\} + \frac{(\lambda - 1)}{r} \left\{ \frac{a}{r} + \frac{b}{r^2} - e^{-rT} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} + at_p + \frac{bt_p^2}{2} \right) \right\} \right. \\
 & \left. + \frac{1}{r} \left\{ e^{-rt_p} \left(aT + \frac{bT^2}{2} - at_p - \frac{bt_p^2}{2} \right) + e^{-rT} \left(\frac{(a + bT)}{r} + \frac{1}{r^2} \right) - e^{-rt_p} \left(\frac{(a + bt_p)}{r} + \frac{b}{r^2} \right) \right\} \right]
 \end{aligned}$$

$$K_2 = \sum_{i=0}^{n-1} e^{-rt_i}, K_3 = \frac{1}{r} \sum_{i=0}^{n-1} \left[e^{-rt_i} \left\{ \left(at_{i+1} + \frac{bt_{i+1}^2}{2} - at_i - \frac{bt_i^2}{2} \right) - \frac{\{(a + bt_i)r + b\}}{r^2} \right\} + \frac{e^{-rt_{i+1}} \{(a + bt_{i+1})r + b\}}{r^2} \right].$$

3.3 Fuzzy Mathematical Model

In this study, we consider C_p, H_p, C_1, H_1, C_b and H_b are imprecise and expressed by triangular fuzzy numbers i.e., $\tilde{C}_p, \tilde{H}_p, \tilde{C}_1, \tilde{H}_1, \tilde{C}_b$ and \tilde{H}_b .

$$\tilde{C}_p = (C_p - \Delta_1, C_p, C_p + \Delta_2), \tilde{H}_p = (H_p - \Delta_3, H_p, H_p + \Delta_4), \tilde{C}_1 = (C_1 - \Delta_5, C_1, C_1 + \Delta_6) \\ \tilde{H}_1 = (H_1 - \Delta_7, H_1, H_1 + \Delta_8), \tilde{C}_b = (C_b - \Delta_9, C_b, C_b + \Delta_{10}), \tilde{H}_b = (H_b - \Delta_{11}, H_b, H_b + \Delta_{12})$$

where $0 < \Delta_1 < C_p$ and $\Delta_1 \Delta_2 > 0, 0 < \Delta_3 < H_p$ and $\Delta_3 \Delta_4 > 0, 0 < \Delta_5 < C_1$ and $\Delta_5 \Delta_6 > 0, 0 < \Delta_7 < H_1$ and $\Delta_7 \Delta_8 > 0, 0 < \Delta_9 < C_b$ and $\Delta_9 \Delta_{10} > 0, 0 < \Delta_{11} < H_b$ and $\Delta_{11} \Delta_{12} > 0$.

Here, we considered the signed distance method for defuzzification as used by Chang [3]. So, the signed distance for C_p is given by

$$d(\tilde{C}_p, \tilde{0}) = [(C_p - \Delta_1) + 2C_p + (C_p + \Delta_2)]/4 = [C_p + (\Delta_2 - \Delta_1)/4]$$

where $d(\tilde{C}_p, \tilde{0}) > 0$ and $d(\tilde{C}_p, \tilde{0}) \in [C_p - \Delta_1, C_p + \Delta_2]$.

Similarly, signed distance for other parameters can be defined.

Now, average total cost for policy 1 is $TC_t = (TC_t^1, TC_t^2, TC_t^3)$, where

$$TC_t^1 = [(C_p - \Delta_1) + (H_p - \Delta_3)F_1 + ((C_1 - \Delta_5) + (H_1 - \Delta_7)F_2)G_1 + ((C_b - \Delta_9) + (H_b - \Delta_{11})F_3)G_2]/T$$

$$TC_t^2 = [C_p + H_p F_1 + (C_1 + H_1 F_2)G_1 + (C_b + H_b F_3)G_2]/T$$

$$TC_t^3 = [(C_p + \Delta_2) + (H_p + \Delta_4)F_1 + ((C_1 + \Delta_6) + (H_1 + \Delta_8)F_2)G_1 + ((C_b + \Delta_{10}) + (H_b + \Delta_{12})F_3)G_2]/T$$

Then, defuzzified total cost per unit time for policy 1 is given by

$$T\tilde{C}_t^f = (TC_t^1 + 2TC_t^2 + TC_t^3)/4 \tag{40}$$

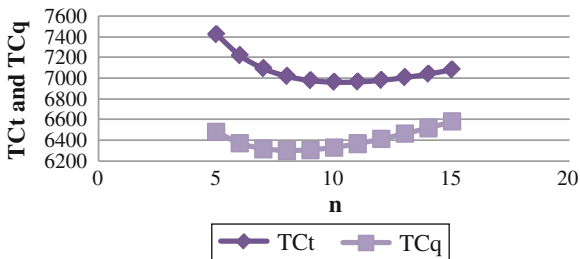
Again, average total cost for policy 2 is $T\tilde{C}_q = (TC_q^1, TC_q^2, TC_q^3)$, where

$$TC_q^1 = [(C_p - \Delta_1) + (H_p - \Delta_3)K_1 + ((C_1 - \Delta_5) + (H_1 - \Delta_7)F_2)G_1 + ((C_b - \Delta_9)K_2 + (H_b - \Delta_{11})K_3)]/T$$

Table 1 Minimum total costs, TC_t and TC_q for different arrangement of m and n

m	n						
	6	7	8	9	10	11	12
1	7,157.37	7,108.51	7,100.05	7,118.64	7,156.00	7,206.75	7,267.23
	(6,432.89)	(6,446.40)	(6,487.65)	(6,546.97)	(6,618.43)	(6,698.24)	(6,783.84)
2	7,218.29	7,092.55	7,017.56	6,977.26	6,961.59	6,963.95	6,979.83
	(6,369.99)	(6,316.22)	(6,298.80)	(6,305.82)	(6,329.89)	(6,366.18)	(6,411.43)
3	7,530.94	7,355.93	7,239.48	7,163.38	7,116.19	7,090.37	7,080.72
	(6,595.95)	(6,501.07)	(6,448.76)	(6,425.40)	(6,422.50)	(6,434.50)	(6,457.60)

Fig. 4 Behavior of the cost functions when $m = 2$ against n for both policies



$$TC_q^2 = [C_p + H_p K_1 + (C_1 + H_1 F_2)G_1 + (C_b K_2 + H_b K_3)]/T$$

$$TC_q^3 = [(C_p + \Delta_2) + (H_p + \Delta_4)K_1 + ((C_1 + \Delta_6) + (H_1 + \Delta_8)F_2)G_1 + ((C_b + \Delta_{10})K_2 + (H_b + \Delta_{12})K_3)]/T$$

Then, defuzzified total cost per unit time for policy 2 is given by

$$T\tilde{C}_q^f = (TC_q^1 + 2TC_q^2 + TC_q^3)/4 \tag{41}$$

4 Numerical Example and Sensitivity Analysis

Let us consider a numerical example with data as in Omar et al. [10] and including some others: $a = 7000, b = 700, C_p = 200, \Delta_1 = 20, \Delta_2 = 40, C_1 = 80, \Delta_5 = 8, \Delta_6 = 16, C_b = 30, \Delta_9 = 3, \Delta_{10} = 6, H_p = 5, \Delta_3 = 0.5, \Delta_4 = 1, H_1 = 3, \Delta_7 = 0.3, \Delta_8 = 0.6, H_b = 8, \Delta_{11} = 0.8, \Delta_{12} = 1.6, \lambda = 1.5, r = 0.09$ and $f(t) = a + bt$. The best solution for policy 1 is $m = 2, n = 10, T = 0.194105, TC_t = 6,961.59$ and for policy 2 is $m = 2, n = 8, T = 0.194625, TC_q = 6,298.80$, it is revealed through Table 1. The total production quantity for policy 1 is 1,371.92 and for policy 2 is 1,375.63. The software MATHEMATICA 8.0 is used for computations. The behavior of the average cost for both policies regarding n is shown in Fig. 4 and sensitivity analysis is given through Tables 2, 3, 4, 5.

Table 2 Effect of parameter C_p on TC_t and TC_q

C_p	m	$n(TC_t)$	$n(TC_q)$	TC_t	TC_q
100	1	6	5	6,352.73	5,682.05
200	2	10	8	6,961.59	6,298.80
300	2	12	9	7,438.53	6,778.28

Table 3 Effect of parameter C_b on TC_t and TC_q

C_b	m	$n(TC_t)$	$n(TC_q)$	TC_t	TC_q
15	2	14	11	6,051.53	5,577.86
30	2	10	8	6,961.59	6,298.80
45	2	9	7	7,667.24	6,858.01

Table 4 Effect of parameter r on TC_t and TC_q

r	m	$n(TC_t)$	$n(TC_q)$	TC_t	TC_q
0.06	2	11	8	7,264.91	6,311.40
0.09	2	10	8	6,961.59	6,298.80
0.12	2	10	8	6,792.03	6,286.21

Table 5 Effect of parameter λ on TC_t and TC_q

λ	m	$n(TC_t)$	$n(TC_q)$	TC_t	TC_q
1.5	2	10	8	6,961.59	6,298.80
2	1	7	6	7,157.82	6,462.51
2.5	1	7	5	7,187.48	6,465.00

Some interesting inferences drawn from the sensitivity tables are given as follows:

1. Tables 2, 3 reveal that as the cost parameters C_p and C_b increase the average costs of the system increase. While, the system favors to the larger number of m and n when C_p increases and to the smaller number of n when C_b increases.
2. Table 4 shows that the total relevant cost of the system decreases with an increase in inflation parameter r and the system favors to the smaller number of n .
3. Table 5 discloses that as the production parameter λ increases, the total average cost of the system increases and optimal solution suggests decreasing m and n .

5 Conclusion

In this article, an inventory model for managing a supply chain system is developed by considering variable production and demand rate. The model is discussed in fuzzy environment and the effect of inflation is also taken into deliberation.

The presented paper is different from the existing one as the previous models ignored the effect of inflation moreover; fixed cost parameters were considered which is not realistic due to the vagueness present in the market. This article is more sensible and helpful for decision-makers as the effect of inflation is considered which is convenient due to the high inflationary surroundings and by taking fuzziness in costs, decision-makers can absorb all the instability in the cost due to the market variation. The further research can be done by considering shortages, lead time and stochastic environment.

Acknowledgments The second author would like to thank to CSIR (New Delhi) for providing financial help in the form of JRF vide letter no. 08/017(0017)/2011-EMR-I.

References

1. Zadeh, L.A.: Fuzzy sets. *Inform. Cont.* **8**(3), 338–353 (1965)
2. Park, K.S.: Inventory models with partial backorders. *Int. J. Sys. Sci.* **13**, 1313–1317 (1982)
3. Chang, H.C.: An application of fuzzy sets theory to the EOQ model with imperfect quality items. *Comp. Oper. Res.* **31**, 2079–2092 (2004)
4. Singh, S.R., Kumari, R., Kumar, N.: Optimisation of fuzzy inventory model for differential items. *Int. J. Oper. Res.* **11**(3), 290–315 (2011)
5. Yadav, D., Singh, S.R., Kumari, R.: Retailer's optimal policy under inflation in fuzzy environment with trade credit. *Int. J. Sys. Sci.* (2013). <http://dx.doi.org/10.1080/00207721.2013.801094>
6. Goyal, S.K.: An integrated inventory model for a single supplier-single customer problem. *Int. J. Prod. Econ.* **15**(1), 107–111 (1976)
7. Banerjee, A.: A joint economic lot size model for purchaser and vendor. *Dec. Sci.* **17**, 292–311 (1986)
8. Banerjee, A., Kim, S.L.: An integrated JIT inventory model. *Int. J. Oper. Prod. Manag.* **15**, 237–244 (1995)
9. Kumar, N., Singh, S.R., Kumari, R.: Three-Echelon supply chain inventory model for deteriorating items with limited storage facility and lead-time under inflation. *Int. J. Serv. Oper. Manag.* **13**(1), 98–118 (2012)
10. Omar, M., Sarker, R., Othman, W.A.M.: A just-in-time three-level integrated manufacturing system for linearly time-varying demand process. *App. Math. Mod.* **37**, 1275–1281 (2013)
11. Buzacott, J.A.: Economic order quantities with inflation. *Oper. Res. Quar.* **26**(3), 553–558 (1975)
12. Bierman, H., Thomas, J.: Inventory decisions under inflationary conditions. *Dec. Sci.* **8**, 151–155 (1977)
13. Yang, H.L., Teng, J.T., Chern, M.S.: Deterministic inventory lot-size models under inflation with shortages and deterioration for fluctuating demand. *Nav. Res. Log.* **48**(2), 144–158 (2001)
14. Singh, S.R., Kumar, N., Kumari, R.: An inventory model for deteriorating items with shortages and stock-dependent demand under inflation for two-shops under one management. *Opsearch* **47**(4), 311–329 (2010)
15. Singh, S.R., Swati, C.C.: An optimizing policy for decaying items with ramp-type demand and preservation technology under trade credit and inflation. *Rev. Busi. Tech. Res* **5**(1), 54–62 (2012)
16. Singh, S.R., Jain, S., Pareek, S.: A warehouse imperfect fuzzified production model with shortages under inflationary conditions. *Adv. Dec. Sci.* (2012). doi:10.1155/2012/638060

R&D and Performance Analysis of Indian Pharmaceutical Firms: An Application of DEA

Varun Mahajan, D. K. Nauriyal and S. P. Singh

Abstract The purpose of the study is to examine the efficiency of R&D and non-R&D firms in Indian pharmaceutical firms from 2000 to 2010 comprising both pre and post product patent periods by applying data envelopment analysis technique. The analysis based on a panel sample data set of 141 firms, drawn from PROWESS database of CMIE, measures efficiency by using one output and four inputs. Net sales revenue is taken as output and raw material cost, salaries and wages, advertisement and marketing cost and capital cost as inputs. The study found that efficiency of R&D firms was higher than that of non-R&D intensive firms for all the years. Nevertheless, both types of firms were found to have a good scope for improvement in their resource use efficiency without affecting their level of output.

Keywords Pharmaceutical industry · R&D · Efficiency · DEA

1 Introduction

The Indian Pharmaceutical industry (IPI) is considered to be one of the most dynamic and a vibrant industry of the developing world as it is one of the largest producers and exporters of the generic drugs in the world. In terms of quality, technology and range of medicines, it is positioned very high in the developing

V. Mahajan (✉) · D. K. Nauriyal · S. P. Singh
Department of Humanities and Social Sciences, Indian Institute of Technology Roorkee,
Roorkee 247667, India
e-mail: mahajan.varun85@gmail.com

D. K. Nauriyal
e-mail: dk_nauriyal@yahoo.com

S. P. Singh
e-mail: singh2006_hss@yahoo.co.in

countries and is ranked third in terms of manufacturing pharmaceutical products by volume. Its annual turnover was estimated to be about Rs. 1,049.44 billion during the year 2010–2011. It is emerging as one of the major contributors to Indian exports with export earnings rising from a negligible amount in early 1990s to Rs. 475.51 billion in 2010–2011. The industry exports about 40 % of the production and the growth rate of Indian pharmaceutical exports far exceeded that of Indian exports as a whole. Over the past decade, it has witnessed 19.22 % compound annual growth rate in exports [1, 2].

It may be interesting to note that Indian pharmaceutical firms (mostly public sector firms) had a less than 20 % share of the market in 1970s which by 2010 rose over 75 %. Up to 1970, there was dominance of MNCs in the Indian pharmaceutical industry and bulk drugs and API (Active Pharmaceutical Ingredient) was imported. But the scenario had drastically altered after the enactment of Indian Process Patent Act of 1970 which led to phenomenal growth of indigenous firms. Since 1972 to 2004 with the process patent regime in place, Indian pharmaceutical producers were able to manufacture bulk drugs and formulations through ‘reverse engineering’ the patented medicines without any obligation to pay royalty. The period between 1995 and 2005 was the transition period as TRIPS agreement required India, as other developing countries, to compulsorily shift to product patent by January 1, 2005 in place of the previous process patent system established by the 1970 Patent Act. On March 22, 2005, India’s Parliament approved the Patent (Amendment) Act 2005, bringing in a system of product patents to be implemented with retrospective effects since January 1, 2005. The new regime protects only products arriving on the market after January 1, 1995. Since then, there is significant rise in R&D intensity, consolidations, mergers and acquisitions among Indian companies though the R&D intensity is still far lower than the multinationals [3, 4].

In view of the changed international business and regulatory environment, the focus of the Indian pharmaceutical firms is also shifting from process improvisation to R&D and new drug discovery as is evident from the fact that these firms are setting up their own R&D setups, allocating higher budgets for R&D and also collaborating with research laboratories, besides updating product quality and drug delivery systems in order to meet regulatory requirements, such as US FDA, of export markets [5]. The increasing R&D expenditure requirements and price competition have now pushing these firms to look beyond producing medicines and explore other avenues to survive and grow in an otherwise increasingly competitive market.

1.1 R&D Expenditure

The empirical research studies have brought to the fore, that R&D expenditure and patenting activity have dramatically increased especially after signing of the TRIPs agreement [3, 4]. TRIPS agreement not only has pushed up R&D

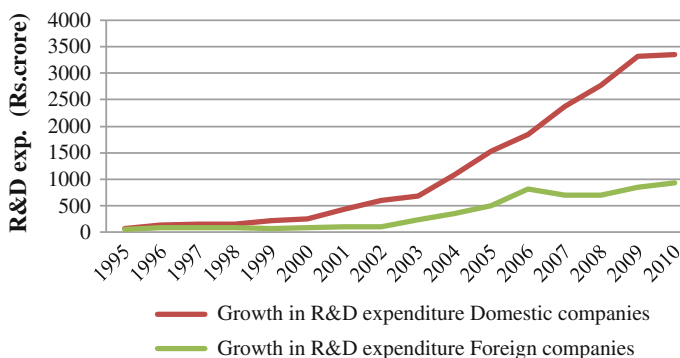


Fig. 1 Growth in R&D expenditure. *Source* GOI report (2012)

expenditures of the Indian pharmaceutical industry but has also altered R&D structure of the industry. Now focus of the firms has shifted towards greater R&D expenditure and the discovery of new chemical entities. Conventionally, most of the Indian pharmaceutical firms were focused on reverse engineering skills rather than new drug discovery and development. Consequently, the R&D intensity was very low as compared to 15–20 % of the sales revenue allocated towards this end by Western innovator companies [5, 6].

New drug discovery and development has never been on the agenda of the Indian pharmaceutical firms, as reflected in their output and from the fact that there was almost marginal investment in R&D with some firms such as Ranbaxy, DRL, Lupin, Cipla, IPCA etc. investing, on an average, around 2 % of their sales revenue on R&D as compared to 15–20 % in regard of the western firms. This large difference in the R&D intensity can be attributed to the different priorities and economic environments in which Indian and foreign firms had to operate. For instance, while Indian firms had been operating in a protected environment where replication of a drug/formulation was possible albeit through a different route, foreign firms had to operate in a far more competitive environment where inventions and innovations were at the premium. With the introduction of Product Patent regime, the top Indian companies realized that for unceasing operation of their business, they had to increase their R& D expenditure in order to survive in the new business environment. Since 1995, there was rapid rise in the R&D spending and it had increased to \$495.3 million by 2005–2006, although there is a very heavy concentration of R&D amongst top 20 firms. The R&D intensity has increased to around 5 % in 2011 from negligible percent in 1990 [2]. It is also evident from Fig. 1 that the R&D expenditure of Indian pharmaceutical firms has increased faster than the multinational subsidiaries operating in India. The lower growth of R&D expenditure in regard of the latter can be attributed to the fact these companies had already established a strong base for R&D, therefore, any dramatic rise in their R&D expenditure could not be expected. Further R&D activities of MNCs' subsidiaries are part of their global network which would

distribute the costs among various subsidiaries, which is not a case with Indian pharmaceutical firms. It may be further pointed out that with patent expiration of many blockbuster drugs, many companies are looking at India as a favorite destination for low cost alternatives especially due to the location of a large number of U.S. FDA (Food and Drug Administration) approved plants in India.

2 Review of Literature

Researchers have applied various parametric and non-parametric approaches to measure efficiency in the Indian Pharmaceutical Industry as well. For instance, Chaudhari and Das [6] applied the parametric frontier approach to estimate efficiency in IPI over the period 1990–2001. Their study shows that the mean efficiency scores of the industry have improved over the sub-period 1999–2001 against the sub period 1990–1998. Neogi and Ghosh [7] estimated the inter-temporal movement of the technical efficiency of the manufacturing firms. The study indicates that there has been a fall in the efficiency of the firms due to globalization. Saranga and Phani [8] analysed a sample of 44 Indian pharmaceutical companies and showed that the DEA models were sensitive to the selected inputs and outputs. Mazumdar and Rajiv [9] also used non-parametric approach of DEA to examine the competitiveness of Indian pharmaceutical industry by computing their technical efficiency for the period 1991–2005. Their analysis reveals a declining trend in output efficiency of firms. Pannu et al. [10] used DEA models to analyse the relative efficiency and productivity change in IPI between 1998 and 2007 which covers the post-TRIPS (1995) and post-Indian Patent Act Amendment (2005) periods. Internationally there are only few empirical studies relating to efficiency in this industry. Gonzalez and Gascon [11] analysed Spanish pharmaceutical industry using DEA BCC model and significant contribution of technical efficiency to productivity growth. Hashimoto and Haneda [12] also used DEA to analyse R&D efficiency at both firm and industry level of Japanese pharmaceutical companies. They empirically show that R&D efficiency of Japanese pharmaceutical industry has almost monotonically gotten worse throughout the study decade. Nauriyal and Sahoo [4] analysed the impact assessment of implementation of new IPR regime on Indian drug and pharmaceutical industry. Panel co-integration result showed that both the R&D and advertisement and marketing expenditures were strong determining forces of the sales revenue of the firms. The elasticity of sales revenue with respect to R&D expenditure was positive and significant.

The scenario for the pharmaceutical sector has changed significantly after Product Patent Act, 2005. In this context; it would be interesting to study efficiencies according to R&D and non-R&D firms especially after signing of TRIPS. Since the level of competition is going to be further increase due to the emergence of other cheaper producers and biotechnology equivalent, it would be interesting to study the efficiency and other parameters of this industry.

3 DEA Models

In case of DEA, a frontier with the input–output bundle of the best-performing firms in the sample is estimated; any shortfall of output that a firm produces or excess inputs used with respect to frontier is termed as inefficiency. Besides identifying the inefficient DMUs, a DEA assessment also helps to determine appropriate measures for improving the performance of the inefficient DMUs [13]. First formulated by Charnes, Cooper, and Rhodes (CCR) [14] in 1978, based on the constant returns to scale (CRS), and further updated by Banker, Charnes, and Cooper (BCC) [15] in 1984, based on the variable returns to scale (VRS), this technique has been extensively used to measure the relative efficiency in many sub-sectors.

$$\left. \begin{aligned}
 \text{Max } E_k &= \frac{\sum_{j=1}^s u_{jk}y_{jk}}{\sum_{i=1}^m v_{ik}x_{ik}} \\
 \text{s.t. } \frac{\sum_{j=1}^s u_{jk}y_{jr}}{\sum_{i=1}^m v_{ik}x_{ir}} &\leq 1, \quad r = 1, 2, \dots, n
 \end{aligned} \right\} \tag{1}$$

$$u_{jk}, v_{ik} \geq 0, \quad \forall \quad i = 1, 2, \dots, m \quad j = 1, 2, \dots, s$$

where y_{jk} is the amount of the j th output produced by the k th DMU; x_{ik} is the amount of the i th input used by the k th DMU; u_{jk} and v_{ik} are the weights given to the output and input; n is the no. of DMUs; s is the no. of outputs; m is the no. of inputs and ϵ is a very small constant value.

It involves finding the values for u_{jk} and v_{ik} , such that efficiency of the k th DMU is maximized, subject to the constraint that all efficiency measure must be less than or equal to one. This fractional problem which is difficult to solve is converted to a linear programming known as multiplier form. The equation form (2) is known as multiplier form. For every inefficient DMU, there are corresponding efficient units that act as benchmarks. The benchmarks can be obtained from the dual problem. Therefore, using the duality in linear programming, one can derive an equivalent envelopment form which is given in equation (3). The envelopment form involves fewer constraints than the multiplier form ($M + N < I + I$). In addition, envelopment model calculates the values of slack in variables and also minimizes them. Hence, it is generally the preferred form to solve.

Banker et al. [15] extended the CRS model into variable returns to scale (VRS) model. The VRS technical efficiency for firm k is estimated by imposing a convexity constraint i.e., $\sum_{i=1}^m v_{ik}x_{ik} = 1$. After that using the duality in linear programming, one can derive an equivalent envelopment form. Model (3) is known as CCR envelopment model with input orientation. The CRS linear programming problem can be easily modified to account for VRS by adding the convexity constraint. The basic difference between BCC and CCR model is that BCC model assumes variable returns to scale (VRS), while CCR model is based on constant returns to scale (CRS). The input-oriented CCR (Multiplier model) is given as follows:

$$\left. \begin{aligned}
 &Max E_k = \sum_{j=1}^s u_{jk}y_{jk} \\
 &s.t. \sum_{i=1}^m v_{ik}x_{ik} = 1 \\
 &\sum_{j=1}^s u_{jk}y_{jr} - \sum_{i=1}^m v_{ik}x_{ir} \leq 0, \quad r = 1, 2, \dots, n \\
 &u_{jk}, v_{ik} \geq \epsilon, \quad i = 1, 2, \dots, m \quad \& \quad j = 1, 2, \dots, s
 \end{aligned} \right\} \tag{2}$$

Input oriented CCR (envelopment) model [1]

$$\left. \begin{aligned}
 &Min Z_k = \phi_k - \epsilon \left(\sum_{j=1}^s s_{jk}^+ + \sum_{i=1}^m s_{ik}^- \right) \\
 &s.t. \sum_{r=1}^n \lambda_{rk}y_{jr} - s_{jk}^+ = y_{jk}, \quad j = 1, \dots, s, \\
 &\sum_{r=1}^n \lambda_{rk}x_{ir} + s_{ik}^- = \phi_k x_{ik}, \quad i = 1, \dots, m, \\
 &\lambda_{rk} \geq 0, \quad r = 1, 2, \dots, n, \\
 &s_{jk}^+, s_{ik}^- \geq 0; \quad \forall j, i, \\
 &\phi_k \text{ is unrestricted in sign}
 \end{aligned} \right\} \tag{3}$$

where ϕ_k is an efficiency ratio, s_{jk}^+ is slack in j th output of the k th DMU; s_{ik}^- is slack in the i th input of the k th DMU. The objective function is to minimize the efficiency score, ϕ_k , and to maximize input and output slacks of DMU. The model is interpreted as the maximum reduction in inputs of k DMU that can be done, given that k DMU has the same reference technology. The constraints (2) and (3) form the convex reference technology. $s_{ik}^-, s_{jk}^+ \geq 0$ shows that input and output slack to be non-negative. It is solved n times for solving the efficiency of all the DMUs.

DEA identifies the efficient decision making unit (DMU) among all DMUs. A DMU is considered Pareto efficient if it is not possible to reduce any input (in case of input-orientation) or increase any output (in case of output-orientation) without reducing output or increasing other input. Efficiency score ranges between zero and unity. The efficient DMUs have the efficiency of unity while inefficient DMUs have less than unity. The k th DMU is Pareto efficient if $\theta_k^* = 1$ and all slacks are zero, i.e., S_{jk}^{+*} and $S_{ik}^{-*} = 0$ for every j and i .

4 Data Sources and Variable Construct

4.1 Data Sources

The main data source of the study is Prowess compiled by Centre for Monitoring Indian Economy (CMIE). It compiles financial information of the companies out of their annual balance sheets. Our sample consists of data relating to financial statements of 141 firms of Indian pharmaceutical industry for which data for all eleven years (2000–2010) were available in the Prowess database. The main reason for choosing this sample is the fact that we have continuous availability of data for a common sample. We have considered the year 2000 as starting year because sufficient data for firms are available in Prowess after 1999. The firms considered in the study together account for about 80 % of the total sales revenue

Table 1 Descriptive statistics of inputs and output (2000–2001 to 2010–2011)

Statistics	NS	RM	S&W	A&M	CC
Mean	1,969.31	578.16	148.90	121.20	101.41
Std. dev.	3,996.29	1,062.11	307.82	324.68	215.85
Range	36,683.46	8,363.92	3,546.37	3,861.28	2,268.21
Minimum	0.13	0.05	0.16	0.05	0.08
Maximum	36,683.59	8,363.96	3,546.53	3,861.28	2,268.20

Note Figures in millions of rupees-deflated to 2,000 price

and 85 % of the input usage for the sector for almost all the years. Thus the sample of firms considered can adequately represent the industry.

4.2 Variable Constructs

The DEA models can generally take two forms viz., output-oriented and input-oriented. In an output-orientation model (output maximization) efforts are made to maximize the output with given inputs. On the other hand, in an input orientation model (input minimization) desired output is produced with minimum utilization of inputs. This model is preferred when output is given and inputs are flexible. The choice of the model depends on the available flexibility either with the inputs or outputs [16, 17]. All input variable considered in this study have required flexibility as compared to the output. Nevertheless, the output variable in our study is net sales revenue that may not have flexibility due to its dependence on external factors such as demand, exports etc., which are not in the control of management.

For computing efficiency, total net sales are taken as output variable, and raw material, salaries and wages, advertising and marketing and capital cost as input variables. All inputs and output are measured in millions of rupees. Descriptive statistics of the input–output variables are shown in Table 1.

4.2.1 Output Variable

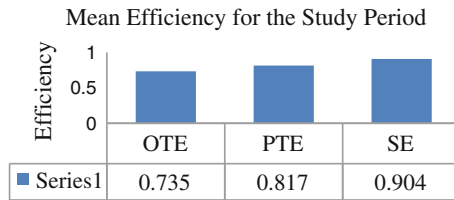
Net sales revenue: In case of Indian pharmaceutical industry, a fair number of studies [8, 10] have made use of sales revenue to examine the progress of IPI. Since the data on output in physical terms are not reported in the balance sheets, this variable, as an output variable, becomes a rational choice.

4.2.2 Input Variables

Raw material cost: It includes cost of all raw materials, spares and packaging.

Salary and wages: It includes total annual expenses incurred by a firm on all employees, including management. Payment of bonus, contribution to employee's provident fund and staff welfare expenses are also included under this variable.

Fig. 2 Mean efficiency of sample companies for the study period. *Source* Computed by author from CMIE, prowess data base



Cost of advertising and marketing: It includes the cost of advertising, marketing, distribution, travel and communication.

Capital cost: It includes rent, interest, depreciation, repairs and maintenance of plant and machinery. It is used as proxy variable for capital.

Since the PROWESS database provides data on current prices, therefore they had to be converted to constant prices (a 2000–2001 prices). The appropriate deflators were used to deflate the data at constant prices. The present study estimates efficiency (overall efficiency, Pure Technical efficiency and scale efficiency) of Indian pharmaceutical industry during 2000–2010, by applying two DEA models, namely, CCR (Charnes, Cooper, and Rhodes, 1978), and BCC (Banker, Charnes and Cooper, 1984).

5 Empirical Results and Discussions

It may be pointed out here that the efficiencies calculated are relative efficiencies and that the efficient firms (whose efficiency = 1) were used as the benchmark. Therefore these efficiencies represent relative-to-best efficiencies. Since CCR model works on the basis of CRS in which scale size of the DMU is not considered, it is relevant to assess pure technical efficiency. Therefore, in order to find out whether inefficiency in any firm is due to inefficient production operation or due to unfavorable conditions displayed by the size of firm, BCC model is also applied. Pure technical efficiency (PTE) refers to the proportion of technical efficiency which is attributed to the efficient conversion of inputs into output, given the scale size. Scale efficiency (SE) measures the impact of scale-size on the efficiency of a firm. It is measured as a ratio of CCR efficiency to BCC efficiency. If the value of SE score is one, then the firm is apparently operating at optimal scale. If the value is less than one, then the firm appears to be either small or big relative to its optimum scale-size. As shown in Fig. 2, the mean OTE of the IPI industry during 2000–2001 to 2010–2011 was 0.735, suggesting thereby that 26.5 % of the inputs can be reduced without changing the output at CRS technology. The CAGR of -0.636% suggests a declining trend in mean OTE for the industry during the study period which is a matter of concern. It can be observed from Fig. 2 that mean SE with a score of 0.904 of the industry is the highest

Table 2 Mean OTE, PTE and SE scores according to R&D-intensity

Year	R&D intensive firms			Non-R&D intensive firms		
	Mean OTE	Mean PTE	Mean SE	Mean OTE	Mean PTE	Mean SE
2000	0.828	0.894	0.923	0.744	0.799	0.933
2001	0.795	0.861	0.925	0.707	0.795	0.896
2002	0.767	0.830	0.928	0.711	0.798	0.892
2003	0.761	0.831	0.918	0.749	0.820	0.913
2004	0.681	0.814	0.842	0.672	0.785	0.861
2005	0.758	0.815	0.934	0.710	0.798	0.898
2006	0.741	0.813	0.916	0.721	0.800	0.904
2007	0.744	0.825	0.905	0.744	0.810	0.923
2008	0.735	0.812	0.908	0.694	0.765	0.914
2009	0.755	0.841	0.900	0.714	0.802	0.896
2010	0.766	0.850	0.906	0.702	0.819	0.861

Source Computed by authors from CMIE, Prowess data base

compared to PTE and OTE. The mean PTE score was around 0.78, indicating thereby that firms were 22 % pure technical inefficient which partly could be attributed to the infrastructural bottlenecks i.e., mainly frequent power cuts leading to delays and productivity losses.

5.1 Technical Efficiency According to R&D

The mean OTE, PTE and SE scores according to R&D expenditure are shown in Table 2. The sample of 141 firms was categorised as R&D and non-R&D firms. There are 68 firms having R&D expenditure and 73 firms as non-R&D.

It can be seen from Table 2 that the efficiency (OTE and PTE) of R&D firms is higher than the non-R&D firms for all the years. The results suggest that the gap between both groups has declined from 2000 to 2007 which implies that difference in efficiency has declined, and thereafter the gap increased. It indicates that R&D intensive firms are performing better in recent years. The mean OTE scores for the study period of both groups are 0.757 and 0.715 indicating that 24.3 and 28.5 % of inputs can be reduced without altering the output. The difference between the mean efficiency scores of both R&D and non-R&D firms is statistically significant which rejects the null hypothesis, as the calculated t-value is greater than the tabulated t-value of a two-tailed *t* test. For R&D firms, highest mean OTE and PTE scores was recorded in the year 2000. While in 2003 non-R&D intensive firms recorded the maximum efficiency score. In OTE score year 2004 was the worst year while in PTE score 2008 was the year when both the groups recorded the minimum value. Consequently inefficient firms can monitor their relative efficiency in the Indian pharmaceutical industry and take required action to improve the efficiency in this competitive regime by reducing input cost of raw material,

salary and wages, advertising and marketing cost and capital cost. Growth results from technological process that depends on competition among research oriented firms that generate innovations. Hence, the growth rate responds to profit incentives in the R&D sector. MNCs have comparatively higher R&D expenditure and use modern technology for production which makes them more efficient. They also allocate more resources for marketing activities which strengthen the firm's brand and product image which results in higher revenue and in turns enhances efficiency.

6 Conclusion

The Indian pharmaceutical industry is going through a stage of transition especially after the implementation of Product Patent Act in 2005. This paper analysed a sample of 141 Indian pharmaceutical firms over 10-years period starting from 2000. The study has endeavored to capture the impact of R&D on efficiency scores of firms. Firms are categorized as R&D intensive and non-R&D intensive firms and then their efficiency scores (both OTE and PTE) are compared. It was found that after 1995, there was significant rise in R&D intensity, consolidations, mergers and acquisitions among Indian companies though the R&D intensity was still far lower than the multinationals. Thus TRIPS agreement appeared to have increased R&D expenditures of the Indian pharmaceutical industry. It was found that the efficiency (OTE and PTE) of R&D intensive firms was higher than non-R&D intensive firms for all the years. The results also demonstrated that the gap between both groups has declined from 2000 to 2007 which indicates that difference in efficiency has declined. The gap, however, started increasing thereafter indicating that R&D firms were performing better in recent years. The mean OTE scores for the study period for both the groups stood at 0.757 and 0.715 indicating that 24.3 and 28.5 % of inputs could be reduced without any change in output. There are also many small firms which have experienced efficiency gains mainly by importing foreign technology and also by efficiently using inputs, through reduction in advertising and marketing expenditures and raw material cost. The findings hold important managerial and policy implications. In order to survive in the post Patent regime, firm's strategy should be on increasing R&D intensity and efficiency by reducing input cost.

References

1. CCI: A brief report pharmaceutical industry in India. Retrieved from www.cci.in/pdf/surveys_reports/indian-pharmaceuticals-industry.pdf (2012). Accessed 10 Feb 2013
2. Government of India: Annual Report, Department of Pharmaceuticals, Ministry of Chemicals and Fertilizers (2012)

3. Nauriyal, D.K.: TRIPS-compliant new patents act and Indian pharmaceutical sector: directions in strategy and R&D. *Indian J. Econ. Bus. Special issue China & India* (2006)
4. Nauriyal, D.K., Sahoo, D.: The new IPR regime and Indian drug and pharmaceutical industry: An empirical analysis. Paper Presented at 3rd Annual Conference of the EPIP Association, Bern, Switzerland—Gurten Park/October 3–4 2008 (2008)
5. Greene, W.: The emergence of India's pharmaceutical industry and implications for the U.S. Generic Drug Market. Office of Economics Working Paper. U.S. International Trade Commission (2007)
6. Chaudhari, K., Das, S.: WTO, the TRIPS and Indian pharmaceutical industry. *J. Quant. Econ.* **4**(1), 97–110 (2006)
7. Neogi, C., Ghosh, B.: Intertemporal efficiency variation in Indian manufacturing industries. *J. Prod. Anal.* **5**, 301–324 (1994)
8. Saranga, H., Phani, B.V.: Determinants of operational efficiencies in the Indian pharmaceutical industry. *Int. Trans. Oper. Res.* **16**, 109–130 (2009)
9. Mazumdar, M., Rajeev, M.: Comparing the efficiency and productivity of the Indian pharmaceutical firms: a Malmquist-Meta-Frontier approach. *Int. J. Bus. Econ.* **8**(2), 159–181 (2009)
10. Pannu, H.S., Kumar, U.D., Farooque, J.A.: Efficiency and productivity analysis of Indian pharmaceutical industry using data envelopment analysis. *Int. J. Oper. Res.* **10**(1), 121–136 (2011)
11. Gonzalez, E., Gascon, F.: Sources of productivity growth in the Spanish pharmaceutical industry (1994–2000). *Res. Policy* **33**, 735–745 (2004)
12. Hashimoto, A., Haneda, S.: Measuring the change in R&D efficiency of the Japanese pharmaceutical industry. *Res. Policy* **37**(10), 1829–1836 (2008)
13. Coelli, T., Rao, D.S.P., Battese, G.: *An introduction to Efficiency and Productivity Analysis*. Kluwer Academic Publishers, London (1998)
14. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **2**, 429–441 (1978)
15. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for the estimation of technical and scale inefficiencies in data envelopment analysis. *Manage. Sci.* **30**, 1078–1092 (1984)
16. Avkiran, N.K.: Investigating technical and scale efficiencies of Australian universities through data envelopment analysis. *Socio-Econ. Plann. Sci.* **35**(1), 57–80 (2001)
17. Ramanathan, R.: *An Introduction to Data Envelopment Analysis: A tool for Performance Measurement*. Sage Publishing, New Delhi (2003)

An Imperfect Production Model with Preservation Technology and Inflation Under the Fuzzy Environment

S. R. Singh and Shalini Jain

Abstract In this paper, we developed a two-warehouse imperfect production model under the fuzzy environment. Demand rate is time dependent. Production rate is demand dependent and greater than the demand rate. Inflation is also taken into this consideration. Shortages are allowed and partially backlogged. The model is developed in both crisp and fuzzy sense. All the costs are taken as fuzzy triangular numbers. Numerical examples for both the models and sensitivity analysis are illustrated this model.

Keywords Imperfect · Inflation · Shortages · Preservation technology · Fuzzy

1 Introduction

Soft Computing is a system solutions based on soft computing techniques. It provides rapid distribution of important results in soft computing technologies, a combination of research in evolutionary algorithms and genetic programming, neural science and neural net systems, fuzzy set theory and fuzzy systems. Soft Computing encourages the integration of soft computing techniques and tools into both everyday and advanced applications.

In practice, product quality is not always perfect but is dependent of the product design and the stability of the production process. When the production process is in an in-control state, the product can meet customer's quality specifications;

S. R. Singh (✉)

Department of Mathematics, D. N. College, Meerut, India
e-mail: shalini.shalini2706@gmail.com

S. Jain

Centre for Mathematical Sciences, Banasthali University, Vanasthali,
Rajasthan 304022, India
e-mail: shivrajpundir@gmail.com

otherwise, products are required to rework or scrap. Lee and Rosenblatt [1] considered the imperfect production process in a traditional EPQ model and derived the optimal number of inspection and production cycle time. The most of the paper considered the production of perfect items. Some of the researchers like Sana et al. [2], Chung et al. [3], Sana [4], Singh et al. [5, 6] extended different types of inventory model by considering the imperfect production.

In most of the models, the inflation and time value of money were disregarded. Because, it was considered by the researchers that the inflation and time value of money did not affect on the inventory but this plays an important role in the inventory model. Due to high inflation, the economic situation of the country may change rigorously and so, it is impossible to ignore the effect of inflation. In this direction, Buzzacott [7] and Misra [8] simultaneously developed EOQ model with constant demand and single inflation rate for all the related cost of the inventory models. Bierman and Thomas [9] extended the inventory model under inflation with some discount rates. Misra [10] presented the production model with different inflation rates for various associated costs. In this direction, some researches were done by Chandra and Bahner [11], Hariga [12], Sarker et al. [13, 14], Sarker and Moon [15].

The primary assumption in most of the inventory problems is that there is an owned warehouse with unlimited capacity. However, in reality, this assumption does not hold as a warehouse normally has limited storage capacity. On the other hand, when the goods of procuring goods is higher than the inventory related cost or the attractive price discount for bulk purchase is available or demand of items is very high, the inventory management procures a large quantity of items at a time. This large quantity of items cannot be stored in this existing warehouse called own warehouse (OW) with limited storage capacity. Then for storing the excess items, one or more warehouse called rented warehouse (RW) is rented either away or near to OW. Generally, the inventory cost of RW is greater than OW. During the decades, the two-warehouse inventory models under different assumptions have been extended by Singh et al. [16, 17], Singh and Singh [18], Dem and Singh [19].

In this paper, we study a two-warehouse imperfect production model with preservation technology investment under the fuzzy environment. Inflation is also considered. Shortages are allowed and partially backlogged. At the end, a numerical example for both the models is illustrating the proposed model, and concluding remarks are provided.

2 Assumptions and Notations

2.1 Assumptions

1. A two-warehouse production model is developed.
2. Perfect and imperfect items are taken into this consideration.
3. Shortages are allowed and partially backlogged.

4. Demand rate is time dependent.
5. Production rate is demand dependent and greater than the demand rate.
6. The model is taken in both fuzzy and crisp sense.
7. Holding cost for RW is greater than that of OW.
8. Preservation technology for decaying items is taken into consideration.

2.2 Notations

d	Demand rate i.e. $d = ae^{bt}$, where a, b, are positive constants and $a > b$
P	Production rate depend on the demand rate i.e. $P = kd$, $k > 1$
g	Original deterioration rate
ξ	Preservation technology (PT) cost for reducing deterioration rate in order to preserve the products, $\xi \geq 0$
$m(\xi)$	Reduced deterioration rate, a function of ξ defined by $m(\xi) = g(1 - e^{-l\xi})$, $l \geq 0$, where l is the simulation coefficient representing the percentage increase in $m(\xi)$
B	Backlogging rate, where $B = e^{-\delta t}$, $\delta > 0$
C_p	Production cost per unit per unit time
h_1	Holding cost for own warehouse (OW) per unit per unit time
h_2	Holding cost for rented warehouse (RW) per unit per unit time
C_0	Inspection cost for OW per unit per unit time
C_R	Inspection cost for RW per unit per unit time
d_1	Damaged item cost per unit per unit time
s	Shortage cost per unit per unit time
c	Lost sale cost per unit per unit time
W	Own warehouse capacity
r	Inflation rate
M	Rework cost per unit per unit time.

3 Model Formulation

In this model, during the time interval $[0, t_1]$, production starts and stock is kept in OW and inventory increases due to the combined effect of demand, production and deterioration. At $t = 0$, perfect and imperfect units are stored in OW until its capacity W is full. Remaining perfect and imperfect items are stored in rented warehouse at $t = t_1$, and production stops at $t = t_1$. Now, during the time interval $[t_2, t_3]$ and $[t_3, t_4]$, inventory decreases due to effect of demand and deterioration. While at time $[t_1, t_3]$ items are depletes due to deterioration occurs. During the time interval $[t_4, T]$, shortages occurs.

$$Q'_1(t) + [m(\xi) - g] Q_1(t) = P - d, \quad 0 \leq t \leq t_1 \tag{1}$$

$$Q'_2(t) + [m(\xi) - g] Q_2(t) = P - d, \quad t_1 \leq t \leq t_2 \tag{2}$$

$$Q'_3(t) + [m(\xi)g] Q_3(t) = -d, \quad t_2 \leq t \leq t_3 \tag{3}$$

$$Q'_4(t) + [m(\xi) - g] Q_4(t) = 0, \quad t_1 \leq t \leq t_3 \tag{4}$$

$$Q'_5(t) + [m(\xi) - g] Q_5(t) = -d, \quad t_3 \leq t \leq t_4 \tag{5}$$

$$Q'_6(t) = -Bd, \quad t_4 \leq t \leq T \tag{6}$$

with boundary conditions:

$$Q_1(0) = 0, Q_2(t_1) = 0, Q_3(t_3) = 0, Q_4(t_1) = W, Q_5(t_4) = 0, Q_6(t_4) = 0$$

Solutions of these equations are:

$$Q_1(t) = \frac{(k - 1)a}{[m(\xi) - g + b]} \left\{ e^{bt} - e^{-[m(\xi) - g]t} \right\} \tag{7}$$

$$Q_2(t) = \frac{(k - 1)a}{[m(\xi) - g + b]} \left\{ e^{bt} - e^{[m(\xi) - g](t_1 - t)} e^{bt_1} \right\} \tag{8}$$

$$Q_3(t) = \frac{a}{[m(\xi) - g + b]} \left\{ e^{[m(\xi) - g](t_3 - t)} e^{bt_3} - e^{bt} \right\} \tag{9}$$

$$Q_4(t) = W \left\{ e^{[m(\xi) - k](t_1 - t)} \right\} \tag{10}$$

$$Q_5(t) = \frac{a}{[m(\xi) - k + b]} \left\{ e^{[m(\xi) - g](t_4 - t)} e^{bt_4} - e^{bt} \right\} \tag{11}$$

$$Q_6(t) = \frac{a}{(b - \delta)} \left\{ e^{(b - \delta)t_4} - e^{(b - \delta)t} \right\} \tag{12}$$

The present worth of Production cost is given by

$$PC = \int_0^{t_2} P e^{-rt} dt \tag{13}$$

The present worth of holding cost in OW is

$$HC_1 = \left[\int_0^{t_1} Q_1(t) e^{-rt} dt + \int_{t_1}^{t_3} Q_4(t) e^{-rt} dt + \int_{t_3}^{t_4} Q_5(t) e^{-rt} dt \right] \tag{14}$$

The present worth of holding cost for RW is

$$HC_2 = \left[\int_{t_1}^{t_2} Q_2(t)e^{-rt} dt + \int_{t_2}^{t_3} Q_3(t)e^{-rt} dt \right] \tag{15}$$

The present worth of damaged cost is

$$DC = \left[\int_0^{t_2} Pe^{-rt} dt - \int_{t_2}^{t_4} de^{-rt} dt \right] \tag{16}$$

The present worth of inspection cost for OW is

$$I_0 = \int_0^{t_1} Pe^{-rt} dt \tag{17}$$

The present worth of inspection cost for RW is

$$I_R = \int_{t_1}^{t_2} Pe^{-rt} dt \tag{18}$$

The present worth of shortage cost is

$$SC = \int_{t_4}^T -Q_6(t)e^{-rt} dt \tag{19}$$

The present worth of lost sale cost is

$$LC = \left[\int_{t_4}^T \{1 - B\}de^{-rt} dt \right] \tag{20}$$

The present worth of rework cost is

$$RC = \left[\int_0^{t_1} Pe^{-rt} dt + \int_{t_1}^{t_2} Pe^{-rt} dt \right] \tag{21}$$

The present worth of total cost is

$$TC = [C_1.PC + h_1.HC_1 + h_2.HC_2 + d_1.DC + C_0.I_0 + C_R.I_R + s.SC + c.LC + M.RC + \xi] \tag{22}$$

3.1 Fuzzy Model

In order to develop the model in a fuzzy environment, we consider the production cost, holding cost for OW, holding cost for RW, damaged cost, Inspection cost for OW and RW, shortage cost, lost sale cost, rework cost as the triangular fuzzy numbers $\tilde{C}_1 = (C_1 - \Delta_1, C_1, C_1 + \Delta_2)$, $\tilde{h}_1 = (h_1 - \Delta_3, h_1, h_1 + \Delta_4)$, $\tilde{h}_2 = (h_2 - \Delta_5, h_2, h_2 + \Delta_6)$, $\tilde{d}_1 = (d_1 - \Delta_7, d_1, d_1 + \Delta_8)$, $\tilde{C}_0 = (C_0 - \Delta_9, C_0, C_0 + \Delta_{10})$, $\tilde{C}_R = (C_R - \Delta_{11}, C_R, C_R + \Delta_{12})$, $\tilde{s} = (s - \Delta_{13}, s, s + \Delta_{14})$, $\tilde{c} = (c - \Delta_{15}, c, c + \Delta_{16})$ and $\tilde{M} = (M - \Delta_{17}, M, M + \Delta_{18})$ such that $0 < \Delta_1 < C_1$, $\Delta_2 > 0$, $0 < \Delta_3 < h_1$, $\Delta_4 > 0$, $0 < \Delta_5 < h_2$, $\Delta_6 > 0$, $0 < \Delta_7 < d_1$, $\Delta_8 > 0$, $0 < \Delta_9 < C_0$, $\Delta_{10} > 0$, $0 < \Delta_{11} < C_R$, $\Delta_{12} > 0$, $0 < \Delta_{13} < s$, $\Delta_{14} > 0$, $0 < \Delta_{15} < c$, $\Delta_{16} > 0$, $0 < \Delta_{17} < M$, $\Delta_{18} > 0$ where $\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5, \Delta_6, \Delta_7, \Delta_8, \Delta_9, \Delta_{10}, \Delta_{11}, \Delta_{12}, \Delta_{13}, \Delta_{14}, \Delta_{15}, \Delta_{16}, \Delta_{17}, \Delta_{18}$ are determined by the decision maker based on the uncertainty of the problem. Thus, production cost (C_1), holding cost for OW (h_1), holding cost for RW (h_2), damaged cost (d_1), Inspection cost for OW (C_0) and RW (C_R), shortage cost (s), lost sale cost (c), rework cost (M) are considered as the fuzzy numbers $\tilde{C}_1, \tilde{h}_1, \tilde{h}_2, \tilde{d}_1, \tilde{C}_0, \tilde{C}_R, \tilde{s}, \tilde{c}$ and \tilde{M} with membership functions.

$$TC = [\tilde{C}_1.PC + \tilde{h}_1.HC_1 + \tilde{h}_2.HC_2 + \tilde{d}_1.DC + \tilde{C}_0.I_0 + \tilde{C}_R.I_R + \tilde{s}.SC + \tilde{c}.LC + \tilde{M}.RC + \xi] \tag{23}$$

$$\begin{aligned} \tilde{C}_1 &= (C_1 - \Delta_1, C_1, C_1 + \Delta_2) \\ \tilde{h}_1 &= (h_1 - \Delta_3, h_1, h_1 + \Delta_4) \\ \tilde{h}_2 &= (h_2 - \Delta_5, h_2, h_2 + \Delta_6) \\ \tilde{d}_1 &= (d_1 - \Delta_7, d_1, d_1 + \Delta_8) \\ \tilde{C}_0 &= (C_0 - \Delta_9, C_0, C_0 + \Delta_{10}) \\ \tilde{C}_R &= (C_R - \Delta_{11}, C_R, C_R + \Delta_{12}) \\ \tilde{s} &= (s - \Delta_{13}, s, s + \Delta_{14}) \\ \tilde{c} &= (c - \Delta_{15}, c, c + \Delta_{16}) \\ \tilde{M} &= (M - \Delta_{17}, M, M + \Delta_{18}) \end{aligned} \tag{24}$$

by Centroid Method, we get

$$\begin{aligned}
 \tilde{C}_1 &= C_1 + \frac{1}{3}(\Delta_2 - \Delta_1) \\
 \tilde{h}_1 &= h_1 + \frac{1}{3}(\Delta_4 - \Delta_3) \\
 \tilde{h}_2 &= h_2 + \frac{1}{3}(\Delta_6 - \Delta_5) \\
 \tilde{d}_1 &= d_1 + \frac{1}{3}(\Delta_8 - \Delta_7) \\
 \tilde{C}_0 &= C_0 + \frac{1}{3}(\Delta_{10} - \Delta_9) \\
 \tilde{C}_R &= C_R + \frac{1}{3}(\Delta_{12} - \Delta_{11}) \\
 \tilde{s} &= s + \frac{1}{3}(\Delta_{14} - \Delta_{13}) \\
 \tilde{c} &= c + \frac{1}{3}(\Delta_{16} - \Delta_{15}) \\
 \tilde{M} &= M + \frac{1}{3}(\Delta_{18} - \Delta_{17})
 \end{aligned}
 \tag{25}$$

$$y_1 = \left[\begin{aligned}
 &(C_1 + \Delta_2).PC + (h_1 + \Delta_4).HC_1 + (h_2 + \Delta_6).HC_2 \\
 &+ (d_1 + \Delta_8).DC + (C_0 + \Delta_{10}).I_0 + (C_R + \Delta_{12}).I_R \\
 &+ (s + \Delta_{14}).SC + (c + \Delta_{16}).LC + (M + \Delta_{18}).RC + \xi
 \end{aligned} \right]
 \tag{26}$$

$$y_2 = [C_1.PC + h_1.HC_1 + h_2.HC_2 + d_1.DC + C_0.I_0 + C_R.I_R + s.SC + c.LC + M.RC + \xi]
 \tag{27}$$

$$y_3 = \left[\begin{aligned}
 &(C_1 - \Delta_1).PC + (h_1 - \Delta_3).HC_1 + (h_2 - \Delta_5).HC_2 \\
 &+ (d_1 - \Delta_7).DC + (C_0 - \Delta_9).I_0 + (C_R - \Delta_{11}).I_R \\
 &+ (s - \Delta_{13}).SC + (c - \Delta_{15}).LC + (M - \Delta_{17}).RC + \xi
 \end{aligned} \right]
 \tag{28}$$

By the method of Defuzzification,

$$TC = \frac{y_1 + 2y_2 + y_3}{4}
 \tag{29}$$

Then we find out the total cost in fuzzy sense.

4 Numerical Example (Crisp Model)

By the previous data, we consider the following data in the appropriate units: (Fig. 1, Table 1)

Fig. 1 Convexity of the Total cost w.r.t. t_2^* and t_3^*

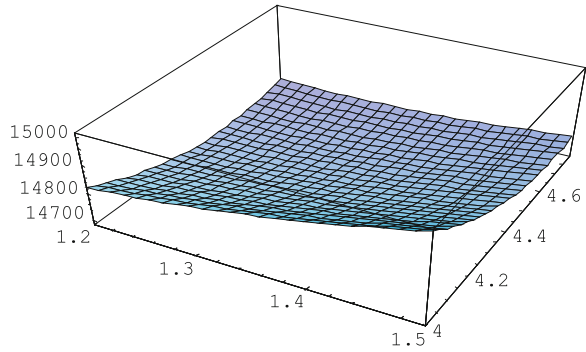


Table 1 Sensitivity Analysis of the model

Parameter	Change in parameter	t_2^*	t_3^*	t_4^*	TC*
r	0.012	1.19088	4.46169	8.7054	14519.4
	0.014	1.16774	4.39003	8.63624	14355.2
	0.016	1.12149	4.31886	8.566	14195.6
	0.018	1.14461	4.2482	8.49649	14040.7
	0.02	1.09841	4.17806	8.42594	13890.6
C_p	6.5	1.1726	4.52162	8.79358	14810.6
	7	1.13124	4.50948	8.81307	14929.5
	7.5	1.08995	4.49739	8.83254	15045.1
	8	1.04873	4.48536	8.85196	15157.4
	8.5	1.00758	4.47338	8.87136	15266.4
a	82	1.21458	4.53581	8.77086	15030.2
	84	1.21512	4.53771	8.76782	15372
	86	1.21563	4.53952	8.76493	15713.8
	88	1.21611	4.54125	8.76217	16055.7
	90	1.21658	4.5429	8.75953	16397.5
b	3.1	1.19435	4.45746	8.69026	14801.8
	3.2	1.17504	4.38396	8.60852	14923.3
	3.3	1.15611	4.31317	8.52876	15052.3
	3.4	1.13755	4.24492	8.45096	15188.6
	3.5	1.11937	4.1791	8.37505	15331.7
C_0	0.9	1.21402	4.53382	8.77404	14688.4
	1	1.21402	4.53382	8.77404	14709.7
	1.1	1.21402	4.53382	8.77404	14731.1
	1.2	1.21402	4.53382	8.77404	14752.4
	1.3	1.21402	4.53382	8.77404	14773.8
c	3.7	1.19554	4.46849	8.69963	14777
	3.8	1.17745	4.40525	8.62691	14892.6
	3.9	1.15975	4.34398	8.55584	15013.5
	4	1.14243	4.28461	8.48636	15139.5
	4.1	1.12547	4.22703	8.41843	15270.4

$$r = 0.01, C_P = 6, a = 80, b = 3, C_R = 1.4, C_o = 0.8, c = 3.6, s = 7, \delta = 0.04, \\ d_1 = 4, W = 40, g = 0.02, l = 0.01, k = 2, h_1 = 5, \zeta = 400, t_1 = 1, T = 10$$

The output results of the decision parameters are:

$$t_2^* = 1.21402, t_3^* = 4.53382, t_4^* = 8.77404, TC^* = 14688.4$$

(Fuzzy Model)

$$r = 0.01, C_1 = (3, 6, 9), a = 80, b = 3, C_R = (1.2, 1.4, 1.6), C_o = (0.6, 0.8, 1.0), \\ c = (3.4, 3.6, 3.8), s = (5, 7, 9), \delta = 0.04, d_1 = (2, 4, 6), W = 40, g = 0.02, \\ l = 0.01, k = 2, h_1 = (2, 5, 8), \zeta = 400, t_1 = 1, T = 10$$

The output results of the decision parameters are:

$$t_2^* = (0.636625, 1.21402, 1.43569), t_3^* = (2.43009, 4.53382, 5.65738), \\ t_4^* = (7.54222, 8.77404, 9.38272), TC^* = (9834.14, 14688.4, 19759.9)$$

4.1 Observations

1. As we increase the inflation rate (r), then t_2^*, t_3^*, t_4^* and Total cost decreases.
2. On increasing the production cost (C_P), t_2^* and t_3^* decreases while t_4^* and total cost increases.
3. On increasing the demand rate (a), t_2^*, t_3^* and total cost increases while t_4^* decreases.
4. On increasing the demand rate (b), t_2^*, t_3^*, t_4^* decreases whereas total cost increases.
5. On increasing the inspection cost for OW(C_o), t_2^*, t_3^*, t_4^* remains same while total cost increases.
6. On increasing the lost sale cost parameter (c), t_2^*, t_3^*, t_4^* decreases whereas total cost increases.

5 Conclusion

We developed the imperfect model with inflation and shortages under fuzzy environment. A solution procedure is presented to determine an optimal replenishment cycle, production cost, inspection cost, damaged item cost and preservation technology cost such that the total cost per unit time is minimized. A

numerical example and sensitivity analysis are presented to illustrate the model. This model can be extended in so many ways such as partial backlogging, learning, permissible delay etc.

References

1. Lee, H.L., Rosenblatt, M.J.: Simultaneous determination of production cycle and inspection schedules in a production system. *Manage. Sci.* **33**, 1125–1136 (1987)
2. Sana, S.S., Goyal, S.K., Chaudhuri, K.S.: An imperfect production process in a volume flexible inventory model. *Int. J. Prod. Econ.* **105**, 548–559 (2007)
3. Chung, K.J., Her, C.C., Lin, S.D.: A two-warehouse inventory model with imperfect quality production processes. *Comput. Ind. Eng.* **56**, 193–197 (2009)
4. Sana, S.S.: A production inventory model in an imperfect production process. *Euro. J. Oper. Res.* **200**, 451–464 (2010)
5. Singh, S.R., Jain, S., Pareek, S.: A Warehouse imperfect fuzzified production model with shortages under inflationary conditions. *advances in decision sciences*. Hindawi Publishing Corporation **638060**, 16 (2012)
6. Singh, S.R., Jain, S., Pareek, S.: An imperfect quality items with learning and inflation under two limited storage capacity. *Int. J. Ind. Eng. Comput.* **4**(4) (2013)
7. Buzzacott, J.A.: Economic order quantities and inflation. *Oper. Res. Q.* **26**, 553–558 (1975)
8. Misra, R.B.: A study of inflationary effects on inventory systems. *Logistics Spectr.* **9**, 260–268 (1975)
9. Bierman, H., Thomas, J.: Inventory decisions under inflationary condition. *Decis. Sci.* **8**, 151–155 (1977)
10. Misra, R.B.: A note on optimal inventory management under inflation. *Naval Res. Logistics Q.* **26**, 161–165 (1979)
11. Chandra, J., Bahner, M.L.: The effects of inflation and the time-value of money on some inventory systems. *Int. J. Prod. Res.* **2314**, 723–730 (1985)
12. Hariga, M.: An EOQ model for deteriorating items with shortages and time-varying demand. *J. Oper. Res. Soc.* **46**, 398–404 (1995)
13. Sarker, B.R., Jamal, A.M.M., Wang, S.: Supply chain models for perishable products under inflation and permissible delay in payment. *Comput. Oper. Res.* **27**, 59–75 (2000)
14. Sarkar, B., Sana, S.S., Chaudhuri, K.S.: A finite replenishment model with increasing demand under inflation. *Int. J. Mat. Oper. Res* **2**, 347–385 (2010)
15. Sarkar, B., Moon, I.K.: An EPQ model with inflation in an imperfect production system. *Appl. Math. Comput.* **217**, 6159–6167 (2011)
16. Singh, S.R., Kumar, N., Kumari, R.: Two warehouse inventory model for deteriorating items with shortages under the conditions of permissible delay in payments. *Int. Trans. Math. Sci. Comput.* **1**, 123–134 (2008)
17. Singh, S.R., Kumar, N., Kumari, R.: Two-warehouse inventory model for deteriorating items with shortages under inflation and time-value of money. *Int. J. Comput. Appl. Math.* **4**(1), 83–94 (2009)
18. Singh, S.R., Singh, T.J.: Two warehouse perishable inventory model with exponential demand and partial backlogging. *Int. Trans. Appl. Sci.* **1**, 33–48 (2008)
19. Dem, H., Singh, S.R.: A two warehouse production model with quality consideration. *International conference on modeling, optimisation and computing (ICMOC-2012)*. *Procedia Eng.* **38**, 3242–3259 (2012)

Assessing the Soft Computing Approach for Problem Solving in the Multi Criteria Decision Making for Performance Assessment in Indian Institutions Through Case Study Approach

Pooja Tripathi, Jayanthi Ranjan and Tarun Pandeya

Abstract In the highly competitive global market the growth of Indian economy substantially depends on its knowledge resources. A competency-based approach towards human resources management (HRM) is one of the key success factors in modern organizations. Today, academia faces much the same situation which manufacturing and service companies “in the real world” have been encountering for decades. It seems that institutions only with explicit quality assurance and the ability to produce talents needed by the industry will be able to survive in the academic market. The need for interaction with the competence data for the continuous monitoring and identification of individual competencies create the opportunity for application of soft computing tool in general and that of Expert Systems in particular. This paper covers the investigation of the outcome of the realization of the need for the soft computing tool for the competence management and performance assessment in the educational institutions to improve quality in the teaching and learning process through the case study methodology on various parameters.

Keywords Soft computing · Expert system · Competence management · Human resource management

P. Tripathi (✉)
Inderprastha Engineering College, Ghaziabad, India
e-mail: trippooja@gmail.com

J. Ranjan
IMT, RajNagar, Ghaziabad, India

T. Pandeya
IGNOU, New Delhi, India

1 Introduction

In the highly competitive global market the growth of Indian economy substantially depends on its knowledge resources. The demand for skilled personnel is increasing, due to more and more multinational companies entering India for their operations. The gap that exists between supply and demand of graduates has taken as a profitable business opportunity, and participation by the private sector in education sector has been very noticeable. The mushrooming of the institutions resulted into imparting a poor quality of technical education which also effected in reducing the competencies of the intellectual capital. There are various issues concerning the setting up and running of technical institutions with the key aspect of quality assurance. in reputed technical institutions. Today, academia faces much the same situation which manufacturing and service companies “in the real world” have been encountering for decades. It seems that institutions only with explicit quality assurance and the ability to produce talents needed by the industry will be able to survive in the academic market. The faculty members impart the content of the curriculum to the students and also enhance their competencies to become employable. The knowledge is created and disseminated through the intellectual creativity and research efforts of the faculty members. A competency-based approach towards human resources management (HRM) is one of the key success factors in modern organizations. A manager should always bear in mind that the definition of competency is fundamental in all HRM activities—starting with recruitment, through training, assessment, and development.

This paper covers the investigation of the outcome of the realization of the need for the soft computing tool for the competence management and performance assessment in the educational institutions to improve quality in the teaching and learning process. A developed web based expert system for Competence Management with special reference to faculty member’s competencies, their job descriptions and the behaviors required in educational institutions was assessed on different parameters to analyze its effectiveness for the purpose through a questionnaire methodology. The developed system helps in organizing the knowledge generated by institutions for the substantial growth of quality in the teaching and learning process in the country.

2 Literature Review

Competencies include the sum of success factors necessary for achieving important results in a specific job in a particular organization (Garrett 2003). These attributes or factors include: personal characteristics, traits, motives, values or ways of thinking that impact an individual’s behaviour. Analysis and assessment of competency begin with job competency mapping or individual competence mapping and continue with competency modelling. Competence, competency

mapping and modelling generally serve to structure a personal competence/y or to construct a job competency.

A competency-based approach towards human resources management (HRM) is one of the key success factors in modern organizations. A manager should always bear in mind that the definition of competency is fundamental in all HRM activities—starting with recruitment, through training, assessment, and development. Competencies are more than just characteristic of a person; therefore, competencies can be defined as knowledge, skills and attitudes, commitments or values that are necessary for a specific position. The competence management data in an organization require continuous analysis, interpretation and preparation of databases at various levels of use and applications in different decision-making systems. The need for interaction with the data for the continuous monitoring and identification of individual competencies create the opportunity for application of Artificial Intelligence in general and that of Expert Systems in particular. Internet/ Intranet technology can change the way that an Expert System is developed and distributed. For the first time, knowledge on any subject can directly be delivered to users through a web based Expert System. Few researchers provided some examples of web-based expert systems in Industry, Medicine, Science and Government, and claimed that “there are now a large number of expert systems available on the Internet.” They argued that there are several factors that make the Internet, by contrast to standalone platforms, an ideal base for KBS (Knowledge Based System) delivery. These factors include: The Internet is readily accessible, web-browsers provide a common multimedia interface, several Internet-compatible tools for KBS development are available, Internet-based applications are inherently portable, and emerging protocols support cooperation among KBS. “Expert System on Competence Management” is an integrated system that addresses most of the important aspects of competence management in organization. The main goal of the developed system is to provide users with recommendations and advice concerning their competence management with the organization goals and visions [1]. The developed web-based expert system for competence management for the educational institution has been based on deductive inference and its theory base is the order of first logic. The knowledge base of this system contains production rules derived from a decision tree. The focus of the project—of a conventional system—is the “data” while the focus project of an expert system (ES) is the domain “knowledge” gathered through the intranet application in the institution for analyzing the competence sets of the individuals and trying to develop the association between various jobs and competencies required for performing the task. The developed web based system collects the required knowledge and provides the framework to develop business strategy to manage the competence and performance assessment of the organization. The authors developed the expert system and collected the results of the implemented expert system in real life situation. This paper is the outcome of the investigation of the effect of the developed expert system and its utilization over various parameters.

3 Research Methodology

An attempt has been made to study the implementation of the tool in a real-life situation. The authors tried to apply the developed tool in a Sample Business School (SBS) in India. The authors have thoroughly studied the principles and policies of SBS, India, and (a pseudo name given to mask the identity of the original institute). This section deals with the Sample Business School operations principally the mission and vision. It highlights the main issues and concern of the SBS related to the dynamic market trend. It specially highlights the shortcomings of the present system: for example the time spent in competence identification, competence assessment, competence acquisition, competence usage through human expert. Thus, the competence management process require vast amount of information handling and if adequate level of knowledge is not applied then task of placing the right people to right task cannot be accomplished. Using the various parameters for the faculty competence assessment, it is seen that several improvements could be made. These parameters help the academic institute to locate, hire, develop and retain the faculty to run the institution. The competence model developed helps an educational institution by increasing the acceptance of its students in the enterprise world wide and the also enhances the research and developments.

A pilot project in an 80-person application development was completed in November 2010, and full implementation is proceeding. The project, called Competence Development and Performance Assessment Expert System (CDPAES) is focused not on entry level competencies, but rather on those needed and acquired to stay on the leading edge of the workplace. The CDPAES initiative is being managed for the “Learning and Developing” within SBS. The main goal is to use the competency model to transfer and build knowledge, not merely to test it but to provide employees with a better idea of what competencies are required of them. Hence, they will be better consumers of educational offerings within and outside the SBS. The project is also expected to lead to better matching of employees to jobs and work teams. Eventually the project may be extended throughout SBS and into other educational institution. There are five major components to the CDPAES project:

- Development of a structure of competency types and levels;
- Defining the competencies required for particular jobs;
- Rating the performance of individual employees in particular jobs based on the competencies;
- Implementing the knowledge competencies in an online system;
- Linkage of the competency model to learning offerings.

4 Developing Competency and Performance Assessment Structure for the Expert System

Before the project began SBS had already defined certain competencies, but they were largely restricted to entry-level skills. For developing the competency structure, the data on the warehousing operations was collected through the author's reflections combined with the inputs from informal discussions with several educators, consultants, and managers, which had taken place at several meetings, seminars and conferences over the last two years, and also combined with a review of literature in the field of education. The authors tried to find out how an educational institution perceives them in term of competitiveness to attract and retain talented faculty. Competency Based Management process describes importance of the working conditions provided, job security feature. The factors like the competitive benefits and salaries for the performers. The various opportunities provided for the educational trainings and advancement growth. It also describes the importance of the manager and employee relationship. One of the other key parameters is how well the management is trying to provide a balance between the work and life. The collected information helped to define an extensive set of essential skills, goals, and objectives for successful faculty in various domains mainly administration, research work, role in higher education, other roles and responsibilities assigned such as evaluation of teaching through students feedback, organizing capabilities, books published and reviewed, written communication, professional academic skills. From the case experiences, standard parameters were developed for the assessment of the faculties in the academic institution. There are various role specific competencies required for a faculty member who functions effectively as a faculty/administrator, faculty/educator, faculty/leader, faculty/researcher, faculty/consultant, faculty/reviewer, faculty/member, faculty/evaluator and faculty/organizer in terms of personality, ability, knowledge and skills. A standard measuring tool was developed to assess the particular competency in the candidate as per the job requirement. The brain storming and the interviews of the senior management and the decision maker for the faculty performance assessment resulted in a tool to indicate the importance of a particular behaviour to assess the presence of particular competency in individual to perform a particular task in the teaching and learning process.

A tool was developed to generate the importance of particular behavioural presence for the competency assessment to perform a particular task. The combination of the data gathered from the assessment and the importance of the competency was used to generate the competency model for the various jobs in the teaching and learning process. The Competency Model for each job in the educational institution was developed which clearly defines the definition of the competency and clearly projects the behavioural traits in the individual for showing the presence the competency and the various skill levels such as 1, 2, 3, 4, 5 as defined in Table 3 to indicate the competency levels as beginner, working, good, very good and outstanding.

5 Evaluation of the Employee on Job Competencies

The evaluation procedures are based on the beliefs that all faculty members will exhibit satisfactory levels of behaviour and that the annual evaluation will be a means of communication through which a faculty member and the department chair develop a clear understanding of goals as well as both strengths and weaknesses. In keeping with these beliefs, the evaluation process will include an assessment of the individual faculty member's performance in each of the various dimensions (such as administrator, leader, educator, organizer, reviewer, member, researcher, consultant) at one of six levels of performance: OUTSTANDING, ABOVE SATISFACTORY, SATISFACTORY, CONDITIONAL, UNSATISFACTORY AND NOT APPLICABLE. The department chair will assume the leadership role in the evaluation of faculty. If a faculty member performs below the satisfactory level in any of the dimensions applicable to his/her role, the faculty member will be rated as conditional or unsatisfactory by the department head in that dimension only. Each of the dimensions will be weighted and overall rating will be determined by the department head. In each role, evaluative indicators are given for three major dimensions of performance: outstanding, above satisfactory, and satisfactory.

5.1 Proposed Online Competence Development and Performance Assessment Expert System

The CDPAES project involved building an online system that contained the competency structure, the job rating system and ratings database, and the competency levels for employees. The authors had built a prototype of the system for the pilot using Microsoft Access; though it had worked well, the system needed greater performance and robustness, and was being ported to SQL Server. The system would have a Web front end for easy access around the world through SBS's Intranet. The purpose of the proposed system is for assisting individual faculty members and administrators in career development, program planning, and evaluation process. Individuals can use these guides to identify their career paths that require competencies they want to develop or already possess. Individuals can plan professional development around competencies that prepare them for academic roles to which they aspire. Similarly, head of the departments can use these checklists and tables when providing career guidance to junior faculty and subsequently as an evaluation tool linked with professional development. Job descriptions could also be prepared using the competencies listings. Institutional faculty development programs could be organized around the competencies recognized as needed by groups of faculty members to more efficiently use faculty development funds.

6 Implementing the Competency Model

The pilot for the project CDPAES had gone well, and now implementation was proceeding with all people and their jobs in the SBS. One issue to be determined was how to migrate the competency model to integrate into the framework. Since framework was also marketed externally, embedding the competency model within it might also create a demand for the competency model in other educational institutions. The model would become a vehicle for institutionalizing innovation in the ever changing knowledge society. The head of the institutions/department determined that employees at SBS needed to master a new form of knowledge then they could force development of the competency by insisting upon its presence in all job competency requirements.

7 Survey

A mail survey was conducted on the issues related to gain the managerial perspectives on the issues relating to the impacts of the developed expert system. The survey was aimed at individuals who were responsible for using the expert system at different levels and for different tasks. The survey questionnaire was sent to the 60 people in the organization.

7.1 Respondents

Overall response rate was approximately 80 %. The response rate was good and it is due to the three factors: (1) The personal meet was made with the person. (2) The respondents interest in the new system introduced in the organization. (3) The questionnaire was very simple and easy to read and the collecting process was very convenient. The survey findings discussed here are based on the responses from (N = 45) those people who rated that the developed system was “good” or “very good” business investment. The respondents were generally those who were using the system for more than six months and were well acquainted with the scope of the developed system. The respondents were selected such that 30 % were manager utilities in the developed system and remaining were using the employee utilities.

8 Results and Discussion

The managers in the survey assessed the importance of the specific implementation factors on a discrete point scales from 1 (“low” importance) to 5 (“high” importance). A rating of “high” importance was interpreted to mean that the

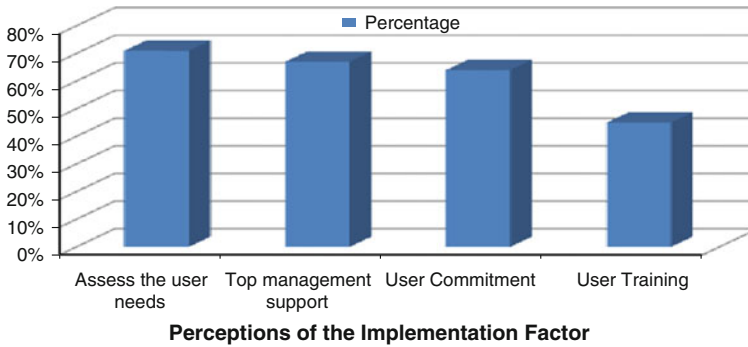


Fig. 1 Perceptions of the implementation factor

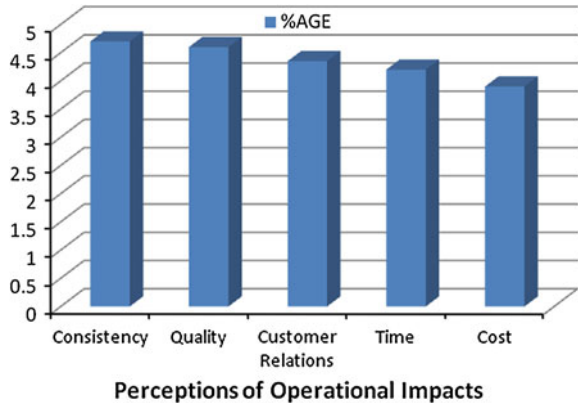
respondent considered the factor to be particularly vital to the developed system. A rating of “low” importance was interpreted to mean that the respondent considered the factor to be relatively less critical success factor. As illustrated in the Fig. 1, the study generally confirmed the importance of the factors identified earlier. The survey results are discussed below along with the implications for the managerial practice.

The assessment of the user needs as shown in Fig. 1 above is rated highest among the all factors, as 71 % of the respondents rated it to be of high importance. This finding indicates that needs assessment is the task of high priority that demands the support of the end user management. The other factors that also play role in the implementation of the developed expert system is the top management support i.e. the receptivity towards the adoption of the innovation in the organization; user commitment towards the usage of the system also plays important role. The user training for the proper access to the system can also affect the type of the data to be stored. It was surprising to get the low rating for the user rating as the developed system was easy to use, making training less important issue.

8.1 Operational Impacts

The respondents were asked to rate the six aspects of their operations that had been impacted by the developed system using a scale of 1 (“negative” impact) to 5 (“positive” impact), with a score of 3 corresponding to “no impact”. Figure 2, summarizes the findings, on a aggregate level the developed system was found to have the favorable impact on all of the listed aspects of operations. According to the respondents, the developed system has the positive impact with respect to: consistency of work output, quality of work output, and relationship with customers, each of these top three ranked impacts of developed system reflect the value added aspects of operations. The productivity related impacts of developed system—time and cost—were ranked lower.

Fig. 2 Perceptions of expert system operational impacts



Although the aggregate findings revealed that the developed system perceived to have the positive impact with respect to all operational impact items. A review of the individual survey responses shows that few respondents have shown the negative impact due to the developed system. Aspects of operations for which respondents reported the negative impacts included (a score of “1” or “2”) which includes the time needed to perform task, and the cost of performing the task. A review of individual questionnaire suggests that for these cases, any negative productivity related impacts were offsets by the strong value added gains of improved quality and consistency.

8.2 Social Impact

The use of developed system on user job content and satisfaction was assessed by the respondents. Although this assessment was subjective, we believe that the respondents provided an accurate appraisal of user impacts as most of them had gained a users perspective through their personal use.

The survey offers support for the notion that Expert System technology can enhance the intellectual content of the Expert System users’ job. A large majority (78 %) of the user believed that “Expert System (ES)” has allowed the user to concentrate on the more intellectually challenging aspects of the job, ES frees the user to perform the daily routine tasks. Nine percent of the sample group reported no impact, while 13 % were “not sure” of the impact. None of the respondents felt that the ES had reduced the intellectual challenge of the user’s job. One explanation for the positive impact of ES technology on job content is that most of the ES represented in this study is explicitly developed to support user as a tool needed to access to expert knowledge (Fig. 3).

Fig. 3 Perception of expert system impact on user job content (with N = 45)

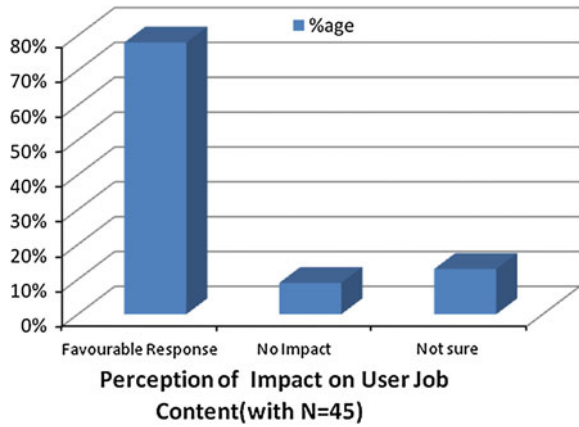
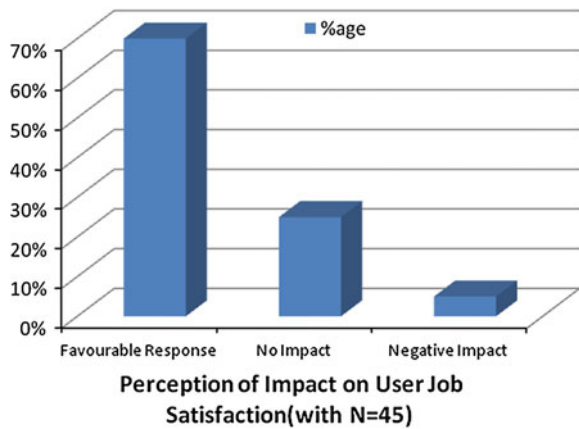


Fig. 4 Perception of expert system impact on user job satisfaction (with N = 45)



8.3 User Job Satisfaction

Seventy percent of the respondents perceived that the ES had resulted in a “positive” or “somewhat positive” impact on the job satisfaction of the users. Twenty-five percent felt that there had been no impact on satisfaction, while 5 % thought that their ES had made a negative impact (see Fig. 4). When coupled with the results related to job content, these findings indicate that ES technology can have favorable social impacts on users.

9 Conclusion

This study builds on the perception and the experiences of the end users. Base on the empirical findings, implications regarding the ES management practices were suggested.

The survey of implementation issues did produce two unexpected results. First, ES user training and management educational activities were not generally considered to be critical. This finding was surprising since these activities are typically cited as an important component of the system implementation process. The second unexpected finding was the relatively low rating for ES technical expertise. As noted earlier, this result may indicate that established technical competence is not always necessary for developing successful ES projects of the type generally represented in the survey (i.e., medium-sized ES developed on a personal computer or workstation platform). The study also provides insight into an area which has not received much attention from empirical researchers: the impacts of ES. The findings related to the social impacts of ES were especially interesting. Although some authors have suggested that ES may have an unfavorable impact on the social aspects of the work place, the results of this study indicate otherwise. In a majority of the cases, the ES was perceived to have a beneficial impact on job content and user job satisfaction. The investigation into operational impacts also yielded some interesting results. While ES were considered to have a positive effect on a variety of operational aspects, the value added benefits of ES (e.g., improvements in consistency and quality) were rated more highly than the productivity-related impacts (e.g., time needed to perform task). In some cases, the value-added benefits of ES appeared to compensate for perceived negative impacts related to productivity. These results underscore the relative importance of ES operational impacts which are intangible in nature. The descriptive study has provided insight into the management of ES. However, there are limitations to a study of this type. Also, the survey methodology does not allow for an evaluation of the dynamics of the implementation process. Additionally, the study generally focussed on medium-sized ES used at the operational level of an organization. The findings from this study may or may not generalize to different types of ES projects. Although there is much to learn about the management of ES, we expect that there will be more opportunities for researchers to explore this area as ES technology continues to migrate into the work place.

References

1. Tripathi, P., Islam, S.N., Ranjan, J., Pandeya, T.: Developing computational intelligence method for competence assessment through expert system: an institutional development approach. In: Computational Intelligence and Computing Research (ICIC), 2010 IEEE International Conference on. 28–29 Dec 2010, pp. 1–5. doi: [10.1109/ICIC.2010.5705747](https://doi.org/10.1109/ICIC.2010.5705747), URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5705747&isnumber=5705719>

2. Garrett, S.: Competency mapping: what is it and how it can be done by individuals. *Career Plan. Adult Dev. J.* **18**(Part 4), 43–59 (2003). Available at: <http://www.careertrainer.com/Request.jsp?lView=ViewArticle&Article=OID%3A112409>. Accessed 12 Mar 2012
3. Lepsinger, R. et al.: *The Art and Science of Competency Models: Pinpointing Critical Success Factors in Organizations*. Jossey-Bass/Pfeiffer, San Francisco (1999)
4. WiseGEEK: What is Competency Mapping? Available at: <http://www.wisegeek.com/what-is-competency-mapping.htm>. Accessed 12 Mar 2012
5. McClelland D.C.: Testing for competence rather than for intelligence. *Am. Psychol.* **28**, 1 (1973)
6. Pagey, U., D’Silva, L.: Competency mapping, *Express computer*, 31. 1. 2005. Available at: <http://www.expresscomputeronline.com/20050131/technologylife01.shtml>. Accessed 12 Mar 2012
7. McLagan, P.: Competencies: the next generation. *Training Dev.* **51**, 40–47 (1997)
8. Hamel, G., Prahalad, C.K.: The core competence of the corporation. *Harvard Bus. Rev.* **68**(3), 79–91 (1990)

Implementing Breath to Improve Response of Gas Sensors for Leak Detection in Plume Tracker Robots

Ata Jahangir Moshayedi and Damayanti Gharpure

Abstract Odor leak localization is playing a critical role in industry, especially in the case of combustible and toxic gas as well as security applications. This research has tried to address the need of a sniffing system to improve the gas sensor (TGS 2620) response. The novel algorithm named “Breath ADJ” based on biological sniffing and smelling is implemented in a sniffing system designed. And tested at various distances from the source. The Breath ADJ algorithm is aimed at realization of breath inhalation and exhalation to suck the odor on the sensor for improving the response and clean off the odor during recovery period of sensor response. The improvement in the system response has been evaluated by three features of slope during the rise of sensor response, Odor reach time and a maximum value of sensor response. The results based on an optimum Breath ADJ algorithm are promising. The algorithm is successful in sensing/detecting odor at a distance of 90 cm from the odor source in an indoor arena and 60 cm in outdoor door arena.

Keywords Gas sensor · Odor pulse leak · Breath ADJ · Leak localization · Plume tracking robot

1 Introduction

The SnO₂ gas sensors due to their response and availability are used in E_{nose} [1], plume tracking [2], hazardous gas alarm detector [3], food industries [4], etc. However, their slow recovery time, drift by changes in the humidity and

A. J. Moshayedi (✉) · D. Gharpure
Department of Electronic Science, University of Pune, Pune 411007, India
e-mail: moshayedi@electronics.unipune.ac.in

D. Gharpure
e-mail: dcg@electronics.unipune.ac.in

temperature [5] in the environment and power consumption are the main demerits of this type of gas sensor. This inherent problem becomes magnified in applications like plume tracking [6] and leak detecting [7] wherein, the Robots are delegated to do the mission. Successful efforts to reduce power consumption are reported in previous work [8], but the sensor response and recovery are still a live challenge for scientists. In addition; the rise and fall time of the sensor response has a direct relation to the sensor base line which varies with changing humidity and temperature [5, 9]. The rising and falling part from sensor response perform a major role in odor detection and it's performance. The sensor can sense the odor faster with a proper response from the sensor side. Explicitly; improvement in this area, will assist sensible reaction from the robot based on gas detectors. Slow rising and falling response may lead to a tragedy in the case of factories and industries as the detector cannot recognize a leak within time [10]. Previously; to solve above problem two types of attempts are introduced. One based on Odor evacuation with the help of fresh air [11] as the traditional and useful method is still used in e_nose systems. The main drawback with this method is the weight of assembly specially in the case of using on robots. Besides, as the environment will have an odor and air mixture the same will be introduced as fresh air for sensors. The second called Mark II [12] which removes the odor and air using a fan behind the sensor. The air is sucked from the back side of the pipe and carries the air containing odor. This action causes fluctuation in sensor response in addition to cleaning the sensors.

Researchers have tried to implement various strategies, inspired by the creature's behavior [2, 13–15], for plume tracking and odor source localization. Breath as the key factor of life, is one of the remarkable parts of creature's behavior. Generally; in creatures, the mixture of two functions, sniffing and breathing, cause the automatic smelling [16]. Breathing can be divided into a continuous cycle of inhaling and exhaling. During breathing also we can smell and sometimes to have more accuracy for smelling one takes a deep breath. The exhaling and inhaling are aimed at removal CO_2 and absorption of O_2 in the blood. In the olfaction, major portion belongs to odor sensory breath. Among creatures, canines are more famous for their odor detection and fast classification. Recent research shows that this ability is not only just due to their number of receptors but it is also linked to the receptor place as well as a systematic way of breathing for smells by rotating the odor in their nasal cavity. The experiments show dogs sniffing contains multiple continuous breath in the range of 0.5–1.5 Hz [17]. A human has the breathing frequency of around 5 Hz. This result has also been simulated with the help of CFD at the 5 Hz frequency [18]. By all these facts; the same procedure can be employed for gas sensors.

En route to improve the gas sensor response, we propose a procedure based on the breath, to flush gas sensors and make the response faster. To best of the author's knowledge of the open literature there has been no such this implementation of breathing for gas sensors yet. The main scope of this paper is trying to improve the ability of gas sensor with new algorithm named breath ADJ. The performance has been studied, using a leak source with respect to distance and

response time as well as the fans power consumption. The rest of this article is structured as follows. First; a brief summary of gas sensor response and features, next; the experiment setup, environment and proposed sniff hardware are described, followed by the breath ADJ algorithm and various experiments carried out for algorithm evaluation. The paper ends with results, discussion, conclusion and future work.

2 Gas Sensor Response and Features

As mentioned before; the rise and fall part have the critical importance in sensor response. The SnO₂ sensor response can be divided into two major parts. Initial “turn on” sensor response, after switching sensor heater on, to reach a temperature of 300, 400 °C by the inrush current. The sensor response part (initial turn on response) due to its dynamic behavior, is generally neglected in all calculation and doesn’t have significant information. That’s why it’s highly recommended to keep the sensors turned on, for some time after switching [19]. Second part consists of: Sensor odor response by change in the sensor resistor (Rs). This response happens subsequently after reaching the proper temperature from sensors heater side and absorption of the odor by the sensor. Usually; we know this response as the sensor response.

As it is shown in Fig. 1, the sensor odor response for the leak pulse itself can be divided into five main parts: Baseline_1, initial change of sensor response, Transient time, Maximum value (Peak value), Falling time, Baseline_2 [20]. Some researchers based on their requirements tried to use one or whole of this response [21, 22]. Among this parts: baseline 1 and baseline 2 depends on temperature and humidity. Maximum value depends on odor concentration as well [9]. The area which is considered in research work is the sensor response after proper “heater On time”. Also; to investigate the sensor response, the maximum value, slope of rising part and the transition point indicating the odor reaching the sensor are the features selected for this work [20].

3 Experiment Setup

3.1 Experiment Environment

The experiments are conducted in two environments as shown in Fig. 2 First; it is done on the table (106 cm × 70 cm × 73 cm in the corner of the room with dimension of 329 cm × 319 cm × 118 cm, named “Outdoor arena”. The room consists of two windows on one side (at a height of 90 cm from the ground), with one of the window, half opened and one door. The second arena is the closed

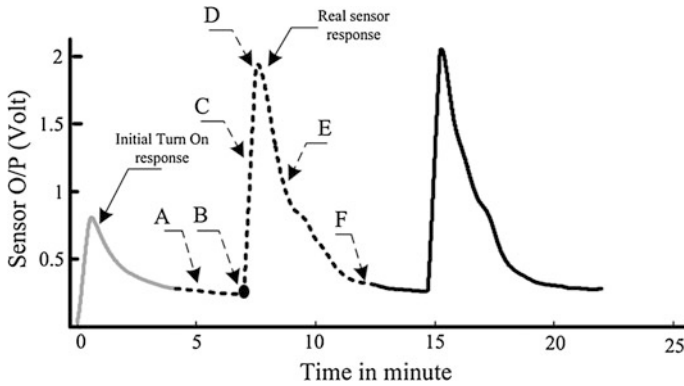


Fig. 1 Sensor initial response and real for the odor pulse, A baseline 1, B initial change in sensor response, C transient time, D max value (peak value), E fallen time, F baseline 2

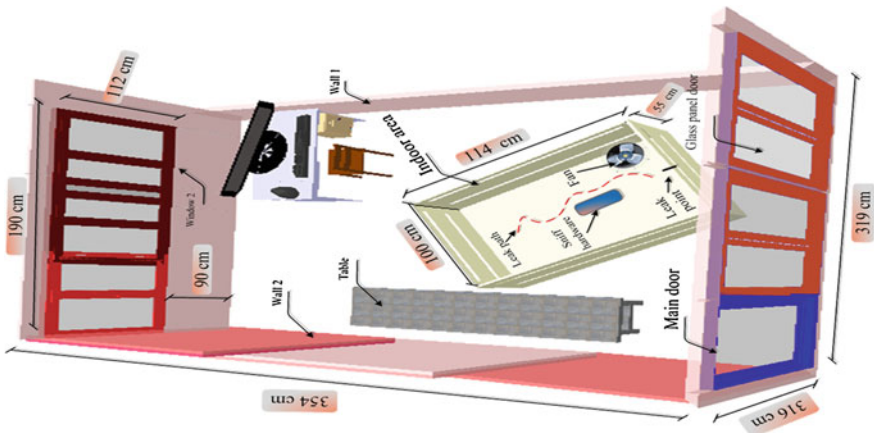


Fig. 2 Experiment arena

chamber designed with the dimension of 114 cm × 100 cm × 55 cm [23] called “Indoor arena”. The ceiling fan and the small fan in the chamber are used for ventilation.

3.2 Proposed Sniff Hardware

The sniffing hardware consists of two fans. One in front and the other at the back (5 V from ADDC company with the diameter of 4 cm and 5,000 RPM) placed at the two sides of aluminum pipe (length = 14 cm, diameter = 6 cm, thickness = 0.0762 cm) (see Fig. 3).

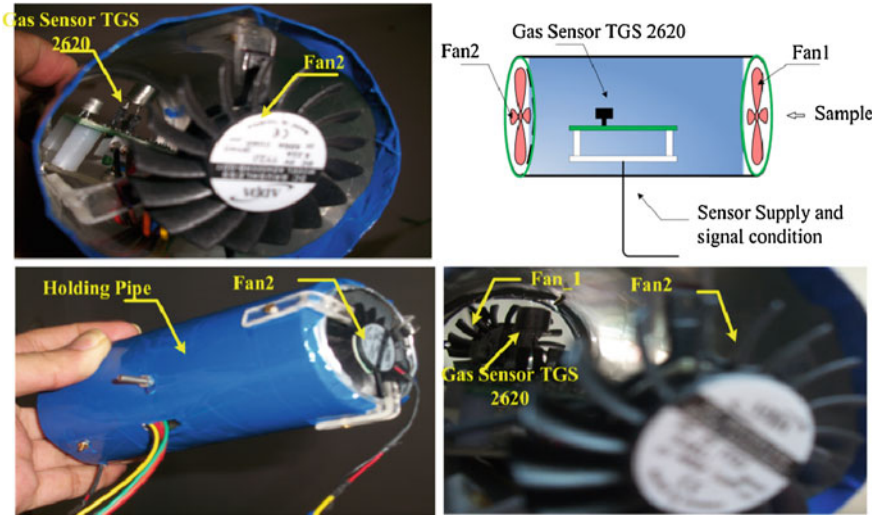


Fig. 3 Sniff assembly (inside the gas sensor enclosure)

In this assembly as it is shown in Fig. 3, the 2620 TGS gas sensor with the compact signal conditioning board is mounted in the middle of aluminum pipe. Besides; for better sensor performance, the sensor head is positioned just ahead of the fans vanes.

3.3 The Leak Odor Delivery

The odor leak delivery is controlled by a microcontroller to switch the solenoid valves (5 V, 50 SIG) from General valve Company to deliver the ethanol odor automatically based on different timing algorithms (see Fig. 4a, b).

As Fig. 4 shows, the aquarium pump (with max speed 3 lit/s) and to control the inlet air, the 3 lit/m Rota meter from the Eureka company has been used. A microcontroller is used to take care of time management based on different experiment algorithms described in the next part. The sensor data is logged by Labview 2012, (PCI 6024 data acquisition card) at the sampling frequency of 500 Hz. During the experiments, the leak source with a nozzle size of 1.5 mm is made based on a bubbling method [24, 25]. The leak is placed at the center of the left wall of the chamber at a height of 11 cm from the ground at the same height of sniffer hardware from the ground. All experiments are carried out with ethanol gas leak of 50 PPM concentration, calculated based on [24]. The leak pulse period was 20 s with the speed of 600 CC/M measured by DF04 form PCI analytical digital flow meter instrument [26].

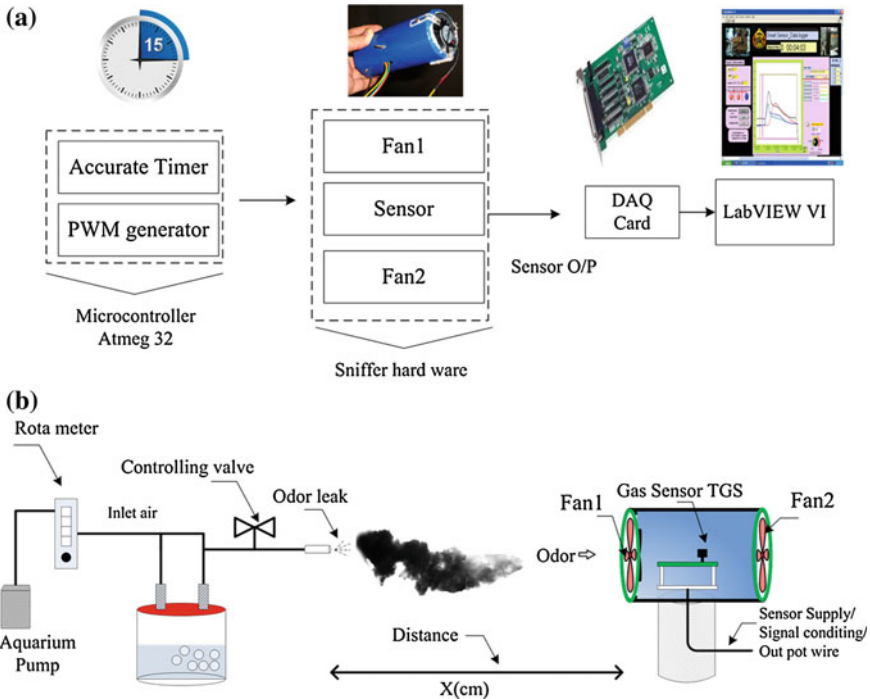


Fig. 4 a odor leak delivery setup block diagram, and b real setup

3.4 Algorithm Exploration

The Fig. 1 shows the real sensor response inside the pipe for the pulse mode odor. As mentioned before; the main target of this research is to use the fan effectively to boost up the gas sensor response inside the pipe. The procedures used to test the effectiveness of the breath implementation idea has been done in three steps: Step I; investigation of sensor response with one of the fans (front side or back side) and another sensor without fans inside the pipe. Step II. Studying the effect of variation in on/off timing of the fans on sensor response. Step III optimizes on/off time for breathing algorithm and comparison with the bare sensor response. The sensor response in all experiments was analyzed based on the three features described in the previous section. In all three steps, the data was logged in 700 files over a period of 3 months and analyzed.

In the first step the sensor response based on the three features (initial change points in response, maximum point response, and maximum to end response), as described before. The experiment is conducted to monitor sensor response inside pipe, under three conditions in both indoor and outdoor arena.

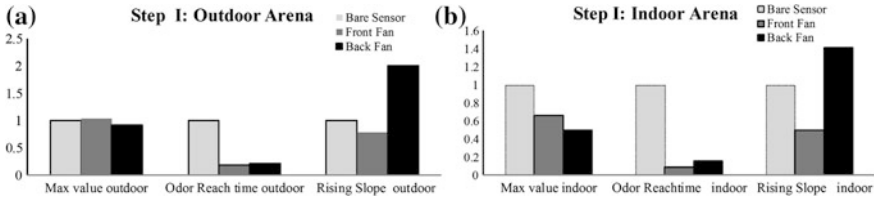


Fig. 5 Sensor response in room and chamber arena

- Bare Sensor response without fans
- Sensor response along with front fan in continuous ON mode
- Sensor response with back fan in continuous ON mode.

The comparison has been done based on the time taken by odor to reach the gas sensor in three cases of using only sensor, front fan and back fan with a distance of 18 cm from the source and a 20 s odor pulse. After 3 min the fans were switched “On” and the doors and windows were opened for ventilation. The total duration for each experiment was 6 min. The result of this step in room environment as well as chamber improvement with respect to bare sensor response is shown in Fig. 5.

As the Fig. 5a shows, improvement in the sensor response in the rising slope and odor reach time specifically for the sensor with back fan. But the maximum value is nearly the same in all three cases. This phenomena may be caused by the natural turbulent wind inside the room. So to avoid this effect and have the better observation the sniffer hardware was shifted to the chamber and the same steps with the same distance are repeated. In this phase the small AC fan inside the chamber, acted as the ventilator to purge the chamber after 3 min of experiment. The experimental result is presented in Fig. 5b. The results demonstrate the same sensor response as outdoors, in case of rising slope and odor reach time as well as maximum value. This may be caused by the inherent airflow in the chamber and fan speed which bring in and remove the odor fast from sensors. The fast input and flushing of odor leads to poor response due to the slow response time of sensors which is minimum 10 s.

Considering the variation in temperature and humidity in the months of May to July, to have a more controlled environment, the rest of the experiments were conducted in the chamber. Further the results depicted in Figure indicate that, for obtaining better sensor performance the use of fans odor delivery/purge time need be optimized. Also the leak injection and fans odor delivering/take out times for the sensor should be synchronized.

According of computational fluid Dynamic [27] the time needed to fill the pipe is calculated from (1):

$$t_{\text{fill}} = \frac{\text{Area} * \text{Length}}{\text{Area} * \text{Velocity}} = \frac{\text{Length}}{\text{Velocity}} \tag{1}$$

Then based on the dimension and velocity of wind which is generated by fans = 0.7 m/s (measured by Lutron AM 4202 digital anemometer), the filling time (t_{fill}) was calculated (2)

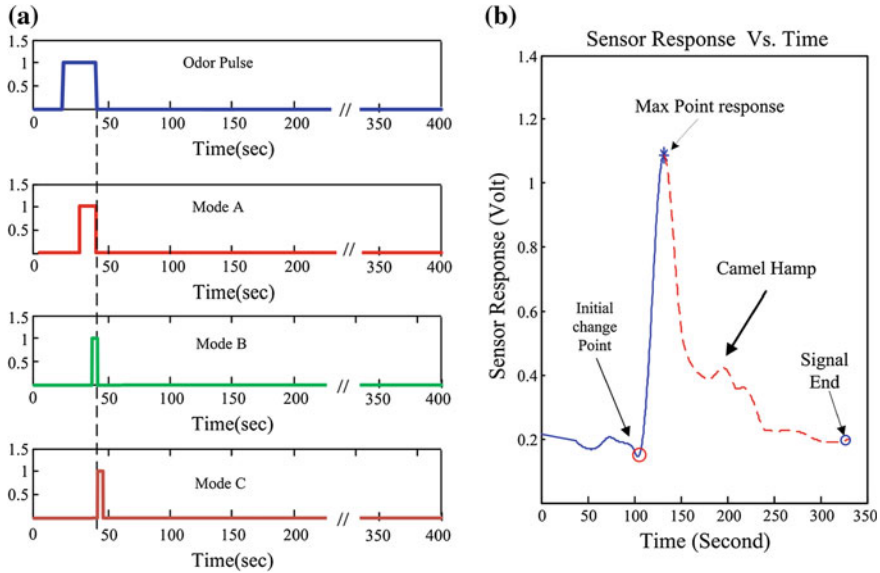


Fig. 6 a Modes timing diagram. b Mode effect on sensor response. *Modes A* Turn ON front/back fan on the last 5 s of odor ON time, *Modes B* Turn ON the front/back fan on the last 1 s of odor ON time, *Modes C* Turn ON the front/back fan 1 s after finishing the odor injection

$$t_{\text{fill}} = \frac{0.14 \text{ m}}{0.7 \text{ m/s}} = 0.2 \text{ s} \tag{2}$$

This time (t) is needed to fill the pipe and bring the odor in front of the sensor but since the sensor response is slow (10 s), the odor delivery time was estimated as minimum 1 s. Experiments were carried out to study sensor response, varying the t_{fill} time and t_{purge} times by switching the fans' On/OFF with respect to the odor pulse duration as shown in Fig. 6.

This was investigated under the three modes with the same distance of 18 cm from 20 s odor leak pulse over a period of 6 min. The Fig. 6 shows the modes in the form of timing diagrams with the 20 s odor pulse as a reference in the first diagram.

The sensor response obtained was analyzed with these modes as the primary stage of breath algorithm, to lead the experiments for proper On/Off timing of fans. The result obtained are shown in Fig. 7a-c.

As the Fig. 7 shows; the sniffer sensor response compared with bare sensor have improved performance in all three modes. The figure indicates the max value for both fans in mode A and B are same and in mode C back fan has the better response. Besides; the odor reaching a time in mode B and C are for both fans nearby and in mode A back fan has better behavior. On the other hand; rising slope for both fans in mode A are close to each other and in modes B and C the back fan performed better. But one drawback of results obtained with back fan is that at times the sensor shows a response like camel Hump (as shown in Fig. 5b).

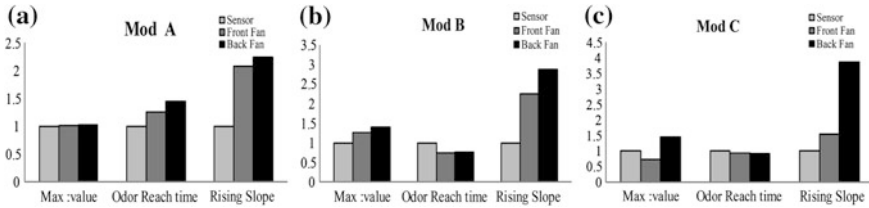


Fig. 7 Modes result base on sensor response features

3.5 Step III: Breath ADJ Optimum Algorithm

The different switching time between fans are experimented with. The odor absorption in proper time and permit the sensor to have a response and remove the odor in the correct time was the main aspect that was investigated in this phase via considering this point. Finally; the optimum breath ADJ algorithms to achieve the best sensor response is implemented as shown in Fig. 8. The odor leak (odor pulse) is injected for a period of 20 s. At the last 8 s of odor leak pulse time, the front fan is switched on to deliver odor near the sensor. After which both fans are switching Off for 26 s, when the odor is allowed to stay in the pipe and then for 8 s the back fan is switched On to evacuate and clean the pipe. Above cycle is repeated for the next batch of odor leak.

4 Algorithm Evaluation

To evaluate the Breath ADJ algorithms, the sensor performance is compared with the bare sensor mounted on a pipe. The Fig. 9 shows the response obtained with the breath ADJ algorithm on the bare sensor as well as a sensor with fans at a distance of 18 cm from the source. The figure depicts the fan On OFF sequence with respect to odor leak timing along with sensor response zones.

The next stage of experiments was to study the system performance in terms of distance from the odor source of sniffer sensor and the bare sensor. The effect of distance from the odor leak source, in the range 18–90 cm in steps of 18 cm are investigated. The result of different distance and three features of gas sensor response are shown in Fig. 10.

The Fig. 10, shows the breath ADJ algorithm has the good performance on sensor responses. The best response was obtained at 18 cm, which is the nearest place with the odor and the minimum value was observed at 90 cm.

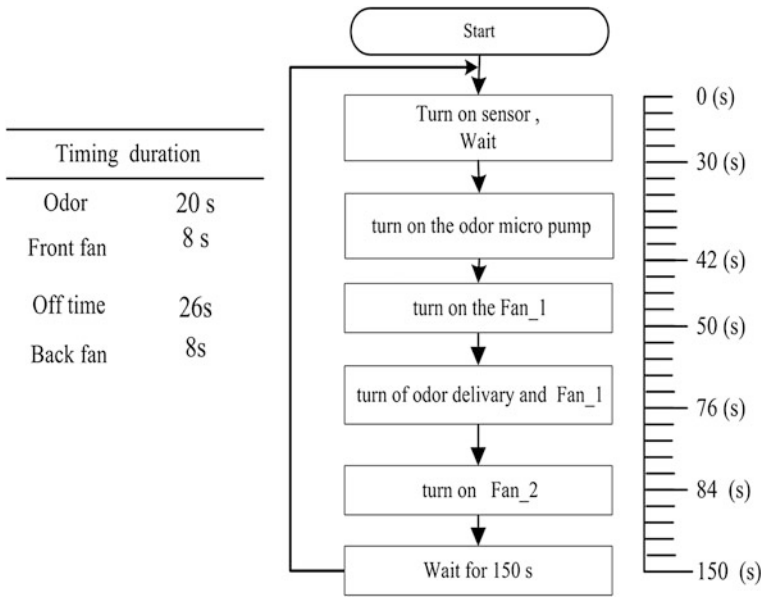


Fig. 8 Breath ADJ algorithm

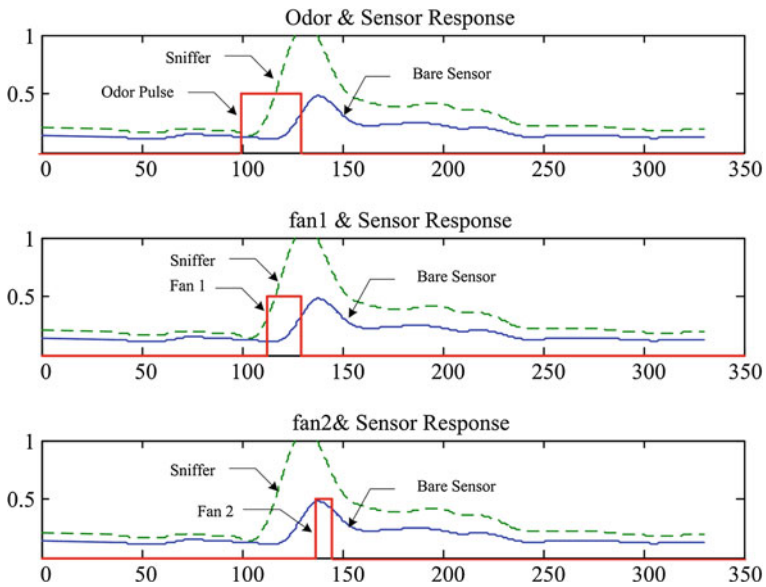


Fig. 9 Sniffer and bare sensor performance for 18 cm

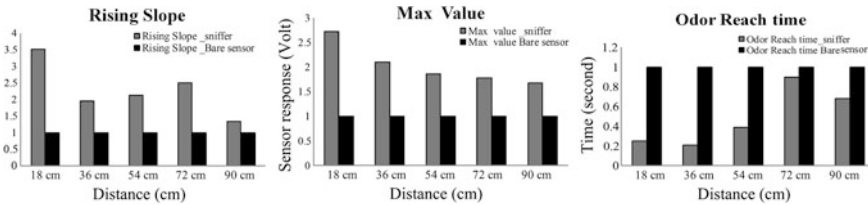


Fig. 10 Sniffer and bare sensor performances location change

5 Conclusion and Future Works

This research focuses on improving the sensor response in case of Leak Odor source localizing for the plume tracking robot applications. This work is attempted to improve the gas sensor TGS 2620 response by the optimum timing for odor absorbing and removing. The breath ADJ algorithm as a novel process with the help of sniffing assembly is proposed. The result of the comparison between three features (rising slope, odor reach time and maximum value) as the most important sensor response are showing the success in improving the sensor response in the case of odor source leak. Figure 10 shows improvement in sensors’ responses with the new algorithm. The algorithm has been tested at various distances and experiments indicate successful detection over a range of 60 cm from the source in an outdoor arena and 90 cm in the indoor arena (see Fig. 11) which is more than reported [12]. The ADJ algorithm can be implemented in plume tracker robots [8] to follow and localize the odors faster which is the next step and future work.

References

1. Arshak, K., Moore, E., Lyons, G.M., Harris, J., Clifford, S.: A review of gas sensors employed in electronic nose applications. *Sens. Rev.* **24**(2), 181–198 (2004)
2. Farrell, J.A., Pang, S., Li, W., Arrieta, R.: Chemical plume tracing experimental results with a REMUS AUV. In: *Proceedings of the OCEANS 2003*, vol 2, pp. 962–968 (2003)
3. Krajci, J.: Gas detection system and method. Google Patents, 06 Feb 2001
4. Schaller, E., Bosset, J.O., Escher, F.: ‘Electronic noses’ and their application to food. *LWT-Food Sci. Technol.* **31**(4), 305–316 (1998)
5. Abidin, M.Z., Asmat, A., Hamidon, M.N.: Identification of initial drift in semiconductor gas sensors caused by temperature variation. In: *IEEE 9th International Colloquium on Signal Processing and its Applications (CSPA)*, pp. 285–288 (2013)
6. Ishida, H., Nakamoto, T., Moriizumi, T., Kikas, T., Janata, J.: Plume-tracking robots: a new application of chemical sensors. *Biol. Bull.* **200**(2), 222–226 (2001)
7. Vítkovský, J.P., Simpson, A.R., Lambert, M.F.: Leak detection and calibration using transients and genetic algorithms. *J. Water Resour. Plann. Manage.* **126**(4), 262–265 (2000)
8. Moshayedi, A.J., Gharpure, D.C.: Priority based algorithm for Plume tracking Robot. In: *1st International Symposium on Physics and Technology of Sensors (ISPTS)*, 2012, pp. 51–54 (2012)

9. Pashami, S., Lilienthal, A.J., Schaffernicht, E., Trincavelli, M.: TREFEX: trend estimation and change detection in the response of MOX gas sensors. *Sensors* **13**(6), 7323–7344 (2013)
10. Monitoring Large Hydrogen Sulphide Releases Why Sensor Recovery Time is Important. http://www.petronline.com/articles/safety/15/edward_naranjo_ph.d/monitoring_largehydrogen_sulphide_releases_why_sensor_recovery_time_is_important_/1305/, Aug 2013
11. Lochmatter, T., Raemy, X., Martinoli, A.: Odor source localization with mobile robots. *Bull. Swiss Soc. Aut. Contr.* **46**, 11–14 (2007)
12. Lilienthal, A., Duckett, T.: A stereo electronic nose for a mobile inspection robot. 1st international workshop on robotic sensing, 2003 (ROSE'03), p. 6–pp (2003)
13. Kanzaki, R., Minegishi, R., Namiki, S., Ando, N.: Insect–machine hybrid system for understanding and evaluating sensory-motor control by sex pheromone in *Bombyx mori*. *J. Comp. Physiol. A*, 1–16 (2013)
14. Edwards, S., Rutkowski, A.J., Quinn, R.D., Willis, M.A.: Moth-inspired plume tracking strategies in three-dimensions. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005 (ICRA 2005), pp. 1669–1674 (2005)
15. Guillot, A.: Biologically inspired robots. In: Springer Handbook of Robotics. Springer, pp. 1395–1422 (2008)
16. Grote, C., Pawliszyn, J.: Solid-phase microextraction for the analysis of human breath. *Anal. Chem.* **69**(4), 587–596 (1997)
17. Craven, B.A., Paterson, E.G., Settles, G.S.: The fluid dynamics of canine olfaction: unique nasal airflow patterns as an explanation of macrosmia. *J. R. Soc. Interface* **7**(47), 933–943 (2010)
18. Elad, D., Lieberthal, R., Wenig, B.L., Einav, S.: Analysis of air flow patterns in the human nose. *Med. Biol. Eng. Comput.* **31**(6), 585–592 (1993)
19. Figaro: Technical Information for TGS2620 (2012)
20. Gibson, T.D., Prosser, O., Hulbert, J.N., Marshall, R.W., Corcoran, P., Lowery, P., Ruck-Keene, E.A., Heron, S.: Detection and simultaneous identification of microorganisms from headspace samples using an electronic nose. *Sens. Actuators, B* **44**(1), 413–422 (1997)
21. Gutierrez-osuna, R., Nagle, H.T., Schiffman, S.S.: Transient response analysis of an electronic nose using multi-exponential models. *Sens. Actuators, B* **61**, 170–182 (1999)
22. Ishida, H.: Robotic systems for gas/odor source localization: gap between experiments and real-life situations. In: Proceedings of the IEEE International Conference on Robotics and Automation, 2007 (ICRA 2007), pp. 3–8 (2007)
23. S. Beirne, B. Corcoran, K. T. Lau, and D. Diamond, “Chemical event tracking using a low-cost wireless chemical sensing network,” in *Sensors, 2008 IEEE*, 2008, pp. 1615–1618
24. Zhang, H., Kulkarni, A., Kim, H., Woo, D., Kim, Y.-J., Hong, B.H., Choi, J.-B., Kim, T.: Detection of acetone vapor using graphene on polymer optical fiber. *J. Nanosci. Nanotechnol.* **11**(7), 5939–5943 (2011)
25. Roussel, S., Forsberg, G., Steinmetz, V., Grenier, P., Bellon-Maurel, V.: Optimisation of electronic nose measurements. Part I: methodology of output feature selection. *J. Food Eng.* **37**(2), 207–222 (1998)
26. Analytical, P.: Digital Gas Flow Meter DFM04, vol. 91(22). p. 9001 (2008)
27. Fox, R.W., McDonald, A.T., Pritchard, P.J.: Introduction to Fluid Mechanics, vol. 2. Wiley, New York (1998)

Use of Simulation and Intelligence Based Optimization Approach in Bioprocess

Pavan Kumar and Sanjoy Ghosh

Abstract Enzyme streptokinase is produced by *streptococcus* sp. in native form and useful to treat acute myocardial infarction, being an essential drug it is necessary to enhance its production utilizing its recombinant strain for thrombolytic therapy. Various fundamental models incorporating indispensable parameters are found to apparently describe the entire existence of employed cells in the bioreactor environment. The unstructured system features can be defined by dynamical system using a composite model. The endeavour would be to establish the fundamental constraints that affect the plasmid instability criterion and hold a relevant role in dynamics of batch and continuous culture system. On performing statistical analysis, screening of production media components and culture condition optimization has been achieved; the data obtained noticeably illustrates the role of few significant parameters governing the culture system. A useful technique has been further implemented using neural network simulation which on the other hand serves as soft computing tool for optimizing factors controlling the process dynamics.

Keywords Streptokinase · Unstructured system · Instability criterion · Culture condition and process dynamics

P. Kumar (✉) · S. Ghosh
Computational Bioprocess Engineering Laboratory, Department of Biotechnology, Indian Institute of Technology Roorkee, Roorkee 247667, India
e-mail: pawanbioinfo22@yahoo.com

S. Ghosh
e-mail: ghoshfbs@gmail.com

1 Introduction

The Streptokinase is regarded to be a well known thrombolytic drug, biochemically a fibrinolytic activator that promotes the formation of enzyme plasmin from plasminogen. The ampicillin resistance gene *skc* have been taken into *E. coli* in order to serve in selecting overproducer clones. Research workers have over-expressed the gene for enzyme streptokinase in *E. coli* [1], the constructs are transformed into strain BL21 and allowed to grow in LB medium contained ampicillin in required concentration [2], until OD 0.6 is attained, at 600 nm. Then induction of the *E. coli* cultures is carried out using 0.8 mM IPTG. The streptokinase samples obtained are analyzed for their expression profiles. Presently the effort is to maximize the sustenance of plasmid vector in cells while culturing in production media, the longer time retention of plasmid with high copy number can provide a high magnitude of production with the help of expression. Streptokinase assay to test its activity of culture supernatant fluids was done by comparison with dilutions of standard purified streptokinase solution, using casein/plasminogen plate technique [3].

Various training procedures proposed for the hybrid modelling approach is generally based on gradient optimization method, so performing repeatedly the exploration analysis with different start vectors of the ANN weights should be taken and the network led to minimal “test error” is to be adopted [4]. A feed forward neural network type with multi-layer perceptron is generally used [5]. The Feed-forward ANNs are made to include modified dynamic characteristics with incorporated estimator model representation. Neural Network based tool can be utilized to serve as a means for modeling non-linear control process [6]. Statistical technique for screening and design experiments should be carried out to determine which of the several experimental variables and their interactions present more significant effects [7]. ANOVA was constructed for the second order response surface model; the significance of each coefficient was determined by Student’s *t* test and *p*-value, to identify the corresponding significant coefficient. A numerical method can be used to solve the regression [8]. Thus statistical technique is also compared to neural network approach in optimization process.

2 Relevance of Unstructured Computational Models

The unstructured model used to depict interaction of factors in bioprocess operation which in turn decide the magnitude which is considered to be of prime importance. Streptokinase production is directly associated to stability and sustenance of plasmid in the recombinant cells [9]. In bioprocess operation, overall probability of plasmid loss can be given by,

$$p_{pl} = f(p_B, p_L, m_C, \mu_L, t)$$

We may represent a part of the dynamic model, as

$$\begin{aligned} dp_{pl}/dt &= (1 - e^{-f_m(p_L/p_B + p_L)})\mu_L \\ dm_C/dt &= Y_{MC}(p_B\mu_B + p_L\mu_L) - D_R m_C \end{aligned}$$

p_B , p_L Concentration of two cell populations (g/l), p_B —plasmid bearing, p_L —plasmid lacking cells, m_C —Metabolite concentration (g/l), Y_{MC} Yield in terms of metabolite, D_L is the dilution rate, f_m Collective probability factor, μ_B and μ_L are specific gr. rates (1/h) of plasmid free cells t Time (h). Above parameters are in turn depends on several other variables being the function of other parameters. Plasmid loss probability (p_{pl}) which is a function directly found to govern the number of plasmid bearing and lacking cells in the preceding and subsequent generations.

2.1 The Dynamical System Simulation

The model simulation for continuous culture operation has been done employing Matlab R2010a. At initial time $t = 0$, for different parameters [10] initial values have been used. The process simulation is done using various standard data values of previous model meant for streptokinase [12] including various other validated model constraints with minor magnitude presumed for different set of parameters on the basis of their role in regulating the dynamics. Figures 1 and 2 shows the relative population dynamics of cells in the bioreactor system.

2.2 Typical Decline of Percentage of Plasmid Bearing ($\mu+$) Cells

It is well evident that the low dilution promotes the generation of higher magnitude of $\mu+$ cells concentration. The event of varying probability of plasmid loss has its sole impact over the entire dynamics of the model by governing the associated substantial factors. Although a threshold amount of metabolite is considered to make the consequence in relative dynamics of two cell types. The dynamics pertaining to plasmid average copy number in batch culture system shows that copy number has somewhat decreasing trend after 4–6 h and it becomes negligible in magnitude after 12 h of culture time.

Fig. 1 Simulation plot shows diversion of plasmid bearing and lacking cells types at $D = 0.15$ during growth dynamics, also matched with experimental findings of certain duration [11] in respect to streptokinase/plasmin activity after induction in mg/l

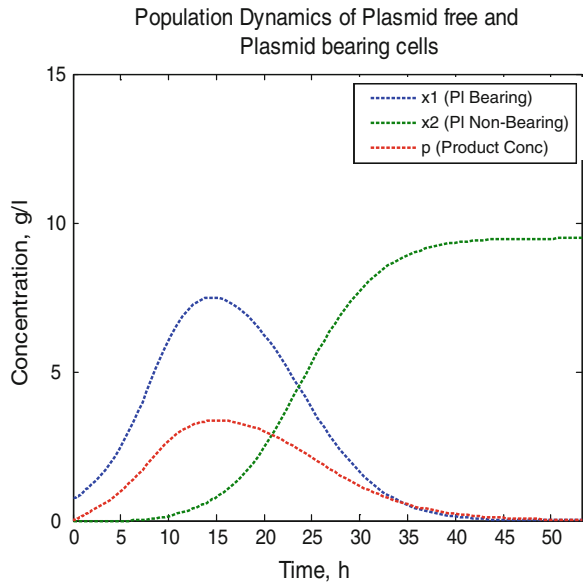
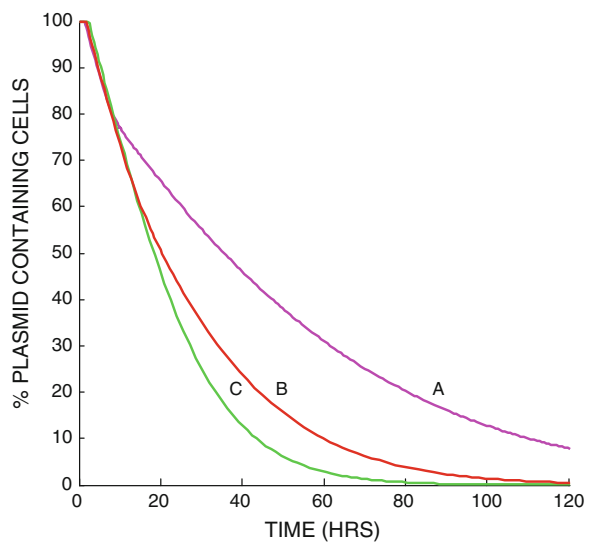


Fig. 2 Decline in percentage of recombinant cells population at different dilution rates $D_A = 0.57$, $D_B = 0.40$ and $D_C = 0.23$, obtained after simulation, the observation is compared with data trend obtained from experimental results [12, 13]



3 Neural Network Based Bioprocess Supervision

Recently artificial intelligence working paradigm has been implemented to bioprocess field in order to achieve a high degree of accuracy in process optimization and prediction. We may obtain the desired output for specific data inputs merely

Fig. 3 A general plan neural network with feed forward type architecture, showing input, hidden and output layer

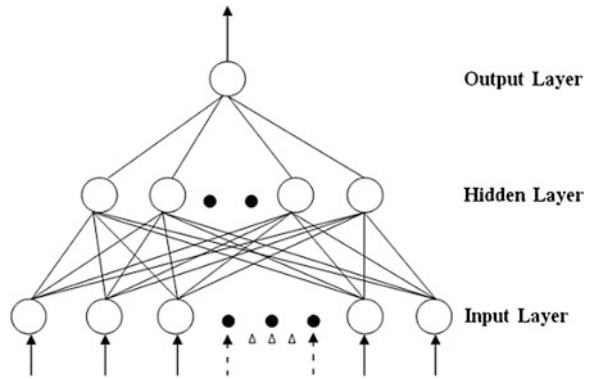
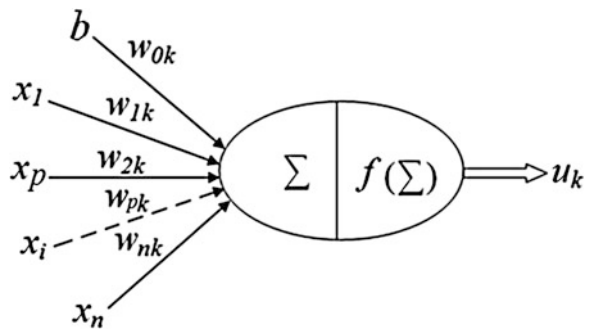


Fig. 4 An overall working framework of feed forward neural network used for non-linear data processing and analysis



by adjusting the strengths of artificially simulated neuron in connecting layer. ANNs are often used for engineering purposes to process information and control automated systems. From Neural Network Architecture, output is given by,

$$u_o = f(x_p, \omega_p)$$

Simply,

$$u_o = \sum x_p \omega_p$$

where, x_p —are inputs, u_o is output vector and ω_p —are the weights.

Activation function, $a_f = 1 / (1 + e^{-up})$.

By adjusting the weights of an artificial neuron we can obtain the desired output for specific inputs. Intelligent descriptions of microbial processes in bioreactors have some decisive advantages over mechanistic model under realistic conditions and has wide range of flexibility.

A general plan and working of Neural Network is framed out in Figs. 3 and 4. Neural network has defined mechanism of adjusting the weights through learning or training. In our case the data is first normalized using upper and lower bound of

the numerical value range and then the normalized data is used to serve as input for the neural network model.

Error Minimization Approach

A multilayer perceptron is designed to process the weighted input. In our case sigmoidal function is used as an activation function in hidden layer using which the output is obtained. The given inputs are different magnitude of culture conditions while the output is taken in terms of streptokinase production with time duration. To ensure optimal memory for the network ten hidden layers are taken after optimizing the number of layers to ensure optimal memory for the neurons. On obtaining the experimental data close to the simulated results, suggests the superior capability of neural network modelling for dynamic behaviour of the system [14]. The functional theme of ANN is clearly configured in framework, shown in Fig. 4.

- Error Function ϵ depends on the assumed weight values
- Gradient Descent: To minimize error by shifting weights along the decreasing slope of error
- The idea is just to iterate through the training set and adjusting the weights to minimize the gradient of the error.

The process of gradient descent is made to minimize the error via iteration of training set and regulate the weights by adjusting it along descending slope of the computed error,

$$\epsilon = \sum_{x_p \in E} (\theta_p - \phi_p)^2$$

The gradient of ϵ :

$$\frac{\partial \epsilon}{\partial \omega} = \left[\frac{\partial \epsilon}{\partial \omega_1}, \dots, \frac{\partial \epsilon}{\partial \omega_p}, \dots, \frac{\partial \epsilon}{\partial \omega_{n+1}} \right]$$

The final output obtained after passing through the hidden layers is expressed as;

$$\begin{aligned} O_q &= f\left(\sum_r w_{rq} u_r + w_{0q}\right) \\ &= f\left(\sum_r w_{rq} \left(f\left(\sum_p w_{pr} x_p + w_{0r}\right)\right) + w_{0q}\right) \end{aligned}$$

Inbuilt tool box in Matlab is employed to simulate the neural network and to make out the output; the statistical significance of the output is evaluated. The error minimization criterion in prediction is well achieved to approach a high degree of accuracy. Back-propagation helps to compute the learning rate η and adjusting the weight;

$$\Delta_i w_{pq}(\delta) = -\eta \frac{\partial E^i}{\partial w_{pq}} + \mu \Delta_i w_{pq}(\delta - 1)$$

Table 1 Levels and range of various factors used in Plackett-Burman

Factors	Medium constituents	Levels used (in g/l)	
		(-1)	(+1)
A	Glucose	3.0	7.0
B	Yeast Extract	3.0	7.0
C	NH ₄ Cl	0.7	1.30
D	Phosphate	5.0	11.0
E	NaCl	0.3	0.7
F	MgSO ₄	0.3	0.7
G	FeSO ₄	0.07	0.13

4 Statistical Screening and Optimization Method

4.1 Statistical Screening of Media Components

The cloned strain of *E. coli* BL21 has been taken as a host organism with plasmid vector pRSET-B. Since the induction of recombinant protein expression is achieved by addition of the non-metabolisable lac-inducer IPTG, it is to be added when OD is found to reach just above a level of 0.6. Design Expert software DX8 is used to prepare the framework for media constituent enrichment. Screening of Production Media constituents for extracellular streptokinase was done using Plackett Burman Method (PB). In this method level and range of factors has been suitably selected, Table 1. Using the initially selected range a layout is prepared to take the response for respective run, given by Table 2.

The seven factors taken are Glucose to FeSO₄ as F_1 – F_7 . As F_i – F_n being encoded to show factors from A to G

$$\text{Effect}(F_i) = \sum (x(+)_i - x(-)_i) / (n/2)$$

where, $n \Rightarrow$ is the number of run (experiments).

Values obtained for Effect as $E(F_1)$ – $E(F_7)$ are depicted in Table 3.

Degree of Freedom = Number of dummy variables found are two in number (here, $E(F_5)$ and (X_7))

$$V_{eff} = \sum E_d^2 / n_d = 1.3 \times 10^{-3} \text{ and } S.E. = \sqrt{V_{eff}} = 0.0360$$

Since, $t(N_i) = E(N_i)/S.E.$; therefore, $t(N_1)$ – $t(N_7)$ has been obtained, p value for the respective factors is shown in Table 3. The level of confidence is greater than 95 % in case p -value is lesser than 0.05, according to the central limit theorem. From the standard table of normal distribution statistic where p corresponds to t , the respective p values are obtained. Now the attempt was to optimize the media constituents using Central Composite design methodology but no significant enhancement in terms of product magnitude were achieved so culture condition would be the forth most goal.

Table 2 Plackett Burman experiment design for production media

Run	A: Glucose g/l	B: Yeast g/l	C:NH ₄ Cl g/l	D:Phosphate g/l	E:NaCl g/l	F:MgSO ₄ .7H ₂ O g/l	G:FeSO ₄ g/l	SK production µg/ml
1	3	3	1.3	11	0.3	0.3	0.13	1.624
2	3	3	0.7	11	0.7	0.7	0.07	1.596
3	3	7	0.7	5	0.7	0.3	0.13	1.554
4	7	7	1.3	11	0.7	0.7	0.13	1.918
5	7	3	0.7	5	0.3	0.7	0.13	1.386
6	7	3	1.3	5	0.7	0.3	0.07	1.344
7	7	7	0.7	11	0.3	0.3	0.07	1.484
8	3	7	1.3	5	0.3	0.7	0.07	1.862

SK stands for streptokinase

Table 3 Respective *p*-values for medium components

Factors	Medium components	Effect	<i>T</i>	<i>p</i> -value
A	Glucose	-0.1260	-3.50	0.0228
B	Yeast Extract	0.2170	6.0278	0.0264
C	NH ₄ Cl	0.1820	5.0556	0.0806
D	Phosphate	0.1190	3.3056	0.0370
E	NaCl	0.0140	0.3889	0.7348
F	MgSO ₄	0.1890	5.2500	0.0344
G	FeSO ₄	0.0490	1.3611	0.3066

4.2 Optimization of Culture Conditions for Streptokinase Production

The optimization of culture condition is performed using Central Composite design (CCD) methodology. Firstly range of different set of factors are taken, shown in Table 4, then statistically predicted (using Design expert 8.0) and ANN predicted values are simultaneously compared for streptokinase production magnitude in respect to various factors taken, as shown in layout Table 5.

In Table 5, the predicted values for each factor has been evaluated by applying neural network method also, and worthy results are obtained. The Table 6 shows ANOVA based interaction analysis among statistical model factors and the significance of the model.

Table 4 The range of various factors taken for CCD

Factors	Levels				
	$-\alpha$	-1	0	+1	$+\alpha$
A: pH	6.00	6.50	7.00	7.50	8.00
B: Agitation	100	150	200	250	300
C: Temperature	31.0	34.0	37.0	40.0	43.0
D: Inoculum conc. (%)	1.25	2.00	2.75	3.50	4.25

5 Results and Discussion

5.1 Statistical Bioprocess Modelling

It is clear from the literature that the parameters pertaining to culture conditions are not yet optimized for the strain. Consequently the screening of various production media constituents and optimization of culture condition parameters are to be assumed for maximum streptokinase/plasmin activity. After screening out vital media components for *E. coli* on the basis of statistical significance, CCD is used to find out the most significant interaction of culture conditions in respect to maximum streptokinase activity.

5.2 Culture Conditions Optimization

The important task of culture condition optimization has to be done for maximum productivity using Design Expert 8.0 software. The Neural Network based simulation of the process has been done in Matlab, the predicted values obtained from neural network has been taken into account together with the statistically predicted output. Neural Network tool box commands with self devised algorithm were used to configure neural network application. In optimization process statistically the contour plots are obtained for all the six interactions among which the two significant plots are shown, Figs. 5 and 6, for interaction among the pair of variables to achieve best plasmin activity.

5.2.1 Model Quadratic Equation Obtained

(Abbreviations used here, $P \Rightarrow$ pH, $A \Rightarrow$ Agitation, $T \Rightarrow$ Temperature, $I \Rightarrow$ Inoculum Conc (%))

Table 5 CCD for culture conditions optimization

Run	Factor 1 A: pH	Factor 2 B: Agitation (rpm)	Factor 3 C: Temp (°C)	Factor 4 D: Inoculum conc. (%)	Response 1 SK production (µg/ml)	Stats pred. values SK production (µg/ml)	ANN predicted SK production (µg/ml)
1	7.5	250	40	2	2.128	2.03	2.1663
2	6.5	250	40	3.5	2.240	2.17	2.2401
3	6.5	250	34	3.5	2.072	2.03	2.0720
4	7	200	37	2.75	2.086	1.91	1.9209
5	6.5	150	40	2	1.239	1.48	1.2403
6	7	200	43	2.75	1.435	1.53	1.4046
7	7	200	31	2.75	1.778	2.02	1.7756
8	7.5	150	34	3.5	2.107	2.12	2.1029
9	7.5	250	34	2	2.723	2.44	2.6931
10	7.5	150	40	2	1.351	1.14	1.3497
11	6.5	250	40	2	2.779	2.52	2.7734
12	7	200	37	2.75	1.365	1.91	1.9209
13	6.5	150	34	2	2.142	2.03	2.1417
14	7.5	150	40	3.5	2.002	1.97	2.0065
15	7	200	37	4.25	2.079	2.19	2.0791
16	6.5	150	40	3.5	2.072	2.10	2.0675
17	7.5	250	34	3.5	2.163	1.83	2.1630
18	7	300	37	2.75	2.107	2.57	2.1070
19	6	200	37	2.75	2.303	2.26	2.3030
20	7.5	250	40	3.5	2.023	1.89	2.0201
21	7.5	150	34	2	1.778	1.76	1.7783
22	6.5	150	34	3.5	2.156	2.17	2.1607
23	7	200	37	2.75	2.053	1.91	1.9209
24	7	100	37	2.75	1.946	1.82	1.9461
25	8	200	37	2.75	1.323	1.71	1.3185

(continued)

Table 5 (continued)

Run	Factor 1 A: pH	Factor 2 B: Agitation (rpm)	Factor 3 C: Temp (°C)	Factor 4 D: Inoculum conc. (%)	Response 1 SK production (µg/ml)	Stats pred. values SK production (µg/ml)	ANN predicted SK production (µg/ml)
26	6.5	250	34	2	2.912	2.86	2.9123
27	7	200	37	2.75	1.907	1.91	1.9209
28	7	200	37	2.75	1.953	1.91	1.9209
29	7	200	37	2.75	2.121	1.91	1.9209
30	7	200	37	1.25	1.953	2.18	1.9495

Table 6 Analysis of variance (ANOVA) results

Source	Sum of squares	df	Mean square	F value	p-value prob > F	Test statistic
Model	3.24	14	0.23	2.63	0.0368	Significant
A-pH	0.45	1	0.45	5.15	0.0385	
B-Agitation	0.85	1	0.85	9.65	0.0072	
C-Temp	0.35	1	0.35	4.00	0.0640	
AB	0.022	1	0.022	0.25	0.6233	
AC	6.202E-003	1	6.202E-003	0.070	0.7942	
AD	0.045	1	0.045	0.51	0.4863	
BC	0.042	1	0.042	0.48	0.5006	
BD	0.94	1	0.94	10.64	0.0052	
CD	0.22	1	0.22	2.56	0.1307	
A ²	7.724E-003	1	7.724E-003	0.088	0.7711	
B ²	0.14	1	0.14	1.53	0.2345	
C ²	0.033	1	0.033	0.38	0.5476	
D ²	0.13	1	0.13	1.42	0.2516	
Residual	1.32	15	0.088	-	-	
Lack of fit	0.93	10	0.093	1.17	0.4569	Not significant
Pure error	0.39	5	0.079	-	-	
Cor total	4.56	29	-	-	-	

Fig. 5 A Contour plot showing the interaction and effect of agitation and pH on streptokinase production. In plot, actual constraints viz. temperature 37 °C, inoculum concentration 2.75 % which is shifted

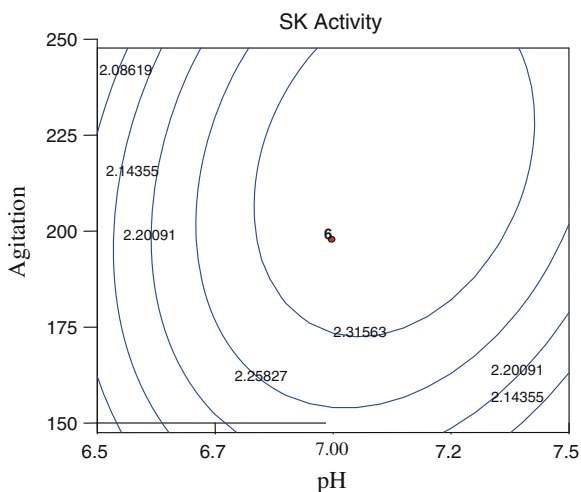
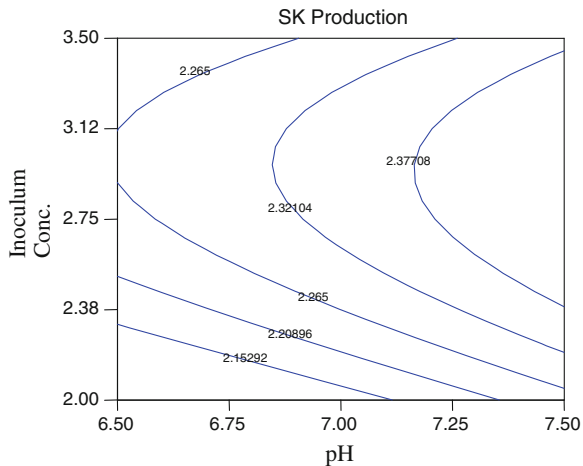


Fig. 6 Contour plot showing the effect of inoculum conc. and pH at temperature 37.86 °C and agitation of 201.43 rpm on streptokinase production



$$\begin{aligned}
 \text{SK Production}(y) = & 6.16592 - 0.81958P + (8.06583E-003A) + 0.12486T - 2.30589I \\
 & - (1.48750E-003PA) - 0.013125PT + 0.14117PI + (3.41250E-004AT) \\
 & - (6.45167E-003AI) + 0.052694TI + 0.067125P^2 + (2.80625E-005A^2) \\
 & - (3.87153E-003T^2) + 0.12006I^2
 \end{aligned}$$

$$\frac{\partial y}{\partial P} = -0.81958 + (2 \times 0.067125P) - (1.48750 \times 10^{-3}A) - 0.013125T + 0.14117I$$

$$\frac{\partial y}{\partial A} = 8.06583 - (1.48750 \times 10^{-3}P) + (2 \times 2.80625 \times 10^{-5}A) + (3.41250T \times 10^{-4}) - 6.45167 \times 10^{-3}I$$

$$\frac{\partial y}{\partial T} = 0.12486 - 0.013125P + (3.41250 \times 10^{-4}A) - 2 \times (-3.87153 \times 10^{-3})T + 0.052694I$$

$$\frac{\partial y}{\partial I} = -2.30589 + 0.14117P - (6.45167 \times 10^{-3}A) + 0.052694T + (2 \times 0.12006I)$$

Now the four simultaneous equations solved for, $\frac{\partial y}{\partial P} = 0$, $\frac{\partial y}{\partial A} = 0$, $\frac{\partial y}{\partial T} = 0$, $\frac{\partial y}{\partial I} = 0$.

Since, $Y = \text{inv}(A) * B$, so using the values of matrix coefficients variables can be obtained as, $P = 8.1351$, $A = 283.0167$, $T = 37.5653$ and $I = 3.5930$. The ANN model is first trained, then testing and validation is done using the part of the same dataset. For our purpose *nn* toolbox is used, available is Matlab. The correlation r^2 in the later case was found to be close to 1.0. The performance efficiency of neural network in respect to error minimization during data training is shown in Fig. 7. Plots shows the Actual versus Predicted obtained from statistical and ANN Based methods. Figure 8, clearly presents the actual and predicted points of the neural network data set, correlation coefficient r^2 for the plot, Fig. 8, is computed to be 0.9291 while it has been estimated as 0.7102 for the statistical model.

Fig. 7 Showing ANN based minimization of error during data training

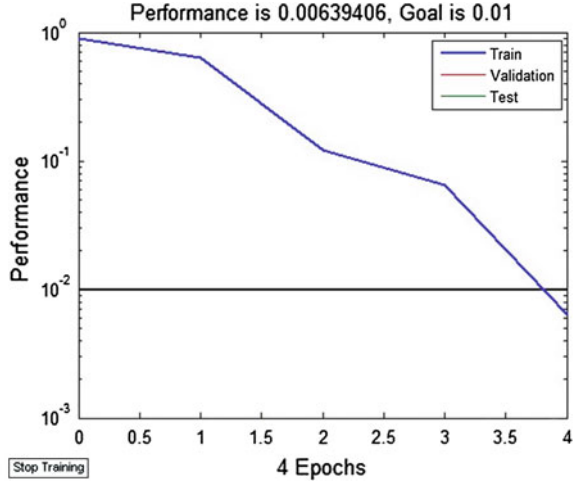
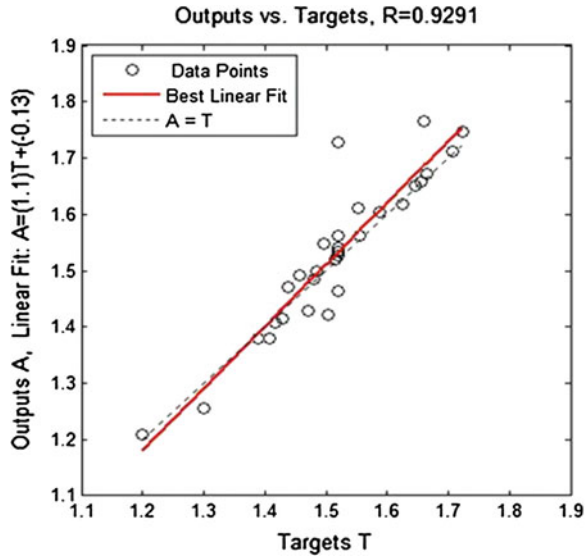


Fig. 8 Plot showing output versus target obtained from neural network model



6 Conclusions

In the dynamical system it is observed that sustenance of plasmid bearing cells occur for a larger time span as moving to a low dilution rate in performing the continuous culture. There is somewhat abrupt increase in the retention of recombinant cells fraction. The average copy number of plasmid is reported to found around 4–6 h of batch culture process. Four production media components

were screened out by Plackett Burman Methodology, these factors viz., glucose (A), yeast extract (B), phosphate (D) and $MgSO_4$ (F) had shown their relevance in recombinant enzyme production having significant role in growth, it would be due to the reason that glucose and yeast extract are prime carbon and nitrogen sources for proper cell growth phosphate and $MgSO_4$ used to support ATP formation, protein synthesis at intracellular level. The next aim had been to optimize the culture-condition governing parameters which had been done subsequently. The validation of the model outcome from CCD was done afterward. Optimization of culture conditions was achieved to a good extent, about 40 % in Central composite Design Technique. Statistically model was found significant and lack of fit was not significant.

So, CCD analysis was performed to estimate the interaction among the relevant factors and evaluate the response that was evident from ANOVA analysis. ANN is simultaneously employed to predict the output and it has shown more accuracy in prediction in comparison to our statistical technique. The prior one having $r^2 = 0.9291$, although 0.7102 obtained statistically. Due to presence of noise very adequate correlation can be sometimes harder to achieve. The culture-condition governing parameters optimization has been subsequently done and success is achieved to a good extent. Efficient modeling may ensure an enhanced production of streptokinase on large scale.

References

1. Thangadurai, C., Suthakaran, P., Barfal, P., Anandaraj, B., Pradhan, S.N., Boneya, H.K., Ramalingam, S., Murugan, V.: Rare codon priority and its position specificity at the 5' of the gene modulates heterologous protein expression in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **376**, 647–652 (2008)
2. Ramalingam, S., Gautam, P., Mukherjee, K.J., Jayaraman, G.: Effects of post-induction feed strategies on secretory production of recombinant streptokinase in *Escherichia coli*. *Biochem. Eng. J.* **33**, 34–41 (2007)
3. Malke, H., Ferretti, J.: Streptokinase: cloning, expression, and excretion by *Escherichia coli*. *Proc. Nat. Acad. Sci.* **81**, 3557–3561 (1984)
4. Simutis, R., Lubbert, A.: Exploratory analysis of bioprocesses using artificial neural network-based methods. *Biotechnol. Prog.* **13**, 479–487 (1997)
5. Lek, S., Guegan, J.F.: Artificial neural networks as a tool in ecological modelling, an introduction. *Ecol. Model.* **120**, 65–73 (1999)
6. Glassey, J., Montague, G.A., Ward, A.C., Kara, B.V.: Enhanced supervision of recombinant *E. coli* fermentations via artificial neural networks. *Process Biochem.* **29**, 387–398 (1994)
7. Bezerra, M.A., Santelli, R.E., Oliveira, E.P., Villar, L.S., Escalera, L.A.: Response surface methodology (RSM) as a tool for optimization in analytical chemistry. *Talanta* **76**, 965–977 (2008)
8. Vellanki, R.N., Potumarthi, R., Mangamoori, L.N.: Constitutive expression and optimization of nutrients for streptokinase production by *Pichia pastoris* using statistical methods. *Appl. Biochem. Biotechnol.* **158**(1), 8315, a–z (2009)
9. Kumar, P., Ghosh, S.: Population dynamics model for plasmid bearing and plasmid lacking cells for streptokinase production in continuous flow stirred tank bioreactor. *Int. J. Eng. Sci. Technol.* **5**(2), 118–127 (2010)

10. Patnaik, P.R.: A heuristic approach to fed-batch optimization of streptokinase production. *Bioprocess. Eng.* **13**, 109–112 (1995)
11. Yazdani, S.S., Mukherjee, K.J.: Continuous-culture studies on the stability and expression of recombinant streptokinase in *Escherichia coli*. *Bioprocess Biosyst. Eng.* **24**, 341–346 (2002)
12. Patnaik, P.R.: Neural optimization of fed-batch streptokinase fermentation in a non-ideal bioreactor. *Can. J. Chem. Eng.* **80**, 920–926 (2002)
13. Ganusov, V.V., Brilkov, A.V.: Estimating the instability parameters of plasmid-bearing cells. 1. Chemostat culture. *J. Theor. Biol.* **219**, 193–205 (2002)
14. Horiuchi, J.-I., Kikuchi, S., Kobayashi, M., Kanno, T., Shimizu, T.: Modeling of pH response in continuous anaerobic acidogenesis by an artificial neural network. *Biochem. Eng. J.* **9**, 199–204 (2001)

Landuse Change Prediction and Its Impact on Surface Run-off Using Fuzzy C-Mean, Markov Chain and Curve Number Methods

Arun Mondal, Deepak Khare, Sananda Kundu, P. K. Mishra
and P. K. Meena

Abstract The landuse change has considerable impact on the surface run-off of a catchment. With the changing landuse there is reduction in the initial abstraction which results in increasing run-off. This also has effect on future because of constant change in landuse due to urbanization. The Soil Conservation Service Curve Number (SCS-CN) model was used in the study for calculating run-off in a sub-catchment of Narmada River basin for the years 1990, 2000 and 2011 which was further validated with the observed data from the gauges. Stream flow of future for 2020 and 2030 was estimated by this method to observe the impact of landuse change on run-off. The landuse classification was done by Fuzzy C-Mean algorithm. The future landuse prediction for 2020 and 2030 was performed with the Markov Chain Model with 2011 validation. Future run-off was generated on the basis of changing landuse which shows increasing rate of run-off.

Keywords Landuse change · Run-off · SCS Curve Number · Fuzzy C-Mean · Prediction · Markov Chain Model

A. Mondal (✉) · D. Khare · S. Kundu · P. K. Meena
Department of Water Resources Development and Management, Indian Institute
of Technology, Roorkee, India
e-mail: arun.iirs@gmail.com

D. Khare
e-mail: kharefwt@gmail.com

S. Kundu
e-mail: sanandakundu@gmail.com

P. K. Meena
e-mail: pramodcae@gmail.com

P. K. Mishra
Water Resources Systems Division, National Institute of Hydrology, Roorkee, India
e-mail: erprabhash@gmail.com

1 Introduction

Assessment of the impact of landuse change on the runoff dynamics of a basin is a motivating field for hydrologists. Different methods have been applied to find out the deficiency of knowledge in this field, but not much is known yet and there is a lack of any general and credible model for predicting the effect of landuse changes [1]. Many results are found, with some even opposing to the results of the others. According to [2], there is no significant rise in the yield of water due to burning of Eucalyptus. While [3] deduced that there is rising water yield because of decrease in forest cover and according to [4] there is change in yield depending on the vegetation cover. Urbanization also leads to increasing run-off and small floods [5]. Various authors who have worked on the impact of landuse change on run-off are [6–9]. Most of the physical models require large amount of data, but [10] said that there are few models which require few data and are not entirely physical models. The prediction of the effect of landuse change also largely depends on the uncertainty of the parameters used in the models [11, 12]. But the uncertainty of such models is more as they have more computation, and are not practical for large catchments.

The Soil Conservation Service Curve Number (SCS-CN) model developed by United State Department of Agriculture (USDA) [13–15] relates the run-off with rainfall parameter and soil cover through Curve Number (CN). This is a very widely used model and gives good result [16–19]. [20] deduced that landuse is an essential parameter for the SCS-CN model and it has been observed that landuse change have significant impact on the run-off and rainfall relation [21] resulted in changed run-off and soil cover [22]. There was a change in the forest catchment due to landuse change [23]. To know the hydrological response in a river basin, it is necessary to evaluate the effect of landuse change and prediction of future landuse change. It is extremely important to know this impact on future run-off.

Change in landuse is particularly important with the growing population which further affects the river basin and run-off. Change detection studies were done by number of researchers [24, 25] who have used satellite images for showing landuse changes. Markov chain model helps in prediction by developing matrix where observed transition probabilities from one state to another is used for future projection [26] and it can be utilized in large spatial scales including different types of landuse [27, 28]. Recent studies with Markov chain model for landuse prediction were performed by [29–31].

In the present study, impact of landuse change in a sub-catchment area of Narmada River basin was done for the years 1990, 2000 and 2011 to estimate the change in run-off condition. With the rising population pressure and demand, there was change in the landuse which is further influencing the total run-off from the catchment and thus affecting the water supply. Again, future landuse prediction was done to estimate the future run-off for the year 2020 and 2030. SCS-CN method was used for the run-off calculation and satellite images were taken for

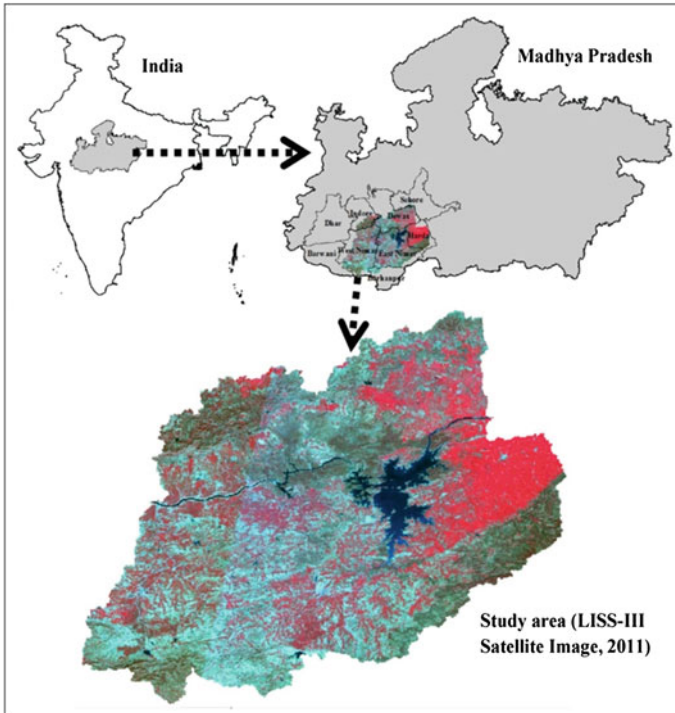


Fig. 1 Study area

landuse classification with Fuzzy C-Mean together with the Markov model to give future landuse projection and its impact on run-off.

The study area is a part of Narmada River basin in Madhya Pradesh extending from $21^{\circ}23'7.7''N$ to $22^{\circ}55'8''N$ and $75^{\circ}21'7''E$ to $77^{\circ}21'17''E$ covering about $20,561 \text{ km}^2$. It includes the districts of Burhanpur, Khandwa, Sehore, Barwani, Harda, Dewas, Indore, West Nimar and East Nimar. The outlet of the area is Mandaleswar gauge. The region has a subtropical climate which is a hot dry summer (April–June) through monsoon rains (July–September) to a cool and dry winter. The rainfall average is approximately 1,370 mm and it decreases from east to west (Fig. 1).

2 Data and Methods

The hydrological data for the gauging stations were used together with the daily station rainfall data for the study area from IMD. The satellite data for 1990, 2000 and 2011 were used for landuse classification from the Landsat TM and LISS-III data.

2.1 SCS-CN Method for Run-off Estimation

Run-off calculation was done with the SCS-CN [32] method at the gauge of Handia and Mandaleswar using surface run-off from Handia to Mandaleswar, total catchment area and base flow. The area was then classified into 5 landuse classes and prediction was done to observe the impact on future run-off. The overall CN (Curve Number) was taken for different landuse classes which helped in determining the run-off for the years 1990, 2000 and 2011.

2.2 Changed Run-off due to Landuse Change

The landuse data and rainfall data were used to calculate the run-off for each year. The calculated data was then validated with the observed data at the gauge points. To get the impact of landuse on run-off, rainfall of 1990 was taken as constant rainfall for 1990, 2000 and 2011, and the change in run-off was observed with respect to the corresponding landuse of these three years. For example, rainfall of 1990 and landuse of 1990 were considered to calculate run-off of 1990, rainfall of 1990 and landuse of 2000 to calculate the run-off of 2000 and so on. The future prediction of 2020 and 2030 run-off was done with the rainfall of 1990 and predicted landuse of these two years with the Markov model.

2.3 Landuse Classification with Fuzzy C-Mean

Landsat TM data sets were taken from USGS (website <http://glovis.usgs.gov>). Data were projected in UTM projection zone 44 and WGS 84 datum. The 1990 image was considered as the base data and 2000 and 2011 images were co-registered by first order polynomial model with the Root Mean Square Error (RMSE) of 0.5 pixel. The images were then geometrically corrected and radiometric normalization was done. All the images were classified using Fuzzy classification into 5 landuse classes. The accuracy assessment was done with all the images after classification.

[33] have described different types of existing clustering techniques. Supervised Fuzzy C-Mean is a classification technique where pixels are categorized after specifying the training sample areas. The fuzzy concept of classification represents a natural model where each pixel may have partial membership corresponding to different landuses. It assigns membership values for each sample of each class that varies from 0 to 1. In the present study following Fuzzy C-mean (FCM) algorithm was used for clustering technique:

$$J_m(U, v) = \sum_{i=1}^C \sum_{k=1}^N u_{ik}^m \|y_k - v_i\|_A^2 \quad (1)$$

where

- Y is $\{Y_1, Y_2 \dots Y_N\} \subset R^n$ = the data,
- c is number of clusters in Y; $2 \leq c \leq n$,
- m is weighting exponent; $1 \leq m < \infty$.
- U is fuzzy c-partition of Y; $U \in M_{fc}$
- v is $(v_1, v_2 \dots v_c)$ = vectors of centers,
- vi is $(v_{i1}, v_{i2}, \dots, v_{in})$ = center of cluster i,
- $\| \cdot \|_A$ is induced A-norm on R^n , and
- A is positive-definite $(n \times n)$ weight matrix.

The membership value in each class of a sample depends on its distance to the corresponding cluster centre. The weight factor m decreases the impact of small membership values. Greater the value of m, smaller will be the influence of samples with small membership values [34].

After classification, accuracy assessment was done with the Kappa statistics [35] on the images of 1990, 2000 and 2011. Markov model was then used to predict future changes in landuse of 2020 and 2030 with validation of 2011 landuse on the basis of 1990 and 2000 landuse.

2.4 Markov Chain Model

Markov model was used here to project future landuse changes and estimation of future run-off from the changed landuse. The Markov model is given as [29]

$$\begin{aligned}
 P\{X_t = a_j | X_0 = a_0, X_1 = a_1, \dots, X_{t-1} = a_i\} \\
 = P\{X_t = a_j | X_{t-1} = a_i\}
 \end{aligned}
 \tag{2}$$

Moreover, it is proper to regard the change process as one which is uniquely distinctive with time $(t = 0, 1, 2 \dots)$.

$P\{X_t = a_j | X_{t-1} = a_i\}$ gives the probability where the process develops transition from state a_i to state a_j within one time period. When the ℓ steps are required for implementing the transition, then the $P\{X_t = a_j | X_{t-1} = a_i\}$ is called the ℓ step transition probability $P_{ij}^{(\ell)}$.

In case $P_{ij}^{(\ell)}$ is independent of time and dependent on states a_i, a_j and ℓ , the Markov chain is then called homogeneous;

$$P\{X_t = a_j | X_{t-1} = a_i\} = P_{ij}
 \tag{3}$$

where P_{ij} value is calculated from the observed data by arranging the number of times the observed data go from state i to j, n_{ij} , and total occurrences of the state a_i , n_i is summed up.

$$P_{ij} = n_{ij} / n_i
 \tag{4}$$

As the Markov chain continues with time, the probability of retaining in state j gets independent of the initial state of the chain after many steps. When this condition occurs, the chain is considered to have reached a steady state. Then P_j , which is the limit probability, is used to determine the value of $P_{ij}^{(\ell)}$,

$$\lim_n P_{ij}^{(n)} = P_j \quad (5)$$

Where

$$P_j = P_i P_{ij}^{(n)} \quad j = 1, 2, \dots, m \text{ (state)}$$

$$P_i = 1 \quad P_j > 0$$

3 Results and Analysis

3.1 Landuse Classification and Prediction with Markov Model

Kappa statistic was calculated for accuracy assessment given in Table 1.

Overall kappa statistics and overall accuracy for 2011 were 0.78 and 84.17 % respectively (Table 2). For the assessment of landuse change in different decades, Landsat TM and LISS-III images were taken for the years 1990, 2000 and 2011. Changes in area that have occurred in different years are provided in Table 3. Large increase in the area of water and built-up is observable from 1990 to 2011. Water area has increased from 1.01 to 4.56 % in 2011 while built-up increased from 3.61 to 10.53 % in 2011. Agricultural land area also increased from 34.87 to 40.26 % in 2011 (Table 3).

The landuse maps of 1990, 2000, 2011, 2020 and 2030 are given in Fig. 2. In 2011, water bodies increases to a great extent due to construction of Indira Sagar Dam after 2000.

The predicted results of 2020 and 2030 (Table 4) indicate rise in the water body and built-up area in next 19 years after 2011 while vegetation, agricultural land and fallow land decreases. Increase in settlement has resulted in the decrease of vegetation and fallow at the cost of built-up. Water area is increasing giving the effect of dam construction as predicted by the model.

3.2 Change in Run-off Values for Changed Landuse

The run-off calculation was done with the rainfall data by SCS-CN method for 1990, 2000 and 2011. The results of 1990 and 2000 were validated with the

Table 1 Kappa statistics

Class name	Kappa statistics (1990)	Kappa statistics (2000)	Kappa statistics (2011)
Water body	0.79	0.82	0.82
Built-up	0.87	0.80	0.83
Vegetation	0.63	0.78	0.76
Agricultural land	0.81	0.81	0.76
Fallow land	0.76	0.75	0.80

Table 2 Overall kappa statistics and overall accuracy

Category	1990	2000	2011
Over all kappa statistics	0.75	0.79	0.78
Over all accuracy (%)	81.67	84.17	84.17

Table 3 Distribution of area of different landuse

Classes	1990 (km ²)	1990 (%)	2000 (km ²)	2000 (%)	2011 (km ²)	2011 (%)
Water body	207.13	1.01	263.38	1.28	937.35	4.56
Built-up	743.25	3.61	1,307.2	6.36	2,164.64	10.53
Vegetation	6,234.55	30.33	5,879.8	28.6	4,695.65	22.84
Agricultural land	7,167.72	34.87	7,361.18	35.81	8,275.70	40.26
Fallow Land	6,205.35	30.18	5,746.44	27.95	4,484.66	21.81
Total	20,558	100	20,558	100	20,558	100

observed river flow. But the river flow became controlled after 2004 due to construction of dams, so 2011 result was not validated with the observed flow. The rainfall of three rain gauge stations was taken from the study area and computation of surface run-off was done. Base flow was used for evaluation of river run-off on daily basis. Comparison between calculated and observed run-off for the year 1990 and 2000 is shown in Fig. 3. The correlation was done with the results which show good relation between the observed and calculated flow.

3.3 Change in Run-off Values for Changed Landuse

The results of the impact of landuse change on run-off show that annual total run-off of the year 1990 was 183.14 mm while total run-off of 2000 was 190.13 mm. Thus the increase is 6.99 mm. From 2000 to 2011 again, the run-off increased up to 215.80 mm with a rise of 25.67 mm. In future prediction of 2020, the annual run-off is estimated as 226.12 mm which again increases in 2030 up to 235.52 mm. Thus the total increase of run-off due to change in landuse from 1990 to 2030 is 52.38 mm if the rainfall for all these years remains same. Thus the result

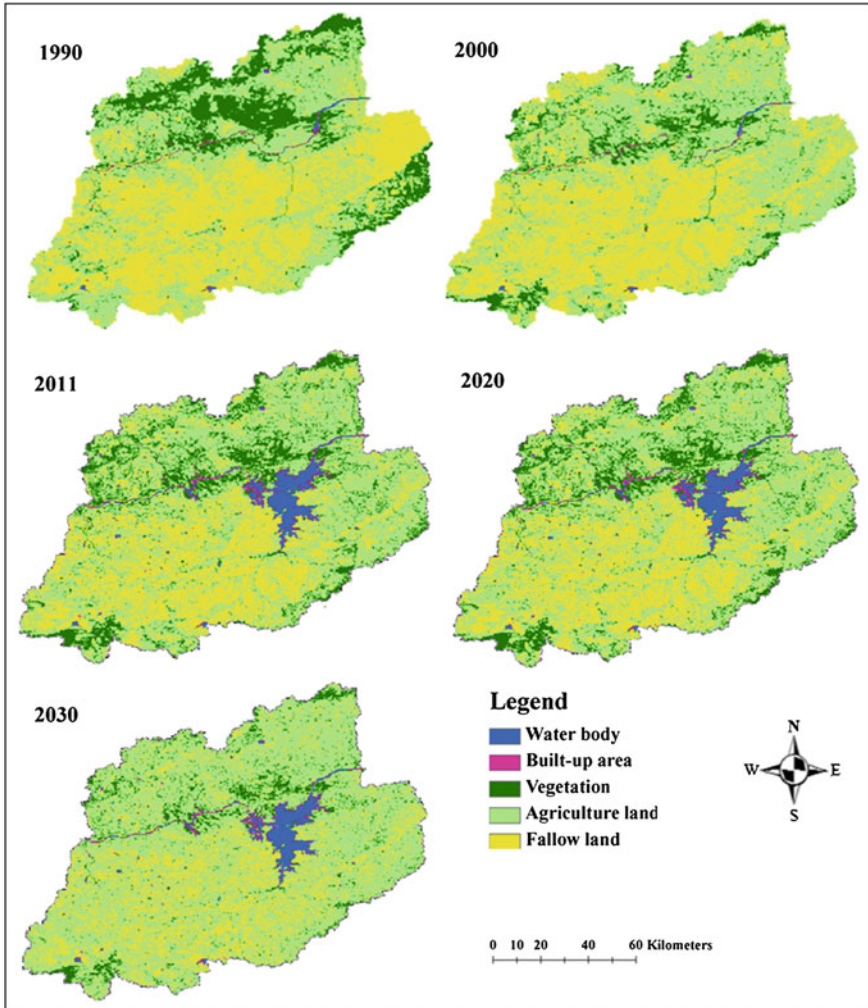


Fig. 2 Landuse map of 1990, 2000 and 2011

Table 4 Future prediction of landuse from Markov Chain for 2020 and 2030

Classes	2011 (km ²)	2011 (%)	2020 (km ²)	2020 (%)	2030 (km ²)	2030 (%)
Water body	937.35	4.56	967.8	4.71	1,009.68	4.91
Built-up	2,164.64	10.53	2,709.8	13.18	3,268.16	15.90
Vegetation	4,695.65	22.84	4,027.05	19.59	3,505.82	17.05
Agricultural land	8,275.70	40.26	9,113.09	44.33	9,445.07	45.94
Fallow Land	4,484.66	21.81	3,740.26	18.19	3,329.27	16.19
Total	20,558	100	20,558	100	20,558	100

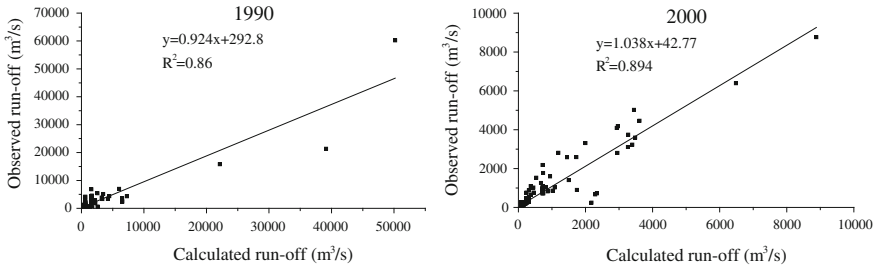


Fig. 3 Correlation of observed and simulated run-off for 1990 and 2000

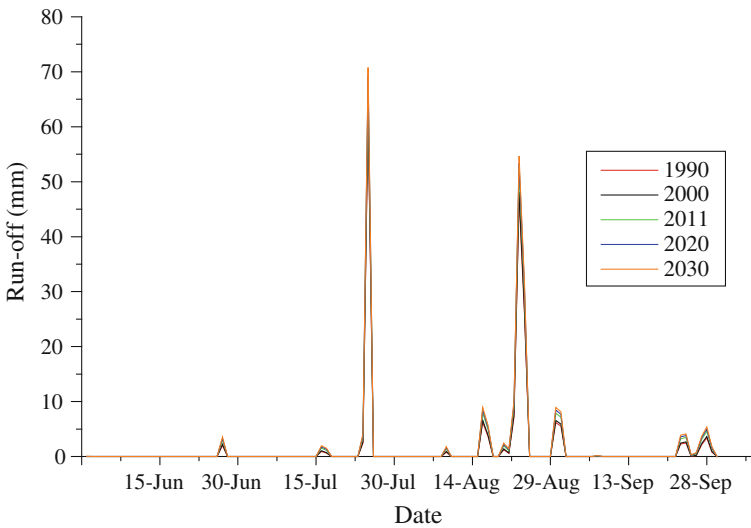


Fig. 4 Impact of landuse change on run-off

provides an estimation of future landuse and future run-off of this sub-catchment which will be useful in further studies (Fig. 4).

3.4 Future Run-off of 2020 and 2030 with Predicted Landuse

The landuse of 2011 was predicted by the landuse of 1990 and 2000 and also classified with the Fuzzy C-Mean algorithm to validate both the results. The landuse of 2020 and 2030 were predicted with the model and the results were further used to generate future run-off with the SCS-CN method. To get the change in run-off with respect to the landuse change, rainfall of 1990 was taken as a constant for all the years and run-off of 1990, 2000 and 2011 were calculated with

the corresponding landuse of these years. Similarly, for 2020 and 2030, the rainfall of 1990 was taken with the predicted landuse. Thus the impact of landuse change on run-off was estimated and analysed from 1990 to 2011 for 21 years and then for next 19 years, keeping rainfall as constant (Fig. 4).

4 Conclusion

There is a constant process of increasing population resulting in rising demand and drastic changes in landuse. All these are affecting the natural resources with changing environment. Landuse change is affecting the surface run-off. This study is a combined use of hydrological methods, modeling and remote sensing together to achieve the goal. The landuse classification of 1990, 2000 and 2011 was done with the Fuzzy C-Mean algorithm which is a soft classification technique and accuracy assessment was performed with the results. The future prediction of 2020 and 2030 landuse was carried out with the Markov model and the validation was done with the 2011 landuse. There was drastic change in the landuse with increasing use of agricultural land and increasing settlement. Forest covers declines from 1990 to 2030 and this change in landuse is observable on the run-off of the study area, even in future. The SCS-CN method was applied for assessing the run-off of the region and the computed result was validated with the observed run-off of 1990, 2000 and 2011. There is a marked increase in run-off from 1990 (67.59 mm) to 2011 (84.85 mm) to 2030 (95.80 mm). These results show that with the increasing settlement and agriculture, and decreasing vegetation, there is an increasing run-off and falling rate of infiltration in near future. On the basis of these results, assessment of future landuse and change in run-off is possible for planning and management of water resource of the area.

Acknowledgments The authors acknowledge the United States Geographical Survey (USGS) and NRSC for providing the Landsat and LISS-III satellite imageries for the study area. The authors are also thankful to the Indian Meteorological Department for providing rainfall and run-off data and to the CSIR for financial support.

References

1. Kokkonen, T.S., Jakeman, A.J.: Structural effects of landscape and landuse on streamflow response. In: Beck, M.B. (ed.) *Environmental foresight and models: a manifesto*, pp. 303–321. Elsevier, Oxford (2002)
2. Langford, K.J.: Change in yield of water following a bush fire in a forest of *Eucalyptus regnans*. *J. Hydrol.* **29**, 87–114 (1976)
3. Hibbert, A.R.: Forest treatment effects on water yield. In: Sopper, W.E., Lull, H.W. (eds.) *International symposium on forest hydrology*, p. 813. Pergamon, Oxford (1967)
4. Bosch, J.M., Hewlett, J.D.: A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *J. Hydrol.* **55**, 3–23 (1982)

5. Hollis, G.E.: The effect of urbanization on flood of different recurrence interval. *Water Resour. Res.* **11**, 431–435 (1975)
6. Fohrer, N., Haverkamp, S., Eckhardt, K., Frede, H.G.: Hydrologic response to landuse changes on the catchment scale. *Phys. Chem. Earth (B)* **26**, 577–582 (2001)
7. De Roo, A., Odijk, M., Schmuck, G., Koster, E., Lucieer, A.: Assessing the effects of landuse changes in the Meuse and Oder catchment. *Phys. Chem. Earth (B)* **26**, 593–599 (2001)
8. Braud, I., Vich, A.I.J., Zuluaga, J., Fornero, L., Pedrani, A.: Vegetation influence on runoff and sediment yield in the Andes region: observation and modelling. *J. Hydrol.* **254**, 124–144 (2001)
9. Schreider, S.Y., Jakeman, A.J., Letcher, R.A., Nathan, R.J., Neal, B.P., Beavis, S.G.: Detecting changes in streamflow response to changes in non-climatic catchment conditions: farm dam development in the Murray-Darling basin. *Aust. J. Hydrol.* **262**, 84–98 (2002)
10. Beven, K.J.: *Rainfall-runoff modelling: the primer*. Wiley, New York (2000)
11. Bronstert, A., Niehoff, D., Burger, G.: Effects of climate and land-use change on storm runoff generation: present knowledge and modelling capabilities. *Hydrol. Process.* **16**, 509–529 (2002)
12. Ranzi, R., Bochicchio, M., Bacchi, B.: Effects on floods of recent afforestation and urbanisation in the Mella River (Italian Alps). *Hydrol. Earth Syst. Sci.* **6**, 239–253 (2002)
13. USDA Soil Conservation Service.: *National Engineering Handbook Section 4 Hydrology. Chapters 4–10* (1972)
14. USDA Soil Conservation Service.: *National Engineering Handbook Section 4 Hydrology. Chapter 19* (1983)
15. USDA-SCS.: *National Engineering Handbook, Section 4 Hydrology*. Washington, D.C. (1985)
16. Stuebe, M.M., Johnston, D.M.: Run-off volume estimation using GIS techniques. *J. Am. Water Resour. Assoc.* **26**, 611–620 (1990)
17. Mishra, S.K., Singh, V.P.: Another look at the SCS-CN method. *J. Hydrol. Eng. ASCE* **4**, 257–264 (1999)
18. Mishra, S.K., Singh, V.P.: SCS-CN based hydrological simulation package. In: Singh V.P., Frevert D.K. (eds) *mathematical Models in Small Watershed Hydrology and applications*, pp. 391–464. Water Resources Publications, Littleton (2002)
19. Schneider, L.E., McCuen, R.H.: Statistical guidelines for curve number generation. *J. Irrig. Drainage Eng. ASCE* **131**, 282–290 (2005)
20. Pandey, A., Sahu, A.K.: Generation of curve number using remote sensing and Geographic Information System. (2002). (<http://www.gisdevelopment.net/application/nrm/water/watershed/watws0015pf.htm>)
21. Yang, T.C., Yu, P.S.: The effect of landuse change on the design hydrograph. *J. Hydrol. Changing Environ.* **3**, 207–216 (1998)
22. Kim, Y., Engel, B.A., Lim, K.J.: Runoff impacts of landuse change in Indian river Lagoon watershed. *J. Hydrol. Eng.* **7**, 245–251 (2002)
23. Okoński, B.: Hydrological response to landuse change in central European lowland forest catchment. *J. Environ. Eng. Landscape Manage.* **15**, 3–13 (2007)
24. He, C.Y., Shi, P.J., Chen, J., Zhou, Y.Y.: A study on landuse/cover change in Beijing area. *Geogr. Res.* **20**(6), 679–687 (2000)
25. Zhang, Q., Wang, J., Peng, X., Gong, P., Shi, P.: Urban built-up land change detection with road density and spectral information from multi-temporal Landsat TM data. *Int. J. Remote Sens.* **23**(15), 3057–3078 (2002)
26. Brown, D.G., Pijanowski, B.C., Duh, J.D.: Modeling the relationships between landuse and land cover on private lands in the Upper Midwest. *USA J. Environ. Manage.* **59**, 247–263 (2000)
27. Jahan, S.: The determination of stability and similarity of Markovian landuse change processes: a theoretical and empirical analysis. *Socio-Econ Planning Sci.* **20**, 243–251 (1986)
28. Muller, R.M., Middleton, J.: A Markov model of land-use change dynamics in the Niagara region, Ontario, Canada. *Landscape Ecol.* **9**, 151–157 (1994)

29. Weng, Q.: Landuse change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and stochastic modeling. *J. Environ. Manage.* **64**, 273–284 (2002)
30. Peterson, L.K., Bergen, K.M., Brown, D.G., Vashchuk, L., Blam, Y.: Forested land-cover patterns and trends over changing forest management eras in the Siberian Baikal region. *For. Ecol. Manage.* **257**, 911–922 (2009)
31. Zhang, R., Tang, C., Ma, S., Yuan, H., Gao, L., Fan, W.: Using Markov chains to analyze changes in wetland trends in arid Yinchuan Plain. *Chin Math. Comput. Model.* **54**, 924–930 (2011)
32. Chow, V.T., Maidment, D.R., Mays, L.W.: *Applied Hydrology*. McGraw-Hill, Singapore (1988)
33. Jain, A., Dubes, R.: *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs (1988)
34. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**(2–3), 191–203 (1984)
35. Lillesand, T.M., Kiefer, R.W., Chipman, J.W.: *Remote sensing and image interpretation*. Wiley, New York (2008)

A LabVIEW Based Data Acquisition System for Electrical Impedance Tomography (EIT)

Tushar Kanti Bera and J. Nagaraju

Abstract A LabVIEW based data acquisition system (LV-DAS) is developed for Electrical Impedance Tomography (EIT) for automatic current injection and boundary data collection. The developed LV-DAS consists of a NIUSB-6251 DAQ card, NISCB-68 connector module and an automatic electrode switching module (A-ESM). A LabVIEW based graphical user interface (LV-GUI) is developed to control the current injection and data acquisition by LV-DAS through A-ESM. Boundary data are collected for a number of practical phantoms and the boundary data profiles are studied to assess the LV-DAS. Results show that the high resolution NIDAQ card of the DAS improves its data acquisition performance with accurate measurement and high signal to noise ratio (SNR).

Keywords Electrical impedance tomography (EIT) · Data acquisition system · LabVIEW · Graphical user interface · Practical phantom · Boundary data profile

1 Introduction

Electrical impedance tomography [1–7] is a computed tomographic technique [8–10] which reconstructs the conductivity distribution of a domain under test (DUT) from a set of boundary voltage data developed by injecting a constant current signal through an array of surface electrodes [1, 11] attached to the domain boundary. EIT is a non linear inverse problem [12–14] which reconstructs the spatial distribution of the conductivity of the DUT from the voltage current data using the image reconstruction algorithm [12–14] which a computer program and developed with forward solver [12–14] and inverse solver [12–14]. An EIT system

T. K. Bera (✉) · J. Nagaraju
Department of Instrumentation and Applied Physics, Indian Institute of Science,
Bangalore, India
e-mail: tkbera77@gmail.com

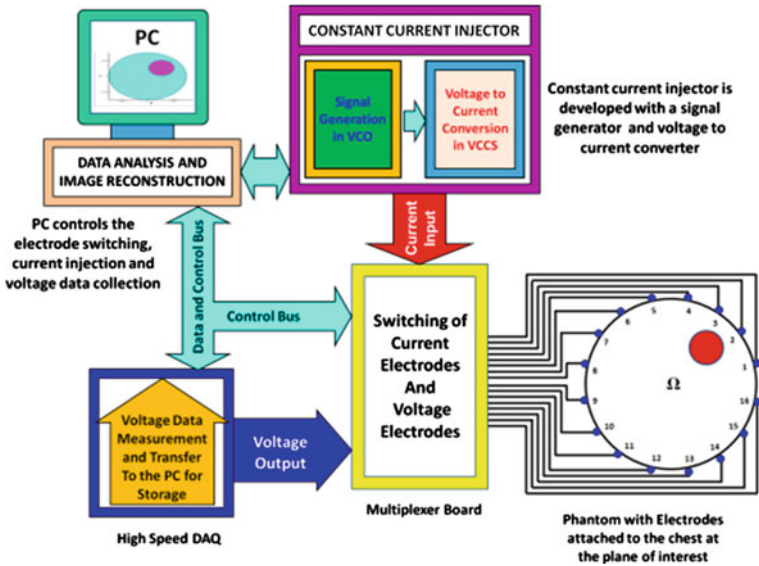


Fig. 1 Schematic of the electrical impedance spectroscopy (EIS) studies on biological materials with four-probe technique using an impedance analyzer [current signal is injected through the outer (shown in red) electrodes and the voltage is measured across the inner (shown in blue) electrodes]

injects a constant current signal to the DUT such as a practical phantom [15–28] or human subject through electrodes or EIT sensors [1, 11, 25] using an EIT instrumentation [26–33] consisting of a current injecting device such as constant current source [34–39] and a voltage measuring device which may be a volt meter or other electronic voltage data measuring circuit (Fig. 1). For better current injection and voltage measurement procedures, a modern EIT instrumentation needs a high precision current injection and high quality high speed DAS for automatic current injection and voltage measurement.

National Instruments Inc. Hardware and LabVIEW software [40–43] can be suitably used for high speed high precision PC based data acquisition in several practical applications [44–55]. A LabVIEW based boundary data acquisition system (LV-DAS) is developed for Electrical Impedance Tomography (EIT). The LV-DAS is developed with NIUSB-6251 DAQ card, NISCB-68 Connector Module and a PC based electrode switching module (A-ESM). A LabVIEW based graphical user interface (LV-GUI) is developed to control the current injection by constant current injector (CCI) and data acquisition by LV-DAS. Practical phantoms are developed and the boundary data are collected from practical phantoms with different current injection patterns [56–58] and the impedance images are reconstructed in EIDORS [59, 60].

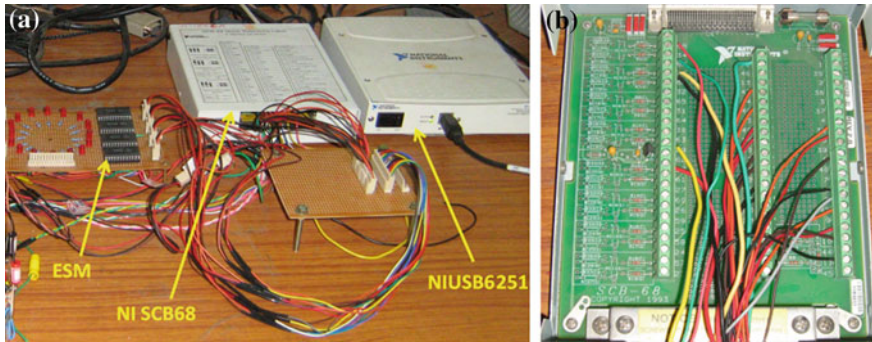


Fig. 2 **a** Schematic of the interface of the USB6251 and the A-ESM through NISCB68 connector module. **b** Wire connection in NISCB68 connector module for the surface electrodes of the phantoms and the USB6251 through A-ESM

2 Materials and Methods

2.1 LabVIEW Based Data Acquisition System (LV-DAS)

The LabVIEW based data acquisition system (LV-DAS) is developed with a NI USB-6251 DAQ card [61, 62], NISCB-68 Connector Module [63] and a PC based electrode switching module (A-ESM) interface with the PC through LV-GUI (Fig. 2a). LV-DAS) is developed for acquiring the voltage data at the phantom boundary using the CCI and A-ESM. The automatic data acquisition is controlled and executed by a LabVIEW based graphical user interface (LV-GUI) which controls the current injection by CCI and data acquisition by LV-DAS through the A-ESM. The LV-DAS is interfaced with PC through the USB (Universal Serial Bus) port of the NI USB-6251 DAQ card. The LV-DAS is also interfaced with phantom or subject with surface electrode array through the A-ESM.

The LV-DAS is developed to perform two works in EIT: digital data generation for operating the analog multiplexers of the A-ESM and analog voltage signal measurement. The developed LV-DAS sequentially generates a 16-bit digital data required for the operating the analog multiplexers of the A-ESM to switch the current and voltage electrodes attached on phantom or subject under test. For a particular current pattern the LV-DAS collects the voltage data from the electrodes connected to the voltage measuring channels (called voltage electrodes) of LV-DAS. The current injection through the CCI, Electrode switching by A-ESM and voltage measurements by LV-DAS all are controlled by LV-GUI through the LabVIEW based data acquisition program (LV-DAP). A-ESM is interfaced with the DAQ card through NI SCB68 [63] as shown in (Fig. 2b) and the LV-DAS generates the 16-bit data to feed it to A-ESM to operate under the control of LV-GUI.

2.2 NI SCB 68 Connector Module

NI SCB68 (Fig. 2b) is a shielded I/O connector block built with 68 screw terminals for easy electronic connection to a NI 68-pin or 100-pin DAQ device for the signal transfer [63]. The SCB-68 is built with a general breadboard area for custom circuitry and sockets for interchanging electrical components. The open component pads of the SCB-68 is permits to add signal conditioning to the analog input (AI), analog output (AO), and PFI0 signals of a 68-pin or 100-pin DAQ.

2.3 Automatic Electrode Switching Module (A-ESM)

Surface electrodes are required to be switched in a particular fashion EIT system for current injection and voltage measurements. In modern EIT systems, generally, high speed analog multiplexers or electronic relays are used to switch the surface electrodes to obtain the current injection with different current patterns. Digital data bits are required to feed the multiplexers in an electrode switching circuit. In the present study, a four multiplexer based A-ESM (Fig. 3) is developed to switch the sixteen electrode of a 16-electrode EIT system. Using the developed A-ESM, the current electrodes (E_i) are switched by the current multiplexer (MUX-I₁ and MUX-I₂) and the voltage electrodes (E_v) are switched by the other multiplexer called the voltage multiplexer (MUX-V₁ and MUX-V₂) as shown in Fig. 3. The A-ESM is developed with four CD4067BE ICs [64] to switch the sixteen surface electrodes ($E_1, E_2, E_3, E_4, E_5, E_6, E_7, E_8, E_9, E_{10}, E_{11}, E_{12}, E_{13}, E_{14}, E_{15}$ and E_{16}) using four probe method with different current injection method as shown in the Fig. 3. The developed A-ESM is controlled by the 16-bit parallel digital data ($D_{15}D_{14}D_{13}D_{12}D_{11}D_{10}D_9D_8D_7D_6D_5D_4D_3D_2D_1D_0$) which are divided in two sets: one for operating the current electrode switching MUXs (MUX-I₁ and MUX-I₂) and other for operating voltage electrode switching MUXs (MUX-V₁ and MUX-V₂).

2.4 LabVIEW Based Graphical User Interface (LV-GUI)

A LabVIEW based graphical user interface (LV-GUI) is developed which controls the current injection by CCI and data acquisition by LV-DAS through the A-ESM. LV-GUI controls the digital data generation, and hence controls the electrode switching for current injection and voltage measurement. Modifying the LabVIEW program, data acquisition with the different current injection patterns are performed for practical phantoms.

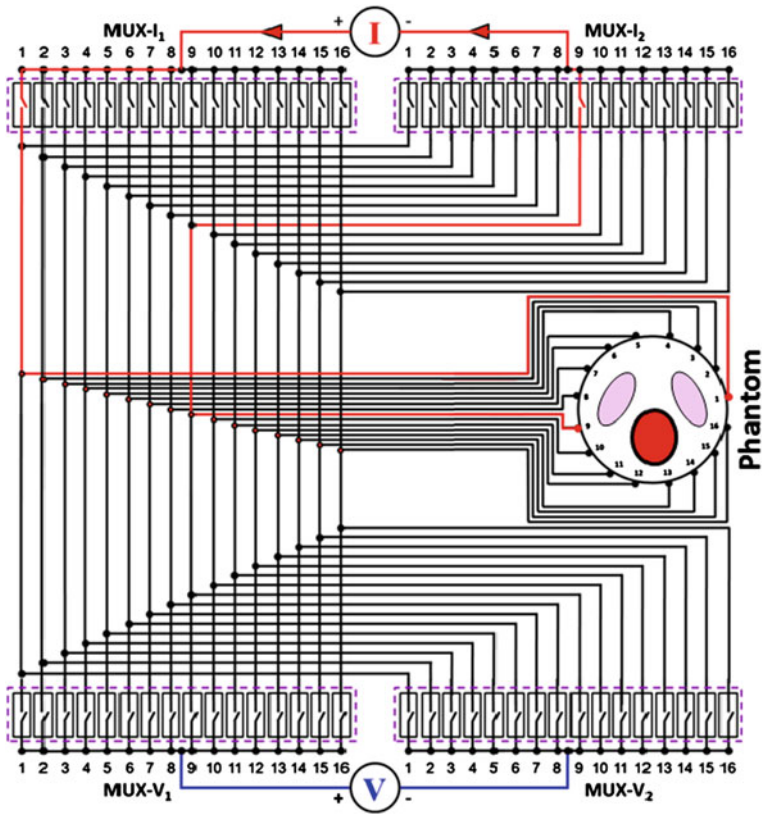


Fig. 3 Schematic of the electrode switching of a sixteen electrode EIT system with an automatic electrode switching module (A-ESM) in opposite current pattern [current (I) is injected through the current electrode switching MUX (MUX-I₁ and MUX-I₂) and voltage (V) is measured across the voltage electrode switching MUX (MUX-V₁ and MUX-V₂)]

2.5 Digital Data Testing Module (DDTM)

The parallel digital bits generated by LV-DAS and tested with digital data testing module (DDTM) [65]. The DDTM is developed with sixteen low power LEDs and sixteen high precision resistors (each resistor (1 kΩ) is connected in series with corresponding LED). After testing of the generation and sequencing of the digital bits they are fed to the electrode switching module (A-ESM) connected to the surface electrodes and the four multiplexers operations are tested simultaneously. To test the MUX switching, the sixteen input pins are connected to the inputs of the DDTM and the single output pin is connected to a battery or a power supply keeping the negative terminal connected to the analog ground.

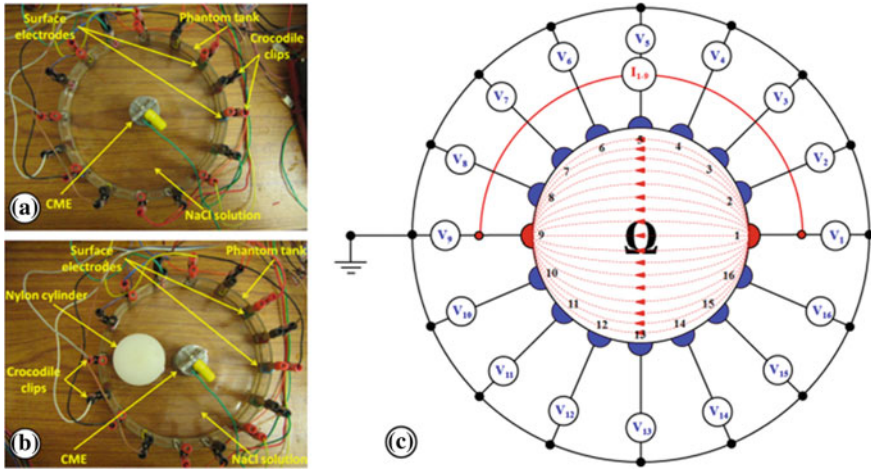


Fig. 4 Practical phantom and boundary data collection process. **a** A NaCl-phantom without inhomogeneity. **b** NaCl-insulator phantom with a nylon cylinder as inhomogeneity near electrode no. 9 and a CME at the phantom center. **c** Schematic of the current injection and boundary data collection in first current projection with opposite current pattern

2.6 Practical Phantom Development

Sixteen electrode NaCl-nylon phantoms are developed (Fig. 4) with a shallow glass tank (150 mm diameter) filled with NaCl solution. Sixteen rectangular stainless steel strips (height 34 mm, width 10 mm) are cut from a type 304 stainless steel sheet and the rectangular SS strips are fixed on the inner wall of the glass tank using insulated steel paper clips. Steel strips are connected with highly conductive copper wires using steel crocodile clips (Fig. 4a) and the wires are connected to the A-ESM. The Tank is filled with 0.9 % (w/v) NaCl solution and a stainless steel cylinder (25 mm diameter) is put at the phantom center (Fig. 4a) to act as a common mode electrode (CME) [15, 23, 26–28, 58] for reducing the common mode error [67, 68] of the electronic circuits. Nylon cylinders (40 mm diameter) are placed at different electrode positions (Fig. 4b) and a number of phantom configurations are obtained. NaCl solution height is maintained as 10 mm so that the active electrode area becomes 10×10 mm. Figure 4a, b shows the phantom without inhomogeneity and phantom with inhomogeneity respectively.

2.7 Current Injection and Boundary Data Collection

In the present study, 1 mA, 50 kHz constant sinusoidal current is injected through current electrodes and the boundary potentials are collected from the other electrodes called voltage electrodes using opposite and neighbouring current injection

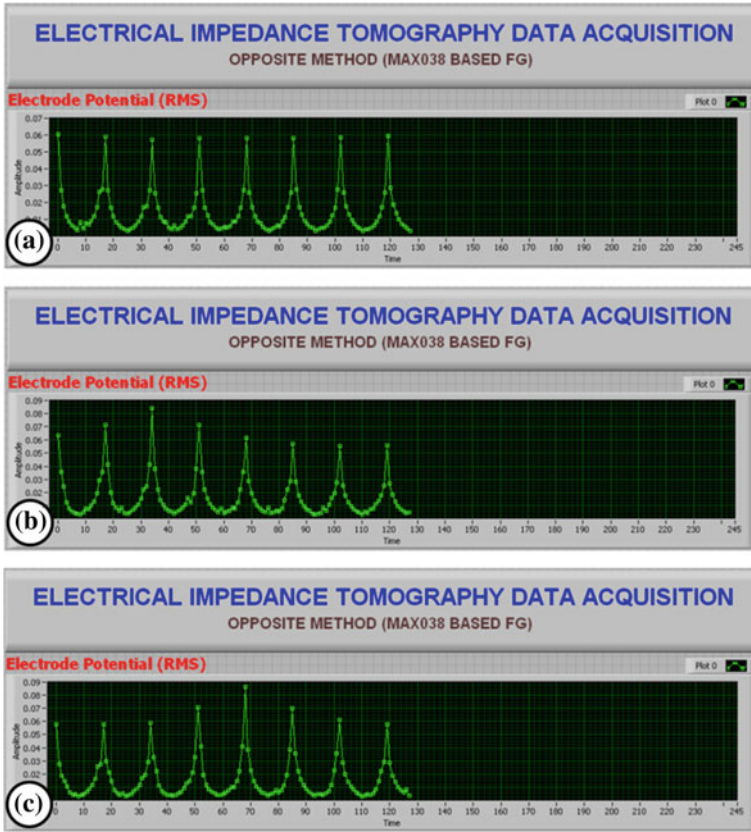


Fig. 5 Boundary data profile of a NaCl phantom acquired by LV-DAS with 1 mA, 50 kHz current in opposite current pattern. **a** Without inhomogeneity. **b** Nylon cylinder (40 mm diameter) near E_3 . **c** Nylon cylinder (40 mm diameter) near E_5

protocol [15, 57–59]. Figure 5 shows the current injection and boundary data collection in the first current projection of opposite current injection method for a sixteen electrode EIT system. For each current projections ($P_1, P_2, P_3, P_4, P_5, P_6, P_7, P_8, P_9, P_{10}, P_{11}, P_{12}, P_{13}, P_{14}, P_{15}$ and P_{16}), the boundary data are collected from all the electrodes to obtain the greatest sensitivity to the resistivity changes in the domain [66]. In the present study with the sixteen electrode EIT system, each of the sixteen current projections (Fig. 4c) yields sixteen boundary potential data ($V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}, V_{11}, V_{12}, V_{13}, V_{14}, V_{15}$ and V_{16}) and hence the complete scan yields 256 boundary potential data (Fig. 4c). Boundary potentials are studied for different phantom configuration and the LV-DAS performance is assessed. Inhomogeneities are placed at different electrode positions and the potential profiles obtained for different inhomogeneity configuration are studied to assess the developed LV-DAS.

3 Results and Discussion

DDTM studies show that the LV-GUI and LV-DAS works perfectly and generates the digital data in correct sequence. Results show that all the MUXs are operated accurately efficiently switch all the surface electrodes of the phantom. Result show that the LV-DAS successfully and efficiently acquires the boundary data from practical phantoms. Figure 5 show the boundary data profiles obtained from a NaCl phantom using the LV-DAS with 1 mA (rms), 50 kHz sinusoidal constant current injection with opposite current pattern. Figure 5a shows the boundary data profile of a NaCl phantom without inhomogeneity acquired by LV-DAS with 1 mA, 50 kHz current in opposite current pattern. Figure 5a shows that the potential profiles obtained for all the projections for a homogeneous medium are symmetric which indicates the proper and efficient operation of LV-DAS. Figure 5b, c shows the boundary data profiles acquired by LV-DAS with 1 mA, 50 kHz current in opposite current pattern for the NaCl phantoms with a nylon cylinders (diameter 40 mm) near electrode no. 3 and 5 respectively. Boundary data obtained for both the phantom show that, for the phantoms with inhomogeneity placed near Nth electrode, the maximum potential (V_{ei}) points [15, 58] in their corresponding boundary data profiles occurs in $[(N-1) \times 16] + N$ th data point in the boundary potential data matrix which exactly correlate with the results reported by Bera and Nagaraju [15, 58]. As described by Bera and Nagaraju [15, 58], the maximum potential (V_{ei}) point in the boundary potential data matrix is the potential of the electrode near the inhomogeneity when the positive terminal of the current source is connected to the same electrode.

The boundary data profiles obtained from a NaCl phantom using the LV-DAS with 1 mA (rms), 50 kHz sinusoidal constant current injection with neighbouring current pattern is shown in Fig. 6. Figure 6a shows the boundary data profile of a NaCl phantom without inhomogeneity acquired by LV-DAS with opposite current pattern by injecting a 1 mA, 50 kHz constant current.

The potential profiles as shown in Fig. 6a demonstrate that, for all the projections the boundary data profiles for a homogeneous medium are symmetric which indicates the proper and efficient operation of LV-DAS. Figure 6b, c shows the boundary data profiles obtained for a 1 mA, 50 kHz current injection by LV-DAS with neighbouring current pattern applied to the NaCl phantoms with a nylon cylinders (diameter 40 mm) near electrode no. 3 and 5 respectively. In the neighbouring current pattern also it is observed that, for the phantoms with inhomogeneity placed near Nth electrode, the maximum potential (V_{ei}) points [15, 58] in their corresponding boundary data profiles occurs in $[(N-1) \times 16] + N$ th data point in the boundary potential data matrix as reported by Bera and Nagaraju [15, 58]. Experimental results show that the EIT impedance image reconstruction can be successfully conducted with LV-DAS.

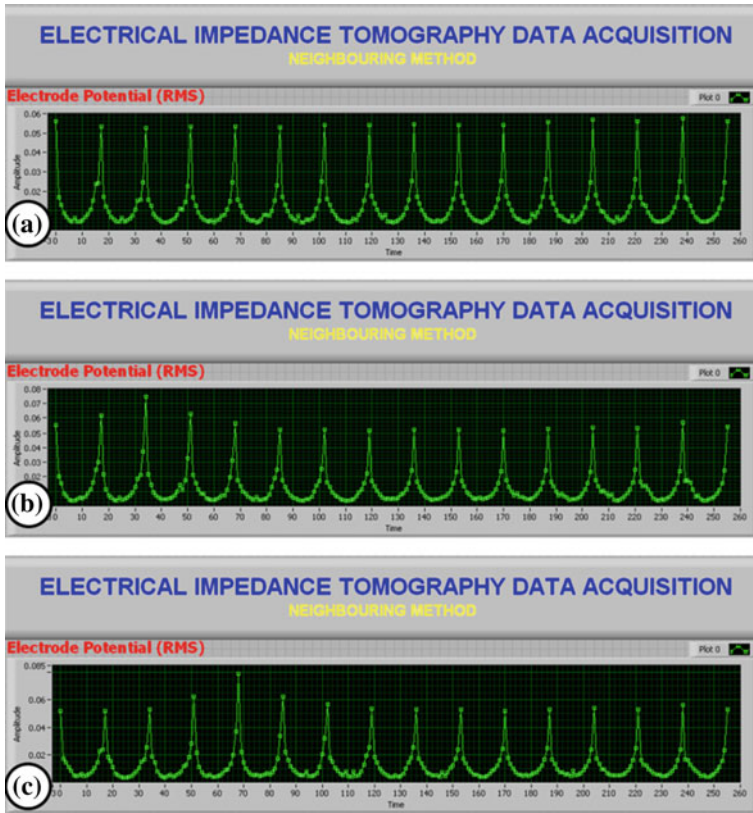


Fig. 6 Boundary data profile of a NaCl phantom acquired by LV-DAS with 1 mA, 50 kHz current in neighbouring current pattern. **a** Without inhomogeneity. **b** Nylon cylinder (40 mm diameter) near E_3 . **c** Nylon cylinder (40 mm diameter) near E_5

4 Conclusion

A LabVIEW based data acquisition system (LV-DAS) is developed for Electrical Impedance Tomography (EIT) and tested with practical phantoms. The LV-DAS is developed with a NIUSB-6251 DAQ card, NISCB-68 Connector Module and an automatic electrode switching module (A-ESM). A LabVIEW based graphical user interface (LV-GUI) is develop to control the current injection by the CCI and data acquisition by LV-DAS through A-ESM. Digital bits generation, MUX operation and the electrode switching are studied with DDTM circuit. Boundary data are collected with different phantom configurations and different current injection protocols and the data profiles are studied to assess the LV-DAS operation and performance. DDTM results show that LV-DAS efficiently and accurately generates the digital bits required for efficient electrode switching. MUXs are operated successfully by the digital bits and the electrodes are switched

accurately for all the current patterns. Boundary data obtained by LV-DAS for NaCl phantoms show that the LV-DAS efficiently collect the boundary data with all the phantom configurations and all the current patterns. The potential profiles obtained for all the projections for a homogeneous medium are found symmetric for both the current patterns which indicates the proper and efficient operation of LV-DAS. Boundary data profile studies also show that, for the phantoms with inhomogeneity placed near N th electrode, the maximum potential (V_{ei}) points in their corresponding boundary data profiles occurs in $[(N - 1) \times 16] + N$ th data point in the boundary potential data matrix which exactly correlate with the results reported in previous studies.

References

1. Webster, J.G.: *Electrical Impedance Tomography*. Adam Hilger Series of Biomedical Engineering, Adam Hilger, New York (1990)
2. Bera, T.K., Nagaraju, J.: *Electrical Impedance Tomography (EIT): A Harmless Medical Imaging Modality*. *Research Developments in Computer Vision and Image Processing: Methodologies and Applications*, Chap. 13, pp. 224–262. IGI Global, USA (2013)
3. Wanga, P., Guo, B., Li, N.: Multi-index optimization design for electrical resistance tomography sensor. *Measurement* **46**, 2845–2853 (2013)
4. Bera, T.K., Nagaraju, J.: Studies on the thin film based flexible gold electrode arrays for resistivity imaging in electrical impedance tomography. *Measurement*. **47**, 264–286 (2014). Impact Factor: 1.130
5. Holder, D.S.: *Electrical Impedance Tomography: Methods, History and Applications*, 1st edn. Institute of Physics Publishing Ltd., UK (2005)
6. Bera, T.K., Nagaraju, J.: A MATLAB based boundary data simulator for studying the resistivity reconstruction using neighbouring current pattern. *J. Med. Eng.* **15** (2013) (Article ID 193578)
7. Bera, T.K., Nagaraju, J.: Elemental resistivity profile analysis of EIT images to assess the reconstructed image quality. *Int. J. Inf. Process.* **7**(1), 1–14 (2013)
8. Bushberg, J.T., Seibert, J.A., Leidholdt Jr., E.M., Boone, J.M.: *The Essential Physics of Medical Imaging*, 3rd edition. Lippincott Williams & Wilkins; Third, North American Edition edition (20 Dec 2011)
9. Hiller, J., Reindl, L.M.: A computer simulation platform for the estimation of measurement uncertainties in dimensional X-ray computed tomography. *Measurement* **45**(8), 2166–2182 (2012)
10. Davis, J., Wells, P.: Computed tomography measurements on wood. *Ind. Metrol.* **2**(3–4), 195–218 (1992)
11. Bera, T.K., Nagaraju, J.: Sensors for electrical impedance tomography. In: Webster, J.G. (ed.) *The Measurement, Instrumentation, and Sensors Handbook*, 2nd edition. CRC Press, Boca Raton, Chap. 61, pp. 61-1–61-30 (2014)
12. Lionheart, W.R.B.: EIT reconstruction algorithms: pitfalls, challenges and recent developments. *Review Article, Physiol. Meas.* **25**, 125–142. PII: S0967-3334(04)70421-9 (2004)
13. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving conductivity image quality using block matrix-based multiple regularization (BMMR) technique in EIT: a simulation study. *J. Electr. Bioimp.* **2**, 33–47 (2011)

14. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving image quality in electrical impedance tomography (EIT) using projection error propagation-based regularization (PEPR) technique: a simulation study. *J. Electr. Bioimp.* **2**, 2–12 (2011)
15. Bera, T.K., Nagaraju, J.: Resistivity imaging of a reconfigurable phantom with circular inhomogeneities in 2D-electrical impedance tomography. *Measurement* **44**(3), 518–526 (2011)
16. Kerner, T.E., Williams, D.B., Osterman, K.S., Reiss, F.R., Hartov, A., Paulsen, K.D.: Electrical impedance imaging at multiple frequencies in phantoms. *Physiol. Meas.* **21**, 67–77 (2000)
17. Griffiths, H.: A cole phantom for EIT. *Physiol. Meas.* **16**(1995), A29–A38 (1995)
18. Kim, B.S., Kim, K.Y., Kao, T.J., Newell, J.C., Isaacson, D., Saulnier, G.J.: Dynamic electrical impedance imaging of a chest phantom using the Kalman filter. *Physiol. Meas.* **27**(5), S81–S91 (2006)
19. Kimoto, A., Shida, K.: Imaging of temperature-change distribution in the brain phantom by means of capacitance measurement. *IEEE Trans. Instrum. Measur.* **49**(3), 591–595 (2000)
20. Li, Yi: Manucher soleimani imaging conductive materials with high frequency electrical capacitance tomography. *Measurement* **46**, 3355–3361 (2013)
21. Sadleiry, R., Foxz, R.: Quantification of blood volume by electrical impedance tomography using a tissue-equivalent phantom. *Physiol. Meas.* **19**, 501–516 (1998)
22. Holder, D.S., Khan, A.: Use of polyacrylamide gels in a saline-filled tank to determine the linearity of the Sheffield Mark 1 electrical impedance tomography (EIT) system in measuring impedance disturbances. *Physiol. Meas.* **15**, A45–A50 (1994)
23. Bera, T.K., Nagaraju, J.: A chicken tissue phantom for studying an electrical impedance tomography (EIT) system suitable for clinical imaging. *Sens. Imaging Int. J.* **12**(3–4), 95–116 (2011)
24. Kao, T.J., Saulnier, G.J., Isaacson, D., Szabo, T.L., Newell, J.C.: A versatile high-permittivity phantom for EIT. *IEEE Trans. Biomed. Eng.* **55**(11), 2601 (2008)
25. Wanga, P., Guo, B., Li, N.: Multi-index optimization design for electrical resistance tomography sensor. *Measurement* **46**, 2845–2853 (2013)
26. Bera, T.K., Nagaraju, J.: A stainless steel electrode phantom to study the forward problem of electrical impedance tomography (EIT). *Sens. Transducers J.* **104**(5), 33–40 (2009)
27. Bera, T.K., Nagaraju, J.: A multifrequency electrical impedance tomography (EIT) system for biomedical imaging. In: *IEEE International Conference on Signal Processing and Communications (SPCOM 2012)*, IISc-Bangalore, India pp. 1–5
28. Bera, T.K., Nagaraju, J.: A simple instrumentation calibration technique for electrical impedance tomography (EIT) using a 16-electrode phantom. In: *Proceedings of The Fifth Annual IEEE Conference on Automation Science and Engineering (IEEE CASE 2009)*, India, 2009, pp. 347–352
29. Bera, T.K., Nagaraju, J.: A study of practical biological phantoms with simple instrumentation for electrical impedance tomography (EIT). In: *Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC2009)*, Singapore, pp. 511–516, 5th–7th May 2009
30. Robitaille, N., Guardo, R., Maurice, I., Hartinger, A.E., Gagnon, H.: A multi-frequency EIT system design based on telecommunication signal processors. *Physiol. Meas.* **30**, S57–S71 (2009)
31. Goharian, M., Soleimani, M., Jegatheesan, A., Chin, K., Moran, G.R.: A DSP based multi-frequency 3D electrical impedance tomography system. *Ann. Biomed. Eng.* **36**, 1594–1603 (2008)
32. Oh, T.I., Koo, H., Lee, K.H., Kim, S.M., Lee, J., Kim, S.W., Seo, J.K., Woo, E.J.: Validation of a multi-frequency electrical impedance tomography (mfEIT) system KHU Mark1: impedance spectroscopy and time-difference imaging. *Physiol. Meas.* **29**, 295–307 (2008)
33. Jennings, D., Schneider, I.D.: Front-end architecture for a multifrequency electrical impedance tomography system. *Med. Biol. Eng. Compu.* **39**(3), 368–374 (2001)

34. Mohamadou, Y., Oh, T.I., Wi, H., Sohal, H., Farooq, A., Woo, E. J., McEwan, A.: Performance evaluation of wideband bio-impedance spectroscopy using constant voltage source and constant current source. *Meas. Sci. Technol.* **23**(10), 105703 (2012)
35. Ross, A.S., Saulnier, G.J., Newell, J.C., Isaacson, D.: Current source design for electrical impedance tomography. *Physiol. Meas.* **24**, 509–516 (2003)
36. Lee, J.W., Oh, T.I., Paek, S.M., Lee, J.S., Woo, E.J.: Precision constant current source for electrical impedance tomography. In: *Proceedings of the 25th Annual International Conference of the IEEE EMBS, Cancun, Mexico*, pp. 1066–1069 (2003)
37. Gnechi, J.A.G.: Voltage controlled current source (VCCCS) for electrical impedance tomography (EIT) measurements in the α and β dispersion frequency ranges. In: *2010 Electronics, Robotics and Automotive Mechanics Conference*, pp. 677–681
38. Bera, T.K., Nagaraju, J.: A multifrequency constant current source suitable for electrical impedance tomography (EIT). In: *Proceedings of 2010 International Conference on Systems in Medicine and Biology. IIT Kharagpur, India*, pp. 278–283, 16–18 Dec 2010
39. Silverio, E.A.A., Silverio, E.A.A.: A high output impedance current source for wideband bioimpedance spectroscopy using 0.35 μm Tsmc Cmos technology. *Int. J. Eng. Appl. Sci.* **1**(2), 68–75 (2012)
40. Corrêa Alegria, F., Martinho, E., Almeida, F.: Measuring soil contamination with the time domain induced polarization method using LabVIEW. *Measurement* **42**, 1082–1091 (2009)
41. Morse D.H., Antolak A.J., Bench G.S., Roberts M.L.: A flexible LabVIEWTM-based data acquisition and analysis system for scanning microscopy. *Nucl. Instrum. Methods Phys. Res. Sect. B: Beam Interact. Mater. Atoms*, **158**(1), 146–152(7) (2 Sept 1999)
42. D’Mello, P.C., D’Souza, S.: Design and development of a virtual instrument for bio-signal acquisition and processing using LabVIEW. *Int. J. Adv. Res. Electr. Electron. Instrum. Eng.* **1**(1), 1–9 (2012)
43. Sumathi, S., Surekha, P.: *LabVIEW Based Advanced Instrumentation Systems*, 1st edn. Springer, Berlin (2007)
44. Czerwinski, F., Oddershede, L.B.: TimeSeriesStreaming.vi: LabVIEW program for reliable data streaming of large analog time series. *Comput. Phys. Commun.* **182**, 485–489 (2011)
45. Bo, L., Liu, X., He, X.: Measurement system for wind turbines noises assessment based on LabVIEW. *Measurement* **44**, 445–453 (2011)
46. Wang, Z., Shang, Y., Liu, J., Xidong, W.: A LabVIEW based automatic test system for sieving chips. *Measurement* **46**(1), 402–410 (2013)
47. Giannone, L., Eich, T., Fuchs, J.C., Ravindran, M., Ruan, Q., Wenzel, L., Cerna, M., Concezzi, S.: Data acquisition and real-time bolometer tomography using LabVIEW RT. *Fusion Eng. Des.* **86**, 1129–1132 (2011)
48. Yue, X., Drakakis, E.M., Lim, M., Radomska, A., Ye, H., Mantalaris, A., Panoskaltis, N., Cass, A.: A real-time multi-channel monitoring system for stem cell culture process. *IEEE Trans. Biomed. Circuits Syst.* **2**(2), 66–77 (2008)
49. Fontenot, R.S., Hollerman, W.A., Aggarwal, M.D., Bhat, K.N., Goedeke, S.M.: A versatile low-cost laboratory apparatus for testing triboluminescent materials. *Measurement* **45**, 431–436 (2012)
50. Andrei, H., Dogaru-Ulieru, V., Chicco, G., Cepisca, C., Spertino, F.: Photovoltaic applications. *J. Mater. Process. Technol.* **181**, 267–273 (2007)
51. Ni, J.-Q., Heber, A.J.: An on-site computer system for comprehensive agricultural air quality research. *Comput. Electron. Agric.* **71**, 38–49 (2010)
52. Ruiz, M., L’opez, J.M., de Arcas, G., Barrera, E., Melendez, R., Vega, J.: Data reduction in the ITMS system through a data acquisition model with self-adaptive sampling rate. *Fusion Eng. Des.* **83**, 358–362 (2008)
53. Mekida, S., Vacharanukul, K.: In-process out-of-roundness measurement probe for turned work pieces. *Measurement* **44**, 762–766 (2011)
54. Giannone, L., et al.: Data acquisition and real-time signal processing of plasma diagnostics on ASDEX upgrade using LabVIEW RT. *Fusion Eng. Des.* **85**, 303–307 (2010)

55. Ionel, R., Vasiu, G., Mischie, S.: GPRS based data acquisition and analysis system with mobile phone control. *Measurement* **45**, 1462–1470 (2012)
56. Bera, T.K., Nagaraju, J.: Studying the 2D resistivity reconstruction of stainless steel electrode phantoms using different current patterns of electrical impedance tomography (EIT). In: *Biomedical Engineering*, Narosa Publishing House, Proceeding of the International Conference on Biomedical Engineering 2011 (ICBME-2011), India, 2011, pp. 163–69
57. Bera, T.K., Nagaraju, J.: Studying the resistivity imaging of chicken tissue phantoms with different current patterns in electrical impedance tomography (EIT). *Measurement* **45**, 663–682 (2012)
58. Malmivuo, J., Plonsey, R.: *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, New York (1995)
59. Polydorides, N., Lionheart, W.R.B.: A Matlab toolkit for three-dimensional electrical impedance tomography: a contribution to the electrical impedance and diffuse optical reconstruction software project. *Meas. Sci. Technol.* **13**, 1871–1883 (2002)
60. Vauhkonen, M., Lionheart, W.R.B., Heikkinen, L.M., Vauhkonen, P.J., Kaipio, J.P.: A MATLAB package for the EIDORS project to reconstruct two dimensional EIT images. *Physiol. Meas.* **22**(107), 111 (2001)
61. Mekida, S., Vacharanukul, K.: In-process out-of-roundness measurement probe for turned workpieces. *Measurement* **44**, 762–766 (2011)
62. Data Sheet, NI USB 6251 OEM, High-Speed M Series Multifunction Data Acquisition (DAQ) Module, National Instruments, USA
63. Data Sheet, NI SCB68, Shielded I/O Connector Block, National Instruments, USA
64. Data Sheet, CD4067BE IC, CMOS Analog Multiplexers/Demultiplexers, Texas Instruments Inc., USA (2012)
65. Bera, T.K., Nagaraju, J.: Surface electrode switching of a 16-electrode wireless EIT system using RF-based digital data transmission scheme with 8 channel encoder/decoder ICs. *Measurement* **45**, 541–555 (2012)
66. Cheng, K.S., Simske, S.J., Isaacson, D., Newell, J.C., Gisser, D.G.: Errors due to measuring voltage on current-carrying electrodes in electric current computed tomography. *IEEE Trans. Biomed. Eng.* **37**(60), 60–65 (1990)
67. Rosell, J., Riu, P.: Common-mode feedback in electrical impedance tomography. *Clin. Phys. Physiol. Meas.* **13**(Suppl. 4), 11–14 (1992)
68. Rahal, M., Rida, I., Usman, M., Demosthenous, A.: New techniques to reduce the common-mode signal in multi-frequency EIT applications. In: *PIERS Proceedings, Marrakesh, MOROCCO*, pp. 1598–1601 (20–23 March 2011)

Crop Identification by Fuzzy C-Mean in Ravi Season Using Multi-Spectral Temporal Images

Sananda Kundu, Deepak Khare, Arun Mondal and P. K. Mishra

Abstract Information regarding spatial distribution of different crops in a region of multi-cropping system is required for management and planning. In the present study, multi dated LISS-III and AWiFS data were used for crop identification. The cultivable land area extracted from the landuse classification of LISS-III image was used to generate spectral-temporal profile of crops according to their growth stages with Normalised Difference Vegetation Index (NDVI) method. The reflectance from the crops on 9 different dates identified separate spectral behavior. This combined NDVI image was then classified by Fuzzy C-Mean (FCM) method again to get 5 crop types for around 12,000 km² area on Narmada river basin of Madhya Pradesh. The accuracy assessment of the classification showed overall accuracy of 88 % and overall Kappa of 0.83. The crop identification was done for one entire Ravi season from 23 October 2011 to 10 March 2012.

Keywords Crop identification · NDVI · Fuzzy C-Mean · Narmada river basin

S. Kundu (✉) · D. Khare · A. Mondal

Department of Water Resources Development and Management, Indian Institute of Technology, Roorkee, India
e-mail: sanandakundu@gmail.com

D. Khare
e-mail: kharefwt@gmail.com

A. Mondal
e-mail: arun.iirs@gmail.com

P. K. Mishra
Water Resources Systems Division National Institute of Hydrology, Roorkee, India
e-mail: erprabhash@gmail.com

1 Introduction

In Indian climatic condition, various types of crops are grown in a place. Monoculture of crops is done only in few places and so most of the time spectral signatures of different crops overlap each other in a same date. Thus it is important to identify these crops and from a single date image it is quite difficult to do crop mapping [1, 2]. To identify the crops, high resolution data of different time series were used to create NDVI and classify the landuse into different vegetation classes [3]. The use of spectral-temporal profile from satellite images to identify crops was initiated in 1980s [4] when a crop profile was presented by using vegetation indicator which measures the reflectance [5]. Many studies were done with the Normalized Difference Vegetation Index (NDVI) by generating profiles from temporal images [6, 7]. NDVI is considered as the most commonly used index for application in agricultural field or for crop identification [8]. Wardlow et al. [1] used MODIS EVI and NDVI (Normalised Difference Vegetation Index) for 12 months data series from the agricultural fields. It is possible to separate the crops during the growth stages at some point from one another. Crop identification from the temporal data series for 8 dates was performed by Doriaswamy et al. [9] with NDVI from MODIS and similar work with MODIS data was done by Ying et al. [10] with 4 images to demarcate wheat crop from others. Based on the crop calendar and growing season of the crops, AWiFS data was used for classification of crops [11].

One of the necessary things is to know the optimum number of dates or images to be used in the crop classification purpose or to have maximum separability. The highest separability with various combinations of dates was found by Murthy et al. [12] to study overlap of spectral signatures among various crops. Similar other works were conducted by Niel et al. [13] and Zurita-Milla et al. [14]. Proper identification of crops is again related to the landuse classifications and problems of mixed pixels. Different techniques used to solve this problem are Fuzzy classifications, Neural Networks, etc. The fuzzy classification is a model where an individual pixel can have partial membership which corresponds to several landuse classes [15]. Wang [16] explained that fuzzy method of classification, due to its multiple membership helps in achieving better accuracy. This method was used by various researchers to improve the identification of crop classes which improves the accuracy of the classification and helps to predict further production [17, 18]. For mapping of vegetation, Vegetation Index was used by combination of two or more spectral bands [19]. Xie et al. [20] also reviewed that NDVI is an important technique for mapping vegetation cover. Kappa coefficient is a discrete multivariate technique used for accuracy assessment [21].

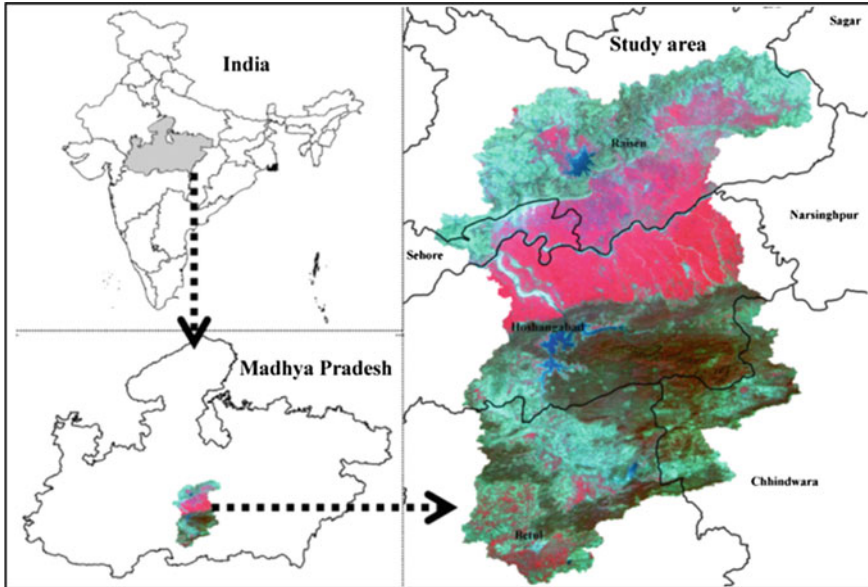


Fig. 1 Study area

2 Study Area

The study area is a part of Narmada River basin in Madhya Pradesh. The area extends from 21°47'24" to 23°25'49"N latitude and from 77°35'04" to 78°44'E longitude with total area of 12,678 km². It covers five districts of Raisen, Ho-shangabad, Sehore, Betur and Chhindwara. The region experiences subtropical type of climate with hot dry summer and a cool dry winter. The average rainfall is about 1,370 mm which decreases from east to west (Fig. 1). The present research involves the landuse classification with the Fuzzy C-Mean algorithm and identification of croplands using 9 temporal satellite images of AWiFS and LISS-III data from October, 2011 to March, 2012.

3 Data and Methodology

AWiFS and LISS-III data sets were taken for the study. All the images were projected with UTM zone 43 projection and WGS 84 datum. The images were registered by first order polynomial model and with the Root Mean Square Error (RMSE) of 0.5 pixel. Then the radiometric normalization was done. The corrected LISS-III images were classified with good accuracy and the agricultural land was extracted from the classified images. The NDVI method was applied on the 9

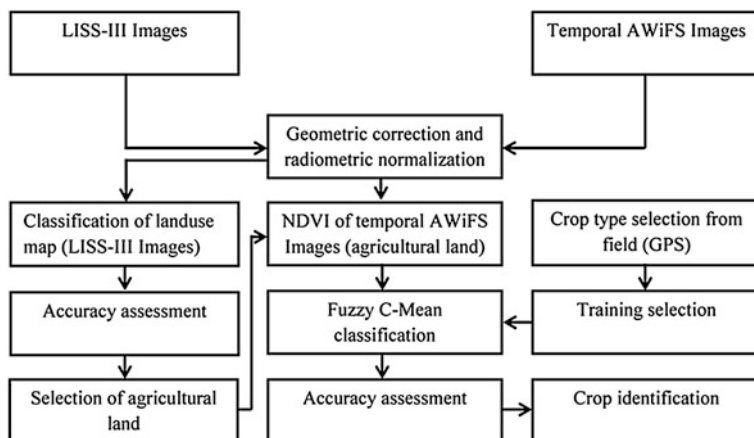


Fig. 2 Methodology

Table 1 Data list

AWiFS				
Sl No.	Date	Band	Wavelength (μm)	Resolution (m)
1	23-Oct-11	2	0.52–0.59 (green)	60
2	11-Nov-11	3	0.62–0.68 (red)	60
3	21-Nov-11	4	0.77–0.86 (near-IR)	60
4	10-Dec-11	5	1.55–1.70 (SWIR)	60
5	24-Dec-11			
6	12-Jan-12			
7	5-Feb-12			
8	20-Feb-12			
9	10-Mar-12			

images of AWiFS data on agricultural land to get the spectral difference of various crops at different growing stages. These images were then stacked to obtain a single image which was classified with Fuzzy C-Mean (FCM) and accuracy assessment with Kappa statistic was done (Fig. 2). The data used are given in Table 1.

3.1 NDVI Method

The temporal AWiFS data of 9 dates give the difference in spectral signature. The NDVI images were generated from the LISS-III and AWiFS images which give unique spectral response of each crop and demarcate them from the non-vegetation classes. These NDVI images were stacked chronologically according to the dates

to get the growth profile of the crops. These profiles were then used for FCM classification with the training data set from the images and samples collected from the field.

In NDVI, absorption of chlorophyll or green pigments in the red spectrum and very high reflectivity in the near infra-red spectrum helps in identifying and differentiating vegetation types. It also indicates the health of the vegetation and hence is important in identifying crops. It is calculated as:

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \tag{1}$$

where, ρ_{nir} is the near infra-red band and ρ_{red} is the red band of the sensor.

3.2 Fuzzy C-Mean Classification

In this study, Fuzzy C-Mean (FCM) algorithm was applied for the classification. It is based on the minimization of the objective function through iteration:

$$J_m(U, v) = \sum_{i=1}^C \sum_{k=1}^N u_{ik}^m \|y_k - v_i\|_A^2 \tag{2}$$

where,

- Y is $\{Y_1, Y_2 \dots Y_N\} \subset R^n$ = the data,
- c is number of clusters in Y; $2 \leq c \leq n$,
- m is weighting exponent; $1 \leq m < \infty$,
- U is fuzzy c-partition of Y; $U \in M_{fc}$,
- v is $(v_1, v_2 \dots v_c)$ = vectors of centers,
- v_i is $(v_{i1}, v_{i2}, \dots, v_{in})$ = center of cluster i,
- $\|\cdot\|_A$ is induced A-norm on R^n , and
- A is positive-definite $(n \times n)$ weight matrix.

The following three constraints are satisfied by the membership values:

$$0 \leq u_{ik} \leq 1; \text{ where } i \in \{1, \dots, C\}; \quad k \in \{1, \dots, N\} \tag{3}$$

$$\sum_{i=1}^C u_{ik} = 1; \quad k \in \{1, \dots, N\} \tag{4}$$

$$\sum_{k=1}^N u_{ik} > 0; \quad i \in \{1, \dots, C\} \tag{5}$$

The objective function is represented as the sum of the square of the Euclidean distances between each input sample and its corresponding cluster centre, and the distances are weighted by the fuzzy memberships. This is an iterative algorithm and uses the following equations:

Table 2 Area covered by different classes

SI No.	Crop type	Area (km ²)	Area (%)
1	Settlement	281.00	2.20
2	Grassland	2,942.60	23.06
3	Forest	918.94	7.20
4	Open/Degraded forest	1,485.06	11.64
5	Double crop land	2,608.29	20.44
6	Ravi crop (wheat)	1,689.17	13.23
7	Water body	520.72	4.08
8	Kharif crop (paddy)	2,272.75	17.81
9	River bank	44.85	0.35
	Total	12,763.37	100

$$\hat{v}_i = \left[\sum_{k=1}^N u_{ik}^m y_k \right] / \sum_{k=1}^N u_{ik}^m \quad (6)$$

$$\hat{u}_{ik} = 1 / \sum_{j=1}^C \left[\|y_k - v_i\| / \|y_k - v_j\| \right]^{2/(m-1)} \quad (7)$$

All input samples are considered and the contributions of the samples are weighted by the membership values for calculating cluster centre. The membership value in each class of a sample depends on its distance to the corresponding cluster centre. The weight factor m decreases the impact of small membership values. Greater the value of m , smaller will be the influence of samples with small membership values [22].

4 Results and Analysis

4.1 Landuse Classification and Accuracy Assessment

The LISS-III image was classified into 9 landuse classes, viz., settlement, grassland, forest, open/degraded forest, double crop land, ravi crop (wheat), water body, kharif crop (paddy) and river bank. The area covered by different classes is given in Table 2. The classified landuse is given in Fig. 3.

4.2 Crop Identification

The landuse classified from LISS-III image was used to extract cultivable area or agricultural land to develop NDVI profile. The NDVI values from 9 images were stacked to form one image. Difference in the spectral-temporal curves is observed in Fig. 4 where spectral reflectance in NDVI of each crop is different from the other.

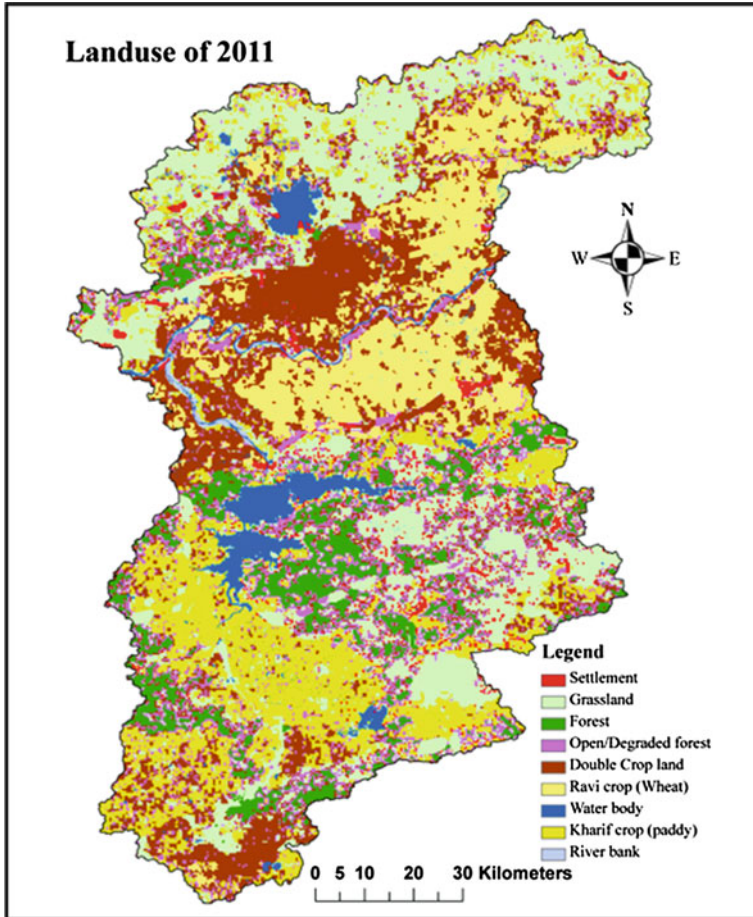


Fig. 3 Landuse classification of 2011

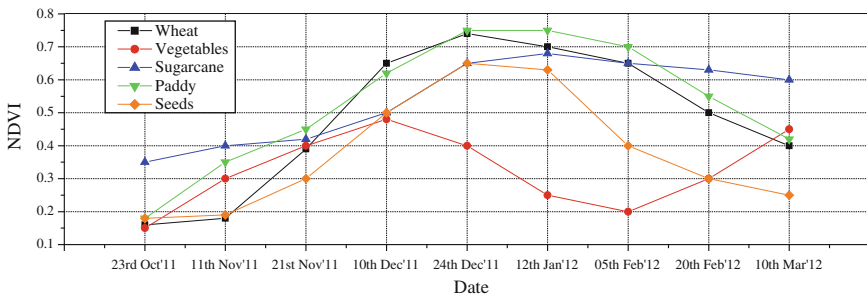


Fig. 4 Spectral curves of different landuse classes

Table 3 Area occupied by different crops

Sl No.	Crop type	Area (km ²)	Area (%)
1	Wheat	1,930.00	44.91
2	Vegetables	577.88	13.45
3	Seeds	1,212.49	28.21
4	Paddy	277.22	6.45
5	Sugarcane	299.86	6.98
	Total	4,297.45	100

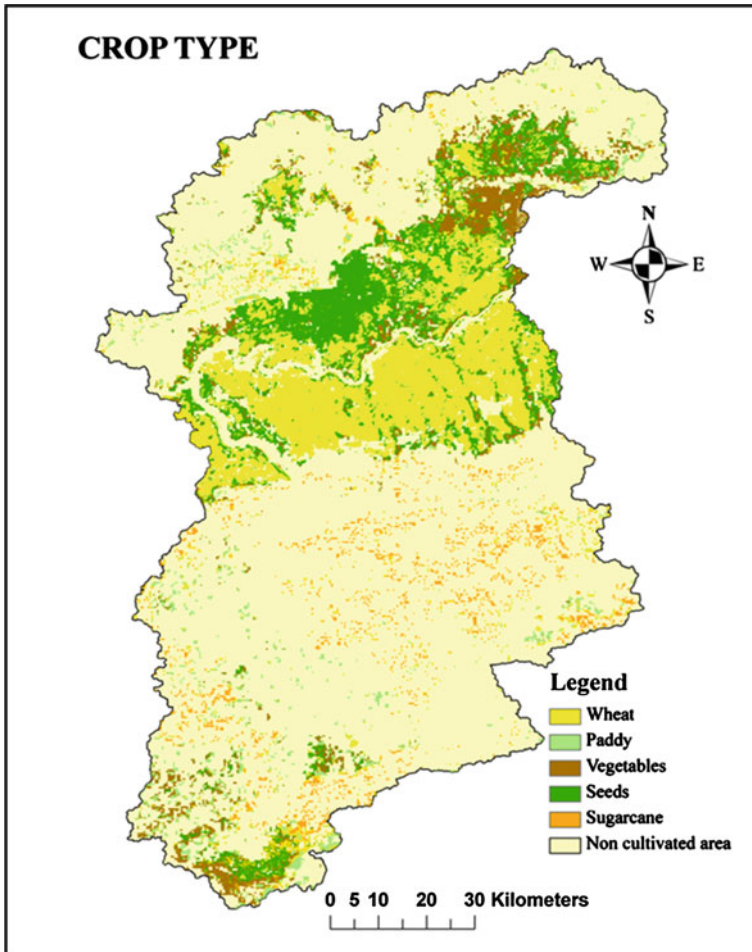


Fig. 5 Crop identification by Fuzzy C-Mean

Table 4 Accuracy assessment of the crops

Rabi crops					
Class name	Reference totals	Classified totals	Number of correct points	Producers accuracy (%)	Users accuracy (%)
Wheat	51	52	46	90.20	88.46
Paddy	10	10	8	80.00	80.00
Vegetables	18	18	16	88.89	88.89
Sugarcane	11	11	9	81.82	81.82
Seeds	35	34	31	88.57	91.18
Total	125	125	110

The stacked NDVI image was then classified by Fuzzy C-Mean into different croplands namely wheat, sugarcane, seeds, vegetables, paddy and non-cultivable areas. Among all the crops, wheat is found to have highest proportion of area with around 45 % of the total area followed by seeds (28.21 %), vegetables, sugarcane and paddy. This difference is reflected in the NDVI values of growth stage from the temporal images. Table 3 signifies the area occupied by different crop types.

Various crops identified by FCM are given in Fig. 5. The crops were identified with the help of crop profiles of NDVI. Total of 125 samples were collected from the field for 5 crops classified in the image. The Producer's and User's accuracy achieved for 5 different crops were 80 % and above (Table 4). The overall accuracy was 88 % and the overall Kappa was 0.83.

5 Conclusions

The availability of temporal images is necessary for obtaining crop profile. The identification of crops and its spatial distribution is required to understand the agriculture of a region and for further planning and management. The given study involves identification of crop profiles through NDVI from 9 temporal AWiFS data for the entire Ravi season from 2011 October to 2012 March. The agricultural land was extracted from the classified landuse of LISS-III image and NDVI profiles were drawn only for this portion of landuse. The profiles indicated different spectral signatures at the growing stages for different crops which helped in identifying the mixed crops of the area which are grown side by side. Further, the Fuzzy C-Mean classification was done with the stacked NDVI image to provide definite crop types of the area and their distribution in Ravi season with accuracy assessment. Thus this method provides spatially distributed crop data within a region.

Acknowledgments The authors are thankful to the National Remote Sensing Centre (NRSC) for providing the AWiFS and LISS-III satellite images for the study area and to the UGC for financial assistance.

References

1. Wardlow, B., Egbert, S., Kastens, J.: Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. *Rem. Sens. Environ.* **108**, 290–310 (2007)
2. Masialeti, I., Egbert, S., Wardlow, B.D.: A comparative analysis of phenological curves for major crops in Kansas. *GIScience Remote Sens.* **47**(2), 241–259 (2010)
3. Vincent, S., Pierre, F.: Identifying main crop classes in an irrigated area using high resolution image time series. In: *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS'03)*, pp. 252–254 (2003)
4. Foerster, S., Kaden, K., Foerster, M., Itzerott, S.: Crop type mapping using spectral–temporal profiles and phenological information. *Comput. Electron. Agric.* **89**, 30–40 (2012)
5. Odenweller, J.B., Johnson, K.: Crop identification using Landsat temporal spectral profiles. *Rem. Sens. Environ.* **14**, 39–54 (1984)
6. Bradley, B.A., Jacob, R.W., Hermance, J.F., Mustard, J.F.: A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data. *Remote Sens. Environ.* **106**, 137–145 (2007)
7. Wardlow, B., Egbert, S.: Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. *Rem. Sens. Environ.* **112**, 1096–1116 (2008)
8. Reed, B.R., White, M.A., Brown, J.F.: Remote sensing phenology. In: Schwartz, M.D. (ed.) *Phenology: An Integrative Environmental Science*, pp. 365–381. Kluwer Academic Publishers, New York (2003)
9. Doriaswamy, P.C., Akhmedov, B., Stern, A.J.: Improved techniques for crop classification using MODIS imagery. In: *IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS'06)*, pp. 2084–2087 (2006)
10. Ying, L., Xiuwan, C., Hongwei, D., Lingkui, M.: An improved multi temporal masking classification method for winter wheat identification. In: *International Conference on Audio Language and Image Processing (ICALIP'10)*, pp. 1648–1651 (2010)
11. Panigrahy, R.K., Ray, S.S., Panigrahy, S.: Study on utility of IRS-P6 SWIR band for crop discrimination and classification. *J. Indian Soc. Remote Sens.* **37**(2), 325–333 (2009)
12. Murthy, C.S., Raju, P.V., Badrinath, K.V.S.: Classification of wheat crop with multi-temporal images: performance of maximum likelihood and artificial neural networks. *Int. J. Rem. Sens.* **24**(23), 4871–4890 (2003)
13. Van Niel, T.G., Mc Vicar, T.R.: Determining temporal windows for crop discrimination with remote sensing: a case study in South-Eastern Australia. *Comput. Electron. Agric.* **45**, 91–108 (2004)
14. Zurita-Milla, R., Gomez-Chova, L., Guanter, L., Clevers, J.G.P.W., Camps-Valls, G.: Multi temporal un-mixing of medium spatial resolution satellite images: a case study using MERIS images for land cover mapping. In: *IEEE Transactions on Geoscience and Remote Sensing Symposium*, pp. 1–10 (2011)
15. Jain, A., Dubes, R.: *Algorithms for Clustering Data*. Prentice-Hall, Englewood Cliffs (1988)
16. Wang, F.: Fuzzy supervised classification of remote sensing images. *IEEE Trans. Geosci. Remote Sens. Symp.* **28**, 194–201 (1990)
17. Dave, R.N.: Characterization of noise in clustering. *Pattern Recogn. Lett.* **12**, 657–664 (1991)
18. Liang, P., Chunyu, Y.: Study on mixed pixel classification method of remote sensing image based on fuzzy theory. In: *IEEE Conference Urban Remote Sensing Joint Event*, pp. 1–7 (2009)
19. Jackson, R.D., Huete, A.R.: Interpreting vegetation indices. *Preventory Vet. Med.* **11**, 185–200 (1991)

20. Xie, Y., Sha, Z., Yu, M.: Remote sensing imagery in vegetation mapping: a review. *J. Plant Ecol.* **1**(1), 9–23 (2008)
21. Lillesand, T.M., Kiefer, R.W., Chipman, J.W.: Remote sensing and image interpretation. Wiley, New York (2008)
22. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy C-means clustering algorithm. *Comput. Geosci.* **10**(2–3), 191–203 (1984)

Small World Particle Swarm Optimizer for Data Clustering

Megha Vora and T. T. Mirnalinee

Abstract Particle swarm is a stochastic optimization paradigm inspired by the concepts of social psychology and artificial intelligence. Population topology plays significant role in the performance of PSO. It determines the way in which particles communicate and share information with each other. Topology can be depicted as a network model. Regular networks are highly clustered but the characteristic path length grows linearly with the increase in number of vertices. On the contrary, random networks are not highly clustered but they have small characteristic path length. Small world networks have a distinctive combination of regular and random networks i.e. highly clustered and small characteristic path length. This paper presents a novel algorithm for data clustering by incorporating the concept of small world in particle swarm optimization. Efficiency of the proposed methodology is tested by applying it on five standard benchmark data set. Results obtained are compared with another PSO variant. Comparative study demonstrates the effectiveness of the proposed approach.

1 Introduction

Data clustering is a process of partitioning a set of data or objects into meaningful groups. Objective of data clustering is to discover natural grouping of a set of patterns, points, or objects. The grouping should be such that data items within one cluster are similar to each other, while those within different clusters should be

M. Vora (✉) · T. T. Mirnalinee
Department of Computer Science and Engineering, S.S.N College of Engineering,
Anna University, Chennai, India
e-mail: meghavora25@gmail.com

T. T. Mirnalinee
e-mail: mrinal23@gmail.com

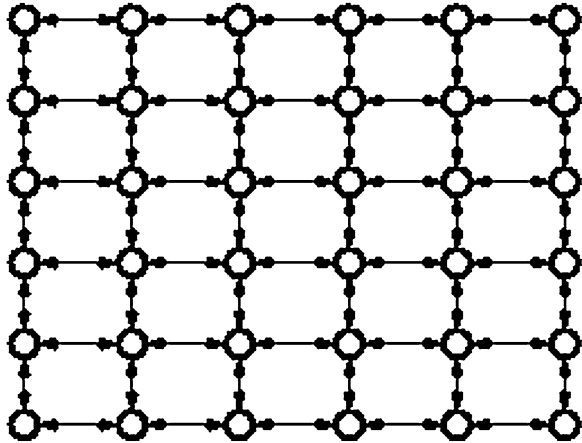
dissimilar. Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, document retrieval, image segmentation, object recognition, spatial data analysis, web mining and pattern classification [1].

In the past few decades numerous clustering methods have been proposed inspired by the biological world. These methods could be based either on evolutionary computation (eg. genetic algorithm), or on intelligent behaviour of social organism (e.g. ant colony, particle swarm). Particle swarm optimization algorithm was originally proposed by Kennedy and Eberhart [2]. It is inspired by the paradigm of birds flocking behaviour. It has received huge attention and popularity due to its algorithmic simplicity and effectiveness for solving global optimization problems. This derivative-free method is particularly suited to continuous variable problems and has received increasing attention in the optimization community. PSO consists of a swarm of particles and each particle flies through the multi-dimensional search space with a velocity. Velocity is constantly updated by the particle's previous best performance and by the previous best performance of the particle's neighbors. Even though PSO is a good and fast search algorithm, it suffers from premature convergence, especially in complex multi-peak-search problems.

The population topology has a significant effect on the performance of PSO. Topology determines the way particles communicate and share information within the swarm. Topology can be depicted as a network model. Regular networks are highly clustered but the characteristic path length grows linearly with the increase in number of vertices. On the contrary, random networks are not highly clustered but they have small characteristic path length. Small world networks have a distinctive combination of regular and random networks i.e. highly clustered and small characteristic path length. Mendes [3] studied the impact of various network topology on particle swarm performance. It was reported that path length led to a more balanced exploration-exploitation trade-off: If it is too small, then information spreads too fast, this means a higher probability of premature convergence. If it is large, then information takes a long time to travel through the graph and thus the population is more resilient and not so eager to exploit earlier on. Even a partial degree clustering (i.e. percentage of vertex's neighbours that are neighbours to one another) helps to disseminate information in the network. Thus we need to have a combination of highly clustered and small characteristic path length network. Small world network fits to this need of ours.

Small world concept was originally proposed by Milgram [4]. It was stated that social network exhibits small-world phenomenon in which any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances. To demonstrate the universality of this phenomenon in network arising in nature and technology, Milgram et al. in 1960s [4, 5] performed a series of striking experiments and reported their results. This was later modeled by Watts and Strogatz [6]. However, the model was insufficient to explain the striking algorithmic component of Milgram's original findings that how the individuals using local information are collectively effective at actually constructing a short

Fig. 1 A swarm of particles
6 × 6

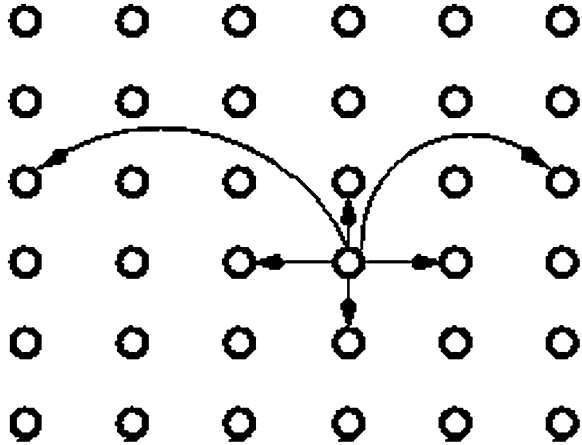


path between two points in a social network. Kleinberg [7] further generalized Watts and Strogatz model and showed that there is a decentralized algorithm capable of finding short paths with high probability. In [8] Small World Particle Swarm Optimization (SWPSO) algorithm was proposed by applying the concept of small world network model given by Jon Kleinberg in PSO. In this paper the concept of SWPSO is applied for data clustering viewing clustering as an optimization problem. Detail description of the algorithm is given in subsequent sections.

2 Small World PSO Based Data Clustering

As mentioned in the introduction, small world network has a distinctive combination of highly clustered and small characteristic path length network. This led to a balanced tradeoff between exploration and exploitation. Here, small world network topology on Von Neumann is network model is considered. This is achieved by adding few random particles (in this case two) and retaining four immediate Von Neumann neighbors of a current particle (see Figs. 1 and 2). These random particles are treated as additional neighbors of current particle. The intuitive advantage of these additional random neighbors is the way small world is formed within the current swarm. Each particle in the swarm represents the candidate solution to the considered problem. Particles travel in the solution space and attempts to move toward a better solution by changing its direction and speed based on its own past experience P_{best} and the experience of the particles in its small world neighbour $SW N_{best}$.

Fig. 2 Small world with two random particles



2.1 Cluster Formation

In the context of clustering N data points into C clusters, each k th particle represents the $X_k = \{x_{ki}, i = 1, 2, \dots, C\}$ cluster centroid vector. x_{ki} refers to the i th cluster centroid vector of the k th particle. The initial population X_k for each k th particle is randomly assigned such that $z_i^{min} < x_{ki} < z_i^{max}$, where z_i^{min} and z_i^{max} are the minimum and maximum value of the i th feature. Similarly for each particle initial velocity V_k is assigned randomly such that $z_i^{min}/2 < v_{ki} < z_i^{max}/2$.

Once initialization is done, for each data vector z_p euclidean distance $d(z_p, x_{ki})$ is calculated to all cluster centroids. Clusters are then formed by allocating the data vector z_p to the cluster centroid from which it has minimum distance. Now, fitness of each particle is computed using β index [9]. It computes the ratio of total variation and within class variation. It is defined as

$$\beta = \frac{\sum_{i=1}^C \sum_{p=1}^{n_i} (z_{ip} - \bar{z})^2}{\sum_{i=1}^C \sum_{p=1}^{n_i} (z_{ip} - \bar{z}_i)^2}, \tag{1}$$

where \bar{z} is the mean of all the data points and \bar{z}_i is the mean of the data points that belong to the cluster C_i ; z_{ip} is the p th data point of i th cluster and n_i is the number of data points in cluster C_i .

Since the numerator is a constant for a given data set, the value of β is dependent only on the denominator. The denominator decreases with the homogeneity in the formed clusters. Therefore, for a given data set, higher the value of β , better is the clustering. The proposed small world based PSO algorithm (SWPSO) for formation of cluster is given below.

1. Initialize the position \vec{X}_k and velocity \vec{V}_k of each k th particle in the swarm where k varies from $(1, 2, \dots, NP)$; NP is the size of swarm

2. Initialize fitness counter $fes = 0$. Define maximum number of fitness evaluation T_FEs
3. Compute the fitness F_k for each k th particle in the swarm using Eq. 1
4. Perform steps 5 to 13 until maximum number of fitness evaluation is not exceeded
5. For each of the k th particle perform steps 6 to 10
6. Compare $Pbest_k$ of the k th particle with its current fitness value F_k . If the current fitness value is better, then assign the current fitness value to $pbest_k$ and assign the current position to \vec{X}_{Pbest_k}
7. Compute the small world best fitness $SW Nbest_k$ by selecting particle with best fitness (here minimum) in the small world neighbourhood of the k th particle
8. Update the previous small world best fitness with the current value of $SW Nbest_k$ if it is minimum. Also, update previous position of the best small world particle with the current one $\vec{X}_{SW Nbest_k}$
9. Update velocity of the particle using Eq. 2

$$\vec{V}'_k = \omega V_k + C_1 R_1 (\vec{X}_{Pbest_k} - \vec{X}_k) + C_2 R_2 (\vec{X}_{SW Nbest_k} - \vec{X}_k) \quad (2)$$

where, ω is inertia weight, C_1 and C_2 are acceleration constant; R_1 and R_2 are random numbers uniformly generated from the interval $[0,1]$

10. If $\vec{X}'_k \in [z_i^{min}, z_i^{max}]$ If $\vec{X}'_k < z_i^{min}$ then $\vec{X}'_k = z_i^{min}/2$ else if $\vec{X}'_k > z_i^{max}$ then $\vec{X}'_k = z_i^{max}/2$;
11. Update the position of the particle using Eq. 3

$$\vec{X}'_k = \vec{X}_k + \vec{V}'_k \quad (3)$$

12. Compute the fitness of the updated particle
13. Increment the fitness counter $fes = fes + NP$.

3 Experimental Results

Performance of the proposed algorithm has been evaluated using five real benchmark data sets taken from the Machine Learning repository [10].

3.1 Description of Data Sets

A Summary about the dataset is given in Table 1. Wisconsin Breast Cancer (WBC) data contains 699 instances distributed in two categories described by nine features of which 16 instances with the missing values are ignored. Thyroid data set has 215 instances of patients with five features describing whether patient has euthyroidism, hypothyroidism and hyperthyroidism (three classes). The

Table 1 Description of data sets [10]

Data set	#Classes	#Features	Size
WBC	2	9	683
Thyroid	3	5	215
CMC	3	9	1473
Wine	3	13	178
Vowel	6	3	871

Contraceptive Method Choice (CMC) dataset is a subset of the 1987 National Indonesia Contraceptive Prevalence Survey. The objects are married women who either were not pregnant or did not know if they were at the time of interview. The problem involves predicting the choice of the current contraceptive method of a woman based on her demographic and socio-economic characteristics. This dataset contains 1,473 objects with nine attributes and three clusters. Wine data has 178 instances of three types of wine with thirteen features. Vowel data set consists of 871 patterns. There are six overlapping vowel classes and three input features.

3.2 Results

Results obtained by the proposed methodology are compared with the with PSOAG [11]. In PSOAG particle swarm optimization with age group topology is proposed. In order to keep population diversity during searching, particles are divided into different age groups by their age. Age measures the search ability of each particle in the local area. Particles in each age group can only select the ones in a younger group or their own group. PSOAG has been shown to be effective when compared with other swarm intelligence based techniques canonical PSO [12], ACO [13], ABC [14], DE [15].

To fairly compare the results obtained by SWPSO with already published results of PSOAG, simulations were done considering a swarm/population of size of 100 (10×10). Population was initialized in the search space using symmetric initialization strategy (i.e. population assumed to be uniformly distributed over the search space). Each run of the algorithm is terminated when the number of function evaluations (T_FEs) exceeds $1e+04$. Acceleration coefficient parameter C_1 and C_2 is set to 1.494. Inertia weight parameter ω is set to 0.72.

Table 2 shows the mean and standard deviation (within parentheses) of clustering accuracy for 20 runs for the two algorithms. Results in bold face indicate the best ones. It is found from the results that the proposed SWPSO based clustering algorithm gives significantly better results in terms of clustering accuracy in two test cases (Thyroid and Vowel) and is comparable in other three (WBC, CMC and Wine). It is to be noted that in all the test cases SWPSO algorithm gives significantly lower values of standard deviation compared to PSOAG. The lower value of

Table 2 Comparison of clustering accuracy with PSOAG

Data set	SWPSO		PSOAG [11]	
	Mean	Stdv	Mean	Stdv
WBC	96.60	0.002	96.31	0.20
Thyroid	85.76	0.019	74.37	10.95
CMC	39.68	0.004	39.87	0.30
Wine	70.22	0.000	70.98	0.33
Vowel	55.67	(0.024)	51.75	(4.25)

standard deviation implies that the proposed SWPSO based clustering algorithm is more stable.

4 Conclusions

In this paper we have proposed a new algorithm for clustering data based on small world particle swarm optimization. To evaluate the performance of the proposed algorithm we have tested it on five benchmark data sets. Comparative study of the proposed algorithm with the PSOAG algorithm justifies its potentiality in terms of both clustering accuracy and stability (standard deviation). In future, we plan to compare the performance of SWPSO for data clustering with some more existing algorithms and by using different validity criteria as fitness measure.

References

1. Jain, A.K., Murty, M.N., Flynn, P.J.: Data clustering: a review. *ACM Comput. Surv.* **31**(3), 264–323 (1999)
2. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. *IEEE Int. Conf. Neural Netw.* **4**, 1942–1948 (1995)
3. Mendes, R.: Population topologies and their influence in particle swarm performance. Ph.D. thesis, University of Minho (2004)
4. Milgram, S.: The small world problem. *Psychol. Today* **2**, 60–67 (1967)
5. Travers, J., Milgram, S.: An experimental study of the small world problem. *Sociometry* **32**, 425 (1969)
6. Watts, D., Strogatz, S.: Collective dynamics of small-world networks. *Nature* **393**, 440–442 (1998)
7. Kleinberg, J.: The small-world phenomenon: an algorithmic perspective. Technical report, Cornell University Ithaca, NY (1999)
8. Saxena, A.K., Vora, M.: Novel approach for the use of small world theory in particle swarm optimization. In: 16th International Conference on Advanced Computing and Communications, pp. 363–366. IEEE (2008)
9. Pal, S.K., Ghosh, A., Uma Shankar, B.: Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. *Int. J. Remote Sens.* **21**(11), 2269–2300 (2000)

10. Newman, D.J., Hettich, S., Blake, C.L., Merz, C.J.: UCI repository of machine learning databases. University of California, Irvine, Department of Information and Computer Sciences. <http://www.ics.uci.edu/~mlearn/MLRepository.html>. (1998)
11. Jiang, B., Wang, N., Wang, L.: Particle swarm optimization with age-group topology for multimodal functions and data clustering. *Commun. Nonlinear Sci. Numer. Simul.* **18**(11), 3134–3145 (2013)
12. Cura, T.: A particle swarm optimization approach to clustering. *Expert Syst. Appl.* **39**(1), 1582–1588 (2012)
13. Shelokar, P.: An ant colony approach to clustering. *Anal. Chim. Acta* **509**(2), 187–195 (2004)
14. Zhang, C., Ouyang, D., Ning, J.: An artificial bee colony approach for clustering. *Expert Syst. Appl.* **37**(7), 4761–4767 (2010)
15. Das, S., Abraham, A., Konar, A.: Automatic clustering using an improved differential evolution algorithm. *IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum.* **38**(1), 218–237 (2008)

Fast Marching Method to Study Traveltime Responses of Three Dimensional Numerical Models of Subsurface

C. Kumbhakar, A. Joshi and Pushpa Kumari

Abstract Fast Marching Method is an efficient numerical technique for estimation and tracking monotonically advancing interfaces such as wavefronts. The method works by addressing the theory of viscosity solutions for Hamilton–Jacobi equations, entropy conditions for propagating interfaces and narrow band technique for recovering shapes from images. The entropy conditions constrain the method to give first arrival wavefronts and make the algorithms based on this method stable in heterogeneous media. The use of narrow band approach makes the method extremely fast in computation. This method has a wide range of applications including computation of wavefronts in seismology, photolithographic developments in microchip manufacturing, problems of search and optimal path planning and many other areas such as graphics and medical imaging. This paper presents Fast Marching Method in context of solving Eikonal equation in three dimensions and its utilization to obtain traveltime responses of a variety of numerical models of subsurface. Study of patterns of such traveltime responses gives a comprehensive idea about anomalous geological structure present in the Earth's subsurface.

Keywords Fast marching method · Seismic tomography · Entropy · Viscosity · Binary tree

1 Introduction

Prediction of seismic ray paths between two points in a medium with lateral velocity variation is required in many application of seismology including body wave tomography, earthquake relocation and migration of reflection data.

C. Kumbhakar (✉) · A. Joshi · P. Kumari

Department of Earth Sciences, Indian Institute of Technology Roorkee, Roorkee, India
e-mail: chinmoy.ism@gmail.com

Accurate prediction of seismic ray paths between two points in heterogeneous media is one of the challenging problems in seismology. The difficulty of computing such ray paths arises from the non-linear relationship between velocity and ray path geometry.

Over the past few decades, the growing demand of accurate and fast computation of seismic ray paths has spawned a number of ray based and grid based techniques. The conventional method for predicting source receiver ray path has been ray tracing [1–3]. In this method trajectory of ray paths between two points are directly computed. This approach is highly accurate and efficient in homogeneous or mildly heterogeneous. However in a heterogeneous medium the ray tracing method often fail to converge to a true ray path and does not give guarantee as to whether ray path is for first arrival seismic energy or for multiples.

A number of methods based on grid base numerical techniques developed addressing the same problems in the early 1990s. Some of these are three dimensional finite difference method [4], two dimensional explicit finite difference method [5] and three dimensional explicit finite difference method [6]. Popovici [6] showed the problem of instability in his method. Schneider [7] reduces the problem of instability and devised a three dimensional robust finite difference technique. However the use of finite difference methods solves the problem of locating ray paths for first arrival seismic energy often in heterogeneous media but again there is no guarantee that the solutions give the first arrivals in a medium with steep velocity gradients [8]. Aside from the problem of computing first arrivals, the problem of stability of algorithms based on conventional numerical schemes in highly heterogeneous media becomes a question [9].

At its core, the problem of locating the first arrival seismic ray path is equivalent to tracking an interface propagating with a velocity normal to itself. Fast marching method originally developed in the field of computational mathematics is highly efficient and accurate technique that deals with the problem of evolution of interfaces in heterogeneous media. Developments of the method can be traced back to the starting work of Sethian [10] on numerical methods for propagating fronts which lead to the development of level set methods [11], and finally to Fast Marching Method [12, 13]. This method is based on construction of entropy-satisfying weak solutions (or viscous limit solution) of Hamilton–Jacobi equation by using numerical schemes borrowed from the techniques of hyperbolic conservation law. Coupling work on entropy and viscosity conditions are used in the method to obtain the weak solutions which allow the algorithms based on this method to remain stable in heterogeneous media by dealing accurately and robustly with all kinds of topological changes in the propagating interfaces. Entropy conditions constrain the method to give first arrival solution and the use of narrow band approach, as given in [9], makes the algorithms extremely fast in computation.

Fast Marching Method has been used extensively in seismic imaging of crust and lithosphere in the recent past [14–17]. Aside from the use in seismic imaging the method has been used in a variety of applications including photolithographic

development [18], Problems of search and optimal path planning [19–21] and graphics and medical imaging [22–24].

The knowledge of approximate model close to the actual model of subsurface of a region is essential in seismic tomography to obtain actual figure of subsurface of that region. Choosing such model of subsurface of a region, particularly in a new region where no work related to subsurface structure has been done is a common problem in seismic tomography. To solve this problem accurately, it is essential to have accurate knowledge of traveltime responses of different models of subsurface. This paper presents the basics of Fast Marching Method and its utilization to obtain the traveltime responses of various numerical models simulating different geological models and their studies.

If it is possible to obtain contours of traveltime from the data recorded in a network of seismographs installed in a region of earth then the contours of different numerical model can be utilised to check for matching with that of observed responses. The matching of the observed response with possible responses of different numerical models may serves as a good idea to choose the approximate model of subsurface.

2 Fast Marching Method

Hamilton–Jacobi equation or Eikonal equation for the propagation of seismic P-waves in an isotropic medium can be written as [25]:

$$|T| = S \quad (1)$$

where, $S(x, y, z)$ is slowness function and $T(x, y, z)$ is a time function called the Eikonal. When T is constant it describes the surface of constant phase (wavefronts). The simplest form of upwind difference scheme for Eq. (1) which takes care of entropy and viscosity is given as [9]:

$$\left[\begin{array}{l} \max(D^{-x}T, -D^{+x}T, 0)^2 \\ +\max(D^{-y}T, -D^{+y}T, 0)^2 \\ +\max(D^{-z}T, -D^{+z}T, 0)^2 \end{array} \right]_{i,j,k}^{1/2} = S_{i,j,k} \quad (2)$$

where, i, j and k denotes the grid points in any orthogonal Cartesian co-ordinate system (x, y, z) and D denotes standard finite difference operators, as for illustration in the x -direction they are:

$$D^{-x}T_{i,j,k} = \frac{T_{i,j,k} - T_{i-1,j,k}}{\Delta x} \quad D^{+x}T_{i,j,k} = \frac{T_{i+1,j,k} - T_{i,j,k}}{\Delta x},$$

where, Δx is the grid spacing in x direction. Equation (2) is upwind scheme [9] in terms that it allows to choose the grid points which respect the direction of flow of information. Intuitively upwind means that if a wavefront progresses from left to

right, one should use a difference scheme that reaches upwind to the left to collect information to construct solutions downwind to the right.

2.1 General Solution of Upwind Difference Scheme

The general solution of Eq. (2) in Cartesian co-ordinates with regular grids can be obtained by considering all the eight quadrants with respect to the grid point where the value is to be computed. The first order general solution to that grid point for traveltime in any quadrant can be written as:

$$T = \frac{b + \sqrt{c}}{a} \quad (3)$$

where,

$$a = \frac{C_x}{\Delta x^2} + \frac{C_y}{\Delta y^2} + \frac{C_z}{\Delta z^2}$$

$$b = \frac{C_x T_x}{\Delta x^2} + \frac{C_y T_y}{\Delta y^2} + \frac{C_z T_z}{\Delta z^2}$$

$$c = S^2 a - \frac{C_x C_y}{\Delta x^2 \Delta y^2} (T_x - T_y)^2 + \frac{C_y C_z}{\Delta y^2 \Delta z^2} (T_y - T_z)^2 + \frac{C_z C_x}{\Delta z^2 \Delta x^2} (T_z - T_x)^2$$

In the above equation C_x , C_y , C_z and T_x , T_y , T_z are the coefficients of the grid points and traveltimes to the grid points respectively, in x , y , and z directions separated by Δx , Δy and Δz with respect to the grid point where value is to be computed. The value of coefficient of any grid is one; if and only if the computed wavefront has already passed through that grid point, and in any other cases it is zero. Since there will be one traveltime value for the unknown grid point computed from one quadrant and there can be a maximum of eight quadrants so there will be a maximum of eight possible solutions for traveltime. Out of all possible solutions the smallest one would be the correct solution for the unknown grid point.

2.2 Implementation of Narrow Band Technique

The upwind difference equation has the property of direction of flow of information or the property of entropy that it always computes the traveltime from smaller value to larger value and never from larger to smaller. To explain this a two dimensional schematic diagram of grid points is shown in Fig. 1, where at certain time of computation the position of a wavefront is shown by dark grey and black grid points in which dark grid point is having minimum traveltime value compared to the traveltime values of all the grid points in the wavefront. All the

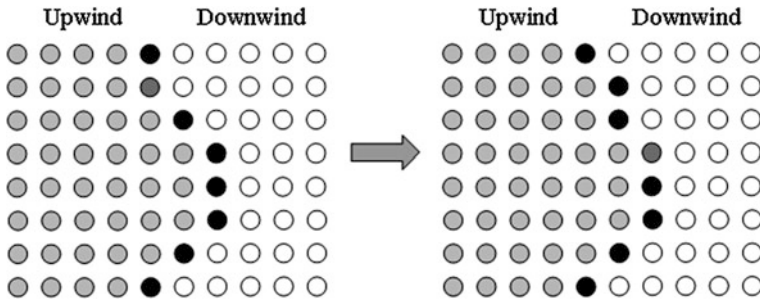


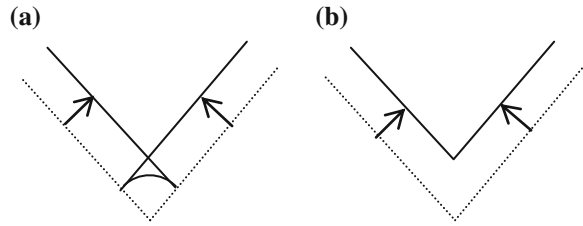
Fig. 1 Narrow band or wavefront evolution technique. *Light gray* grid points are *Alive* points, *black* and *dark grey* grid points are *Close* points in which dark grey grid point is having minimum traveltime and white points are *Far* points. The upwind side contains all the *Alive* grid points and downwind side contains all the *Far* grid points

grid points in the wavefronts are tagged as *Close* points, all the grid points (light grey grid points) through which wavefront has passed are tagged as *Alive* points and all the grid points (white grid points) where traveltime has not been computed are tagged as *Far* points. The narrow band evolved in downwind fashion by finding the *Close* point with minimum traveltime (which is true first-arrival traveltime), tagging it as *Alive*, updating any adjacent *Close* points and computing for the first time any adjacent *Far* points using Eq. (3). If *Far* point is computed it is tagged as *Close*. Using this technique the shape of narrow band approximates the shape of first arrival wavefront. Choosing the *Close* point with minimum traveltime means that causality is satisfied. Use of binary mini heap sorting to locate the global minimum within narrow band decreases the computation time significantly and makes this method very fast. If there are N total grid points in the grids and NB is the maximum number of points in the narrow band, then using the mini heap sorting the scheme solves the problem in $N \log(NB)$ operations [8].

2.3 Role of Entropy in Fast Marching Method

Seismic waves during propagation through heterogeneous media often self-intersect and form swallowtails as shown in Fig. 2, where first arrival wavefronts contain kinks or singular points which are normally spanned by later arriving wavefronts. In such situation the gradient term in the Eikonal equation becomes multi-valued which can lead to the instability of numerical schemes that do not respect the singular points in traveltime fields. To deal accurately and robustly with such situation, it is important to apply a condition in numerical scheme of wave propagation that discards traveltimes from both sides of discontinuities as well later arrival wavefronts in computing future wavefronts. The upwind numerical scheme with entropy conditions serves both the purposes by constructing weak solutions to proceed past the singular points.

Fig. 2 The wavefronts with *dotted line* is initial wavefront with discontinuities. Swallowtail formation is shown by *solid line* (a); first arrival wavefront is shown by *solid line* (b)



The entropy conditions for propagating wavefront may be stated as: once a wavefront passes through a region it will never come into the same region again. Application of such entropy condition decreases the total variation of fronts. A comprehensive analysis of application of such entropy condition can be found in [9]. Since fast marching method uses entropy satisfying upwind difference scheme, hence the method gives first arrival stable solution in heterogeneous media.

3 Numerical Case Studies

The Fast marching method has been used in a variety of numerical models, simulating different artificial geological models, to compute the traveltimes responses. The traveltimes contours for these models are plotted correspondingly to study their patterns. It is to be noted that traveltimes contours were computed keeping source within the symmetric positions of the models for better study however the same kind of study can be done for the source position anywhere within the model. It is also to be noted that although the traveltimes computations were done in three dimensions but for the sake of discussion only two dimensional sections have been plotted.

3.1 Study of Head Waves

The aim of this numerical experiment is to show that beyond the crossover distance the first arrivals are critically refracted head waves and hence to justify that the solutions give the first arrival waves. For this the model parameters are given in Table 1. With these parameters the Fast Marching Method was imposed and the results were obtained. A portion of vertical section of the wavefronts through the source is shown in Fig. 3.

From the considered model parameters, the critical offset is 54.82 km which is marked by yellow line. It can be observed that beyond the critical offset head waves starts generating. The crossover distance from the model parameters is 211.5 km. From the figure it is clear that it is the first offset where the head waves come first. It can also be observed that beyond this distance all are head waves

Table 1 Model parameters to study head waves

Parameters	Values
Model dimension	(500, 20, 350) km
Number of grid points	(251, 11, 176)
Grid spacing	(2, 2, 2) km
Source position	(20, 10, 350) km
Medium velocity	4 km/s in upper layer 7 km/s in lower layer

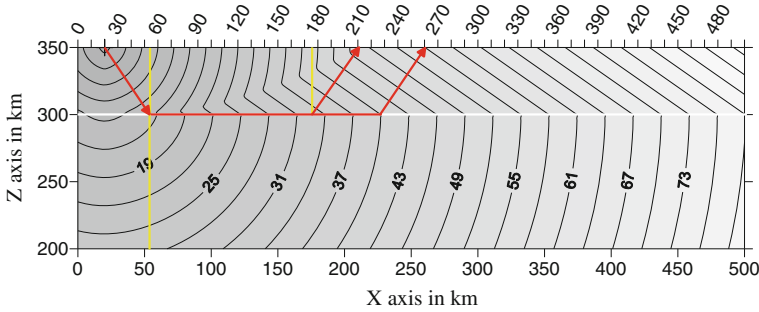


Fig. 3 A portion of *vertical slice* of the wavefronts through the source; illustrating that beyond the crossover distance it is the critically refracted head waves that are the first arrivals. The *red lines* in the figure are the ray paths for critically refracted head waves. The *white line* indicates the boundary separating two media and the *yellow lines* are normals at the boundary at different position shown in the figure. The longest *yellow line* marks the critical offset along X-axis. The time interval between two successive wavefronts is 3 s

which are coming to the surface. This is possible only when the algorithm based on Fast Marching Method gives first arrival solution. This statement is further supported by the fact that all the head waves are parallel and their angle with vertical axis is critical angle which in turn means that, among all the head waves, the critically refracted head waves are coming to the surface. As we know that critically refracted head waves are the first wave among any other head waves so it again justifies that the solution gives first arrivals.

3.2 Study of Anticline and Syncline Folds

In this case the traveltime contours for anticline and syncline folds as shown in Fig. 4 are obtained and discussed. The parameters for the model considered are given in Table 2.

The traveltime contours for the numerical model defined in Table 2 for source positions at (50, 50, 1) km and (100, 50, 1) km are obtained using Fast Marching Method for the same model with and without applying dip in the Y-direction which is the direction of strike. The contours are shown in Fig. 5a–d.

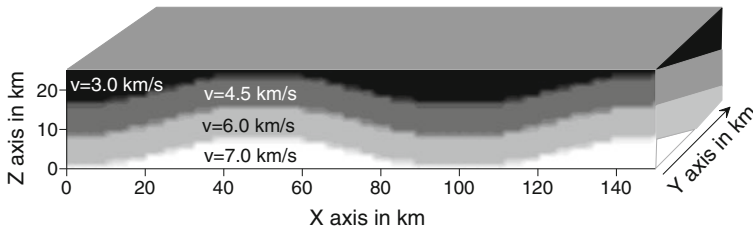


Fig. 4 Anticline and syncline model; in one case the Y-direction is considered without dip and in another case the Y-direction is considered with a dip of 6.7° . Direction of strike is along Y-axis in the model

Table 2 Model parameters to study anticline and syncline folds

Parameters	Values
Model dimension	(150, 100, 25) km
Number of grid points	(151, 101, 26)
Grid spacing	(1, 1, 1) km
Source position	(50, 50, 1) km and (100, 50, 1) km
Medium velocity	As shown in Fig. 3

Figure 5a and b are obtained without applying dip in the Y-direction in the model. Figure 5a is obtained when the source is at (50, 50, 1) km; either when source lies below the anticline structure. Figure 5b is obtained when the source is at (100, 50, 1) km; either when source lies below the syncline structure. From the observation of these two it is clear that contours are elliptical in shape with major axis along strike direction in case of anticline structure and along dip direction in case of syncline structure. Further it can be observed that the spacing between adjacent contours increases in the dip direction in case of anticline structure and decreases in the same direction in case of syncline structure. Thus it is observed that the shape of traveltime contours is different for anticline and syncline folds.

Figure 5c and d are obtained when a dip of amount 6.7° is applied in the Y-direction of the model. Figure 5c is for anticline structure and Fig. 5d is for syncline structure. It is observed from both the figures that due to application of dip in the direction of positive Y-axis the spacing between adjacent contours decreases in the same direction for anticline and syncline structures. Thus it is observed that due to change of understructure the traveltime contours change symmetrically for anticline and syncline folds. The anomalous behavior of contours on the upper portion of Fig. 5c, d is due to appearance of head waves.

3.3 Study of Faults

In this section, we present a numerical study of two types of faults, a normal fault and a Graben-Horst type of faults. The model parameters for model dimensions, number of grid points and grid spacing for both types of faults are same as given in

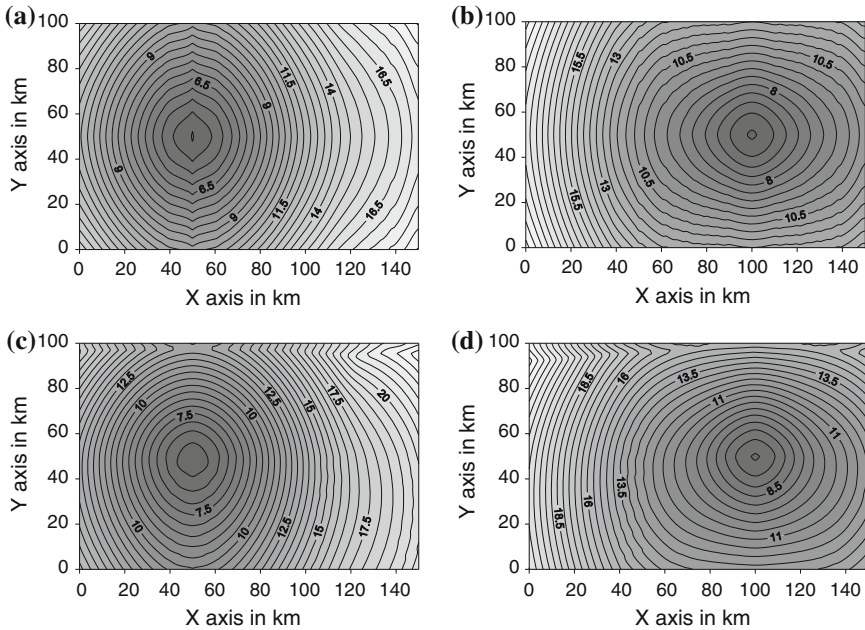


Fig. 5 The computed traveltime contours; **a** when source is at (50, 50, 1) km and **b** when source is at (100, 50, 1) km obtained using the model in Fig. 3 without applying dip in the Y-direction. The computed traveltime contours; **c** when source is at (50, 50, 1) km and **d** when source is at (100, 50, 1) km obtained using the model in Fig. 4 after applying dip of 6.7° in the Y-direction

Table 2. The vertical section of velocity structure for both normal fault and Graben and Horst types of faults are shown in Fig. 6a and c respectively. Traveltime responses for both types of fault are obtained using Fast Marching Method keeping the source at (100, 50, 1) km in case of normal fault and at (75, 50, 1) km in case of Graben-Horst type of fault, as shown in Fig. 6b and d respectively. Contours of traveltimes in Fig. 6b for both normal fault and in Fig. 6d for Graben-Horst type of fault clearly simulate the position of faults considered in the models; hence these kinds of approaches can be utilized to identify the position of real fault of actual earth.

4 Conclusions

The first marching method is an efficient tool to obtain accurate traveltime responses of heterogeneous media. The efficiency lies in the fact that it gives guarantee for first arrival solutions because it addresses the entropy conditions for propagating wavefronts; it is stable because it approximates the discontinuity of traveltime field with weak solution; and it is computationally fast because it utilizes the narrow band technique with heap sorting.

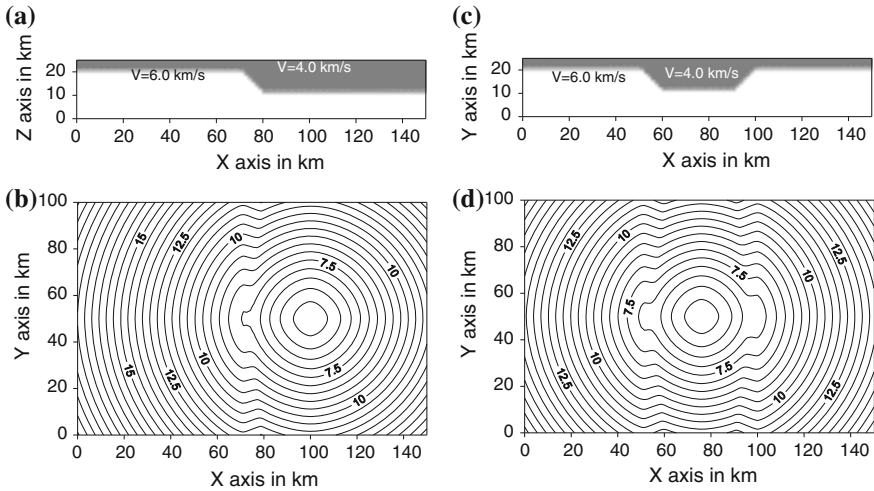


Fig. 6 Vertical section of the normal fault model in the X–Z plane, with X-axis as dip direction (a); contours of traveltime for the normal fault considered in case study (b); vertical section of Horst and Graben type of fault model in the X–Z plane (c); contours of traveltime for Graben and Horst type fault considered in case study (d)

The identification of nature of geological structures of a region is essential in tomographic problems in seismology. This study gives the comprehensive idea about traveltime contours associated with different geological models. The traveltime contours of different geological models can be compared with the traveltime contours derived from traveltime record from a seismic network in a region of earth to understand the approximate nature of geological structure of that region. The head waves which are extremely useful for locating of thin layer are handled by this method which is one of the biggest advantages of this method.

References

1. Julian, B.R., Gubbins, D.: Three-dimensional seismic ray tracing. *J. Geophys.* **43**, 95–113 (1977)
2. Virieux, J., Farra, V.: Ray tracing in 3-D complex isotropic media: an analysis of the problem. *Geophysics* **56**, 2057–2069 (1991)
3. Červený, V.: *Seismic Ray Theory*. Cambridge University Press, Cambridge (2001)
4. Vidale, J.E.: Finite-difference calculations of traveltimes in three dimensions. *Geo-physics* **55**, 521–526 (1990)
5. van Trier, J., Symes, W.: Upwind finite-difference calculation of traveltimes. *Geophysics* **56**, 812–821 (1991)
6. Popovici, A.M.: Finite-difference traveltime maps. *Stanford Explor. Proj. Rep.* **70**, 245–256 (1991)

7. Schneider Jr, W.A.: Robust and efficient upwind finite-difference traveltime calculations in three dimensions. *Geophysics* **60**, 1108–1117 (1995)
8. Rawlinson, N., Hauser, J., Sambridge, M.: Seismic ray tracing and wavefront tracking in laterally heterogeneous media. *Adv. Geophys.* **49**, 203–273 (2008)
9. Sethian, J.A., Popovici, A.M.: 3-D traveltime computation using the fast marching method. *Geophysics* **64**, 516–523 (1999)
10. Sethian, J.A.: Numerical methods for propagating fronts. In: Concus, P., Finn, R. (eds.) *Variational Methods for Free Surface Interfaces*, pp. 155–164. Springer, New York (1987)
11. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys.* **79**, 12–49 (1988)
12. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1591–1595 (1996)
13. Lelièvre, P.G., Farquharson, C.G., Hurich, C.A.: Computing first-arrival seismic traveltimes on unstructured 3D tetrahedral grids using the fast marching method. *Geophys. J. Int.* **184**, 885–896 (2011)
14. Rawlinson, N., Sambridge, M.: The fast marching method: an effective tool for tomographic imaging and tracking multiple phases in complex layered media. *Explor. Geophys.* **36**, 341–350 (2005)
15. Saygin, E., Kennett, B.L.N.: Ambient seismic noise tomography of Australian continent. *Tectonophysics* **481**, 116–125 (2010)
16. Saygin, E., Kennett, B.L.N.: Crustal structure of Australia from ambient seismic noise tomography. *J. Geophys. Res.* (2012). doi:[10.1029/2011JB008403](https://doi.org/10.1029/2011JB008403)
17. Bodin, T., Sambridge, M.: Seismic tomography with the reversible jump algorithm. *Geophys. J. Int.* **178**, 1411–1436 (2009)
18. Sethian, J.A.: Fast-marching level-set methods for three-dimensional photolithography development. In: *Proceedings of SPIE 2726, Optical Microlithography IX* (1996). doi:[10.1117/12.240962](https://doi.org/10.1117/12.240962)
19. Petres, C., Pailhas, Y., Patron, P., Petillot, Y., Evans, J., Lane, D.: Path planning for autonomous underwater vehicles. *IEEE Trans. Robot.* **23**, 331–341 (2007)
20. Garrido, S., Moreno, L., Blanco, D.: Exploration of 2D and 3D environments using Voronoi transform and fast marching method. *J. Intell. Robot. Syst.* **55**, 55–80 (2009)
21. Garrido, S., Moreno, L., Blanco, D., Jurewicz, P.: Path planning for mobile robot navigation using Voronoi diagram and fast marching. *Int. J. Robot. Autom.* **2**, 42–64 (2011)
22. Wen, T., Zhu, Q., Qin, W., Li, L., Yang, F., Xie, Y., Gu, J.: An accurate and effective FMM-based approach for freehand 3D ultrasound reconstruction. *Biomed. Signal Process.* **8**, 645–656 (2013)
23. Wada, F., Hamdi, M.A.: 3D segmentation of intravascular ultrasound images: a fast-marching method. *J. Radiol. Diagn. Imag.* **1**, 29–36 (2013)
24. Yan, J., Zhuang, T., Zhao, B., Schwartz, L.H.: Lymph node segmentation from CT images using fast marching method. *Comput. Med. Imag. Graph.* **28**, 33–38 (2004)
25. Chapman, C.: *Fundamentals of Seismic Wave Propagation*. Cambridge University Press, Cambridge (2004)

A Heuristic for Permutation Flowshop Scheduling to Minimize Makespan

Deepak Gupta, Kewal Krishan Nailwal and Sameer Sharma

Abstract As the problem related to minimizing makespan as objective is NP-hard for more than two machines in flowshop scheduling, therefore need for heuristics have been felt to yield optimal or near optimal solutions in polynomial time. In the present paper, we propose an alternative heuristic algorithm which is compared with the benchmark Palmer's, CDS and NEH algorithm for the processing of n-jobs through m-machines. The proposed heuristic gives solution for solving n-job and m-machine flowshop scheduling problem with minimizing makespan as criteria. Comparisons have been made for tested instances.

Keywords Flowshop · Scheduling · Makespan · Processing

1 Introduction

Scheduling is a decision-making process that is used on a regular basis in many manufacturing and services industries. It deals with the allocation of resources to tasks over given time periods and its goal is to optimize one or more objectives [25]. In the flowshop scheduling problem, n-jobs have to be processed through m-machines and is typically a combinatorial optimization problem with same technological route for each job. In the flow shop problem there are n-factorial

D. Gupta (✉)

M. M. University, Mullana, Ambala, Haryana, India

e-mail: guptadeepak2003@yahoo.co.in

K. K. Nailwal

A. P. J. College of Fine Arts, Jalandhar, Punjab, India

e-mail: kk_nailwal@yahoo.co.in

S. Sharma

D. A. V. College, Jalandhar, Punjab, India

e-mail: samsharma31@yahoo.com

different job sequences possible for each machine, and therefore $(n!)^m$ possible different schedules. It is very difficult to find the effectiveness measure of all the sequences which are $(n!)^m$ (a large number) and select one most suitable sequence which optimizes the required performance measure [1]. The performance measure makespan is defined as the total completion time when the set of all jobs finish processing on all the machines. If a schedule is such that the order in which the jobs go through the first machine is maintained throughout the system is called a permutation schedule and this problem minimizing makespan is denoted by $F_m/prmulC_{max}$.

In the present paper, we present an efficient heuristic for solving n-job, m-machine flowshop scheduling problem with the criteria of minimizing makespan. Our heuristic also improves upon the benchmark heuristic (Palmer, CDS, NEH) in view of the number of alternatives for jobs processing on machines keeping makespan minimum. The remaining paper is organized as follows: in Sect. 2 we provide the review of literature, Sect. 3 gives the assumptions and notations, Sect. 4 explains the proposed heuristic algorithm, Sect. 5 illustrates the numerical, Sect. 6 gives the comparison of the proposed heuristic algorithm with Palmer's, CDS and NEH algorithm and finally, Sect. 7 presents the conclusion.

2 Literature Review

The objective of classical flowshop problems in scheduling is mainly the minimization of makespan. The notation for n-job, m-machine flowshop problem with objective as minimizing makespan is firstly given by Conway et al. [6]. The term permutation flowshop was coined by Pinedo [25]. Thus, the permutation flowshop scheduling problem with makespan as criteria is denoted by $F_m/prmulC_{max}$. With the algorithm of Johnson [15] $F_2/prmulC_{max}$ can be solved in polynomial time. But, it was proved later that $F_m/prmulC_{max}$ is strongly NP-hard for $m > 2$; see e.g. Garey et al. [9]. Also, Johnson algorithm for three machine problem is based on some assumptions made on the processing time of various jobs on machines. Therefore, there became a search by many researcher so as to find the optimal solution for general problem of n-job, m-machine. Palmer [23] proposed a algorithm for n-job, m-machine by calculating job priority function named as slope index to give priorities to the jobs to move from one machine to another and then arranging them in the descending order of their slope index. The literature reveals that Palmer's heuristic approach is based on Page [22] analogy between scheduling and sorting problems. Bonney and Gundry [3] used two slope indices for each job. A simple heuristic was proposed by Campbell et al. [4], which is based on the Johnson algorithm and is popularly known as CDS heuristic. The RA heuristic of Dannenbring [7] gave two neighborhood schemes for improved solutions using Johnson algorithm and attempts to combine the advantages of both Palmer and CDS heuristic. Also, King and Spachis [18] evaluated the performance of some earlier heuristic in literature. An improved functional heuristic was developed by Gupta [12, 13]. The algorithm based on the priority to be given to jobs with large total

processing time was proposed by Nawaz et al. [20] (also known as NEH). Talliard [29] gave 260 problems which are considered the benchmark problems in basic scheduling and NEH heuristic performs well for these problems. Also, Park et al. [24] concluded that NEH outperforms significant heuristics in literature. An excellent review and classification of heuristics for permutation flowshop scheduling with makespan objective can be seen in the paper of Framinan et al. [8]. This paper establishes a general framework in which the existing heuristics can be fitted. Some other reviews include [11, 28]. Chakraborty and Laha [5] gave a modified NEH heuristic algorithm for n-job, m-machine scheduling problem and has the same complexity as that of NEH. Kalczynski and Kamburowski [16] defends that NEH is the best heuristic comparing its results with the metaheuristics. Kalczynski and Kamburowski [17] proposed an improved NEH heuristic based on the concept of Johnson's algorithm. In literature moving from simple heuristics to exact solution include Ignall and Schrage [14] which developed branch and bound technique for minimizing makespan in 3-machine scheduling problem and was independently given by Lomnicki [19]. Also, Bestwick and Hastings [2] gave new bound procedures with the introduction of the concept of dummy operations. The other include tabu search [10, 21], genetic algorithm [27], and ant colony algorithms [26].

3 Assumptions and Notations

The proposed algorithm is based on the assumptions: (1) All the jobs and machines are available at the beginning of the processing. (2) Pre-emption of jobs is not allowed. (3) Machines never breakdown and are available throughout the scheduling process. (4) All processing times of the machines are deterministic, finite and independent of sequence of the jobs to be processed. (5) Each job is processed through each of the machine once and only once. A job is not available to the next machine until and unless processing on the current machine is completed. (6) Setup times are sequence independent and are included in the processing times or otherwise can be ignored. (7) An infinite in-process storage buffer is assumed. If a given job needs an unavailable machine then it joins a queue of unlimited size waiting for that machine.

Given n-jobs $J_i(1 \leq i \leq n)$, these are to be processed on m-machines $M_j(1 \leq j \leq m)$ in the same technological order, which minimizes the makespan C_{\max} . The processing time of the job J_i on the machine M_j is denoted by $a_{i,j}$.

4 Algorithm

The proposed heuristic algorithm is a constructive method which improves performance at each stage, thus building the n-job sequence giving minimum makespan. The proposed heuristic algorithm is explained as:

1. For each job J_i , calculate the total processing time T_i as given by the expression

$$T_i = \sum_{j=1}^m a_{i,j}$$

2. Arrange the jobs in a list according to decreasing values of T_i .
3. If number of jobs is even, go to step 4. and if odd go to step 5.
4. If the number of jobs i.e. $n = \text{even}$ ($2p$ say), then divide the list into two partial sequences having equal number of p -jobs each i.e. $\{J_1, J_2, \dots, J_{p/2}\}$ and $\{J_{p/2+1}, J_{p/2+2}, \dots, J_{2p}\}$. Calculate the value of C_{\max} for all the $\lfloor p \rfloor$ partial sequences of jobs in both the groups. Take the sequence of jobs having minimum value of C_{\max} from both the partial sequences. Now, we have at least two sequences (in case of tie) otherwise one sequence with minimum value of C_{\max} from each of the partial sequences. In case of tie we have to consider both the sequence of jobs. Make all the permutations of jobs with above two partial sequences with the restriction that the relative positions of the jobs in any partial sequence is not altered in any later sequence for a particular group. Take the sequence having minimum value of C_{\max} . This sequence of jobs will have the total minimum makespan.
5. If the number of jobs i.e. $n = \text{odd}$ ($2p + 1$), then divide the list into two partial sequences containing p -jobs each as $\{J_1, J_2, \dots, J_p\}$ and $\{J_{p+2}, J_{p+3}, \dots, J_{2p+1}\}$, leaving the middle job i.e. J_{p+1} . Calculate the value of C_{\max} for all the $\lfloor p \rfloor$ partial sequences of jobs in both the groups. Take the sequence of jobs having minimum value of C_{\max} from both the partial sequences. Now, we have at least two sequences (in case of tie) otherwise one sequence with minimum value of C_{\max} from each of the partial sequences. In case of tie we have to consider both the sequence of jobs. Make all the permutations of jobs with above two partial sequences with the restriction that the relative position of the jobs in any partial sequence is not altered in any later sequence for a particular group. Take the sequence(s) having minimum C_{\max} value. Now, insert the job J_{p+1} in all possible ways in this sequence. Take the sequence(s) having minimum C_{\max} value. This sequence(s) of jobs will have the total minimum makespan.
6. Store the results obtained in step 4 or step 5 according as n is even or odd.

5 Numerical Illustration

To evaluate the proposed algorithm two numerical illustrations are given for n (even and odd).

When n is even: Consider the following 4-job and 5-machine problem. For four jobs, the T_i values are shown in Table 1.

The sequence of jobs in list according to decreasing values of T_i is (J_2, J_3, J_1, J_4) . Divide the list into two groups giving two partial sequences having equal number of jobs namely (J_2, J_3) and (J_1, J_4) . Calculate the value of C_{\max} for each of the partial schedule. Note $C_{\max} = 38$ for the partial schedule (J_3, J_2) in the first group.

Table 1 Calculation for T_i values

Job J_i	Machine M_j					T_i
	M_1	M_2	M_3	M_4	M_5	
J_1	7	5	2	3	9	26
J_2	6	6	4	5	10	31
J_3	5	4	5	6	8	28
J_4	8	3	3	2	6	22

Table 2 Calculation for T_i values

Job J_i	Machine M_j			T_i
	M_1	M_2	M_3	
J_1	8	5	2	15
J_2	10	6	9	25
J_3	6	2	8	16
J_4	7	3	6	16
J_5	10	4	7	21

Similarly, $C_{max} = 32$ for the partial schedule (J_1, J_4) . Now, make all the possible arrangements with partial schedules (J_3, J_2) and (J_1, J_4) keeping the job J_3 always before the job J_2 , job J_1 before the job J_4 i.e. the relative position of jobs remains unaltered for a particular group in the arrangement. The possible arrangements with this restriction will give the six sequences (in case of partial schedule containing two jobs each) namely (J_3, J_2, J_1, J_4) , (J_3, J_1, J_2, J_4) , (J_3, J_1, J_4, J_2) , (J_1, J_4, J_3, J_2) , (J_1, J_3, J_4, J_2) , (J_1, J_3, J_2, J_4) . The sequence (J_1, J_3, J_4, J_2) and (J_1, J_3, J_2, J_4) gives the minimum value of $C_{max} = 51$. Hence, the proposed algorithm yields two alternate schedules with minimum makespan.

When n is odd: Consider the following 5-job and 3-machine problem. For four jobs, the T_i values are shown in Table 2.

The sequence of jobs in list according to decreasing values of T_i is $(J_2, J_5, J_3, J_4, J_1)$ [one may take $(J_2, J_5, J_4, J_3, J_1)$ sequence also as T_i is same for the job J_3 and J_4]. Divide the work content list into two groups giving two partial sequences namely (J_2, J_5) and (J_4, J_1) leaving the middle job J_3 . Calculate the value of C_{max} for each of the partial schedule. Note minimum $C_{max} = 32$ for the partial schedule (J_2, J_5) . Similarly, minimum $C_{max} = 22$ for the partial schedule (J_4, J_1) . Now, make all the possible arrangements with partial schedules (J_2, J_5) and (J_4, J_1) keeping the job J_2 always before the job J_5 , job J_4 before the job J_1 i.e. the relative position of jobs remains unaltered for a particular group in the arrangement. The possible arrangements with this restriction will give the six sequences namely (J_2, J_5, J_4, J_1) , (J_2, J_4, J_5, J_1) , (J_2, J_4, J_1, J_5) , (J_4, J_1, J_2, J_5) , (J_4, J_2, J_1, J_5) , (J_4, J_2, J_5, J_1) . The sequences (J_2, J_5, J_4, J_1) , (J_2, J_4, J_5, J_1) and (J_4, J_2, J_5, J_1) gives the minimum value of $C_{max} = 42$. Now insert the job J_3 at all the possible locations in the partial sequences (J_2, J_5, J_4, J_1) , (J_2, J_4, J_5, J_1) and (J_4, J_2, J_5, J_1) . The job J_3 is permanently placed at the location where it yields minimum value of C_{max} . The sequences obtained with minimum value of $C_{max} = 48$ namely $(J_3, J_2, J_5, J_4, J_1)$,

Table 3 Comparative results

S. No.	No. of jobs (i)	No. of machines (j)	Makespan/No. of alternatives			
			Palmer's heuristic	CDS heuristic	NEH heuristic	Proposed heuristic
1	5	3	48/02	48/01	48/02	48/10
2	4	3	26/02	26/01	26/03	26/03
3	4	4	33/01	31/01	31/01	31/01
4	4	5	53/01	51/01	51/02	51/02
5	7	7	71/01	65/01	65/01	65/02

$(J_2, J_3, J_5, J_4, J_1), (J_2, J_5, J_3, J_4, J_1), (J_2, J_3, J_4, J_5, J_1), (J_2, J_5, J_4, J_3, J_1), (J_2, J_4, J_3, J_5, J_1), (J_2, J_4, J_5, J_3, J_1), (J_3, J_2, J_4, J_5, J_1), (J_3, J_4, J_2, J_5, J_1), (J_4, J_3, J_2, J_5, J_1)$. Hence, the proposed algorithm yields ten alternate schedules with minimum makespan.

6 Comparisons with Benchmark Heuristics

The proposed heuristic is compared with well established heuristics known so far. These heuristic algorithms are: Palmer's, CDS and NEH heuristic. The results obtained with the proposed heuristic and with established heuristics are shown in Table 3.

6.1 Palmer's Heuristic

Historically, the first of heuristic approaches ranking the jobs according to their processing times is due to Palmer [23]. He specified a job priority function named 'slope index'. This heuristic consists of two steps as follows:

Step 1: Compute slope A_i for job J_i as $A_i = - \sum_{j=1}^m \{m - (2j - 1)\} a_{i,j}$

Step 2: Ordering the jobs in the sequence based on descending order of A_i values gives the minimum makespan.

6.2 CDS Heuristic

Campbell, Dudek and Smith (CDS) heuristic algorithm uses Johnson rule for minimizing makespan in a deterministic flowshop problem. The CDS heuristic algorithm simply converts a given n-job, m-machine problem ($m > 2$) into k number of 2-machine, n-job sub-problems, where $k = m - 1$. Each sub-problem

is solved using Johnson rule for C_{max} . The sequence of the sub-problem which yields minimum value of C_{max} after applying Johnson rule is selected for scheduling jobs on the machines.

6.3 NEH Heuristic

Nawaz, Ensfor and Ham (NEH) algorithm constructs the sequence of jobs in iterative manner. The total processing times T_i is calculated using the expression

$$T_i = \sum_{j=1}^m a_{i,j}.$$

The jobs are then arranged in decreasing order of the values of T_i .

Calculate C_{max} for each of the different arrangement of these two jobs. The arrangement of jobs having small value of C_{max} is selected for subsequent iteration. Then, next job from the list is chosen. This job is alternatively placed at all possible locations in the partial sequence. This job is permanently placed at the position where it yields lowest value of C_{max} for the partial schedule. Similarly, next job from the list is chosen, and placed one by one at all possible positions of the partial sequence to find C_{max} value of the partial sequence. This job is placed permanently at the position where partial sequence has minimum C_{max} value. This process is continued till all jobs from the list are covered, thus giving the final sequence of jobs with minimum makespan.

Tested Problems:

(1)

$J_i \setminus M_j$	M_1	M_2	M_3
J_1	8	5	2
J_2	10	6	9
J_3	6	2	8
J_4	7	3	6
J_5	10	4	7

(2)

$J_i \setminus M_j$	M_1	M_2	M_3
J_1	6	5	4
J_2	8	1	4
J_3	3	5	4
J_4	4	4	2

(3)

$J_i \setminus M_j$	M_1	M_2	M_3	M_4
J_1	4	3	7	2
J_2	3	7	8	5
J_3	1	2	4	3
J_4	3	4	3	7

(4)

$J_i \setminus M_j$	M_1	M_2	M_3	M_4	M_5
J_1	7	5	2	3	9
J_2	6	6	4	5	10
J_3	5	4	5	6	8
J_4	8	3	3	2	6

(5)

$J_i \setminus M_j$	M_1	M_2	M_3	M_4	M_5	M_6	M_7
J_1	3	5	7	1	6	9	4
J_2	2	5	8	1	6	7	9
J_3	4	8	1	6	7	9	1
J_4	5	7	6	1	8	4	3
J_5	1	2	8	4	6	7	4
J_6	3	5	4	6	8	1	2
J_7	5	2	8	4	6	3	2

7 Conclusion

It is clear from Table 3 showing comparative results of the proposed heuristic with the benchmark heuristics that the makespan calculated from the proposed heuristic algorithm is either less or equals to the makespan calculated by other heuristic algorithms. Also, the proposed heuristic wins in the number of alternative sequences which optimizes the makespan. Hence, the proposed heuristic algorithm is an efficient alternative to Palmer’s, CDS and NEH heuristic algorithms.

References

1. Baker, K.R.: Introduction to Sequencing and Scheduling. Wiley, New York (1974)
2. Bestwick, P.F., Hastings, N.A.J.: A new bound for machine scheduling. *Oper. Res. Q.* **27**, 479–487 (1976)
3. Bonney, M.C., Gundry, S.W.: Solutions to the constrained flowshop sequencing problem. *Oper. Res. Q.* **27**, 869–883 (1976)
4. Campbell, H.G., Dudek, R.A., Smith, M.L.: A heuristic algorithm for n-jobs, m-machine sequencing problem. *Manage. Sci.* **16**, 630–637 (1970)
5. Chakraborty, U.K., Laha, D.: An improved heuristic for permutation flowshop scheduling. *Int. J. Inf. Commun. Technol.* **1**, 89–97 (2007)
6. Conway, R.W., Maxwell, W.L., Miller, L.W.: Theory of Scheduling. Addison-Wesley, Reading (1967)
7. Dannenbring, D.G.: An evaluation of flowshop sequencing heuristics. *Manage. Sci.* **23**, 1174–1182 (1977)
8. Framinan, J.M., Gupta, J.N.D., Leisten, R.: A review and classification of heuristics for permutation flow-shop scheduling with makespan objective. *J. Oper. Res. Soc.* **55**, 1243–1255 (2004)
9. Garey, M.R.D., Johnson, D.S., Sethi, R.: The complexity of flow shop and job shop scheduling. *Math. Oper. Res.* **1**, 117–129 (1976)
10. Grabowski, J., Wodecki, M.: A very fast tabu search algorithm for the permutation flow shop problem with makespan criterion. *Comput. Oper. Res.* **31**, 1891–1909 (2004)
11. Graham, R.L., Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G.: Optimisation and approximation in deterministic sequencing and scheduling: a survey. *Ann. Discret. Math.* **5**, 287–326 (1979)
12. Gupta, J.N.D.: A functional heuristic for the flow-shop scheduling problem. *Oper. Res. Q.* **22**, 39–47 (1971)
13. Gupta, J.N.D.: A heuristic algorithm for the flow shop scheduling problem. *R.A.I.R.O. Rech. Operationnelle* **10**, 63–73 (1976)
14. Ignall, E., Schrage, L.: Application of the branch and bound technique to some flow shop scheduling problems. *Oper. Res.* **13**, 400–412 (1965)
15. Johnson, S.M.: Optimal two- and three-stage production schedules with setup times included. *Naval Res Logistics Q.* **1**, 61–68 (1954)
16. Kalczynski, P.J., Kamburowski, J.: On NEH heuristic for minimizing the makespan in permutation flowshops. *OMEGA, Int. J. Manage. Sci.* **35**, 53–60 (2007)
17. Kalczynski, P.J., Kamburowski, J.: An improved NEH heuristic to minimize makespan in permutation flow shops. *Comput. Oper. Res.* **35**, 3001–3008 (2008)
18. King, J.R., Spachis, A.S.: Heuristics for flowshop scheduling. *Int. J. Prod. Res.* **18**, 345–357 (1980)
19. Lomnicki, Z.A.: A branch and bound algorithm for the exact solution to some flow shop scheduling problems. *Oper. Res. Q.* **16**, 89–100 (1965)
20. Nawaz, M., Ensore Jr., E.E., Ham, L.: A heuristic algorithm for the m machine, n job flow shop sequencing problem. *OMEGA, Int. J. Manage. Sci.* **11**, 91–95 (1983)
21. Nowicki, E., Smutnicki, C.: A fast tabu search algorithm for the permutation flow-shop problem. *Eur. J. Oper. Res.* **91**, 160–175 (1996)
22. Page, E.S.: An approach to the scheduling of jobs on machines. *J. Roy. Stat. Soc.* **23**, 484–492 (1961)
23. Palmer, D.S.: Sequencing jobs through a multi-stage process in the minimum total time-A quick method of obtaining a near optimum. *Oper. Res. Q.* **16**, 101–107 (1965)
24. Park, Y.B., Pegden, C.D., Ensore, E.E.: A survey and evaluation of static flowshop scheduling heuristics. *Int. J. Prod. Res.* **22**, 127–141 (1984)
25. Pinedo, M.: Scheduling: Theory, Algorithms, and Systems. Prentice-Hall, Upper Saddle (2002)

26. Rajendran, C., Ziegler, H.: Ant-colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. *Eur. J. Oper. Res.* **155**, 426–438 (2004)
27. Reeves, C.R.: A genetic algorithm for flowshop sequencing. *Comput. Oper. Res.* **22**, 5–13 (1995)
28. Ruiz, R., Maroto, C.: A comprehensive review and evaluation of permutation flow shop heuristics. *Eur. J. Oper. Res.* **165**, 479–494 (2005)
29. Taillard, E.: Benchmarks for basic scheduling problems. *Eur. J. Oper. Res.* **64**, 278–285 (1993)

A Survey on Web Information Retrieval Inside Fuzzy Framework

Shruti Kohli and Ankit Gupta

Abstract With the emergence of web as one of the primary mode of information sharing and searching, it is a challenge posed to the researchers and developers to design the information retrieval system which can effectively and efficiently returns the query result as per user's requirement. This survey paper tends to find out some challenges posed by information retrieval and how the concept of fuzzy helps to solve those challenges.

Keywords Fuzzy logic · Web intelligence · Information retrieval

1 Introduction

Since its inception in early 1960s, Internet has come down a long way. What started as a defense research project now has become a necessity of life. After the introduction of world wide web (popularly known as WWW) by its creator Tim Berner Lee in 1990s, it has dominated the Internet. The WWW has revolutionized the way we gather, process and use information. It has reforms the way how we handle daily aspects of our life like business, education, commerce etc. [1]. Now a days, the indexed web contains approximately 12.6 billion web pages with close to 3.5 billion internet users [2, 3]. In 2009, more than 1.6 trillion searches were made by the user globally. According to a report by Mckinsey and company, overall search value while using internet is estimated at around \$780 billion [4].

S. Kohli (✉) · A. Gupta

Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India
e-mail: kohli.shruti@gmail.com

A. Gupta

e-mail: theankit@yahoo.com

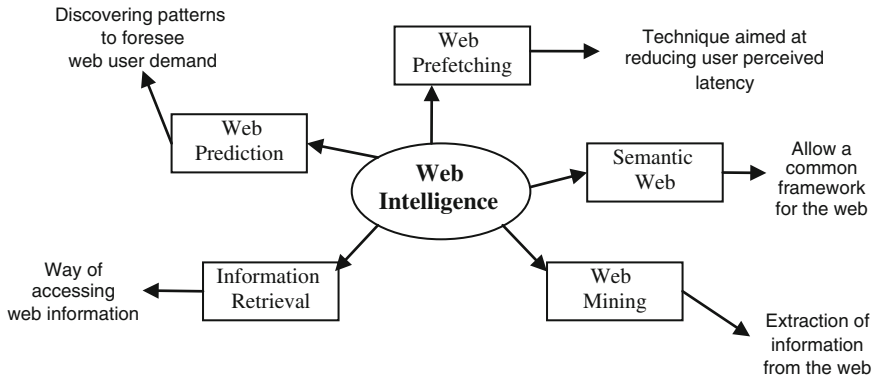


Fig. 1 Web intelligence

In the year 2000, a new term Web Intelligence was coined by Zhong et al. [1] who described Web Intelligence as a “*new direction for scientific research and development to explore the fundamental roles as well as practical impacts of Artificial Intelligence and advanced information technology on the next generation of web empowered products systems and services and activities*”. Some key domains of Web intelligence are web prediction, Web mining, Web prefetching, Semantic web, Information retrieval etc. (Fig. 1). These five domains of web intelligence, although responsible for different areas of functioning, work together for a better overall experience for the end user.

Some survey papers, [5–9], have already been presented in the past for web mining and web prediction domain but they had very little reference of Information Retrieval domain of Web intelligence. Kobayashi and Takeda [8] has written excellently on Information retrieval on the web but their work was mainly focused on growth of internet and different technologies used for Information Retrieval besides this, the paper was published when the web was in its primitive stage. This survey paper has been written while keeping in mind the Information Retrieval domain of Web Intelligence.

The rest of the paper is organized as follows: Sect. 2 of this paper describes the basic fundamentals, challenges of Information retrieval, Sect. 3 describes different soft computing methodologies used for information retrieval, Sect. 4 describes different fuzzy logic techniques used for effective and smooth retrieval of information and Sect. 5 concludes this paper.

2 Information Retrieval (IR)

Information Retrieval is a way to access information available on the internet, intranet, databases and data repositories. An Information Retrieval system takes a user’s query as input and returns a set of documents sorted by their relevance to the

query. IR system are usually based on the segmentation of documents and queries into index terms and their relevance is computed according to the index terms they have in common as well as according to other information such as the characteristics of the documents(e.g. number of words, hyperlinks between papers or bibliographic reference) or some probabilistic information [10].

2.1 Information Retrieval: An Introduction

An Information Retrieval model typically consists of four small module [7, 8]:

1. *Indexing*: Described as “*Generation of Document representation*”. An Index term is a collection of terms with pointers to places where information about the document can be found. Four approaches to indexing documents on the web are (i) Human or manual indexing, (ii) Automatic Indexing, (iii) Intelligent or agent based indexing, (iv) Metadata, RDF and Annotation based indexing
2. *Querying*: Expression of user preferences through natural language or terms connected by logical operators
3. *Evaluation*: Performance of matching between user query and document representation,
4. *User Profile construction*: Storage of terms representing user preferences, specially to enhance the system retrieval in future accesses by the user.

Formally, an Information Retrieval system based on the conventional fuzzy set model can be defined as quadruple $\langle T, Q, D, F \rangle$ [11] where,

1-T represents set of index terms which are used to represent queries and documents,

2-Q is a set of queries that can be recognized by the system. Each query $q \in Q$ is a legitimate Boolean expression composed of index terms and the logical operators AND, OR and NOT.

3-D is a set of document, $D = \{d_1, d_2, \dots, d_n\}$. Each document $d_i \in D$ is represented by $((t_1, e_{i1}), (t_2, e_{i2}), \dots, (t_m, e_{im}))$ where e_{iji} denotes the weight of term t_j in document d_i and may take any value between 0 and 1, $1 \leq i \leq n$, and $1 \leq j \leq m$. These degree of strength e_{ij} of term t_j in document d_i are determined either subjectively, by the author of the document, or objectively, by some algorithmic procedure [12].

4-F is a retrieval function

$$F : D \times Q \rightarrow [0, 1],$$

which assigns to each pair (d, q) a number in the closed interval $[0, 1]$. This number is a measure of similarity between the document d and the query q and is called the document value for document d with respect to query q .

A fundamental aim of any information retrieval system is to search the relevant information from a pool of knowledge while trying to reject any non relevant information as possible. There are two basic measures of efficiency of any Information retrieval as follows:

$$\text{Recall} = \frac{\text{retrieved relevant object}}{\text{total number of relevant objects}} \quad (1)$$

$$\text{Precision} = \frac{\text{retrieved relevant object}}{\text{total number of retrieved objects}} \quad (2)$$

There are two types of Information Retrieval systems proposed in the literature, (i) Classical and (ii) Web Based. Following section will give a short difference between the two.

2.2 Difference Between Classical and Web Based IR

Classical methodology of IR basically deals with the offline data of various entities like companies, organizations etc. The advantages here are that the most of the users of this data are familiar with the systems as most of the users are more or less a part of the concerned entities. So most of the IR systems are based on this approach of previous knowledge of expected query types and indexing types and so it is relatively easy to develop models keeping above things in mind.

Problem with web based IR system is that they mostly deals with online data retrieval and the major users of this system are those who are neither much familiar of the data nor the system. Challenge lie here to depict the expected query type from the inexperienced, unfamiliar users and then map it to a particular document index.

Much research has been done to develop a classical IR system based on Fuzzy Logic but the Research to develop Web based IR system using fuzzy is still in early stage. This paper will try to deal with some major research work related to Web based IR system using Fuzzy.

2.3 Challenges of Web Based Information Retrieval System

Unlike Classical Information Retrieval System, a web based Information System faces a wide variety of challenges. These challenges can be broadly categorized into three basic domains as follows:

- (1) Data Domain: The data domain signifies the location and format characteristics of the data. Data is distributed, heterogeneous, semi-structured and time varying.

- (2) Information Domain: This domain specifies the nature of the data and the corresponding information. Data is imprecise, incorrect, inconsistent and uncertain.
- (3) User Domain: This domain characterizes the problems that arise because of the involvement of the users with the system. Most of the users are unfamiliar with the system, their information needs change frequently and most of their search queries are based on personalized goals.

Lucarella [13], pointed out that imprecise and uncertain information comes from three major aspects in an IR system environment including the representation of users, queries, the representation of documents and the relevance relationship between user's queries and documents.

Keeping above things in mind, the IR systems are developed around the concept of relevance [14].

3 Information Retrieval and Soft Computing

There are various Artificial Intelligence methodologies which work together to achieve the consolidated goal of Web Intelligence. A domain of AI is Soft computing having different areas like Fuzzy Logic, Artificial Neural Networks, Rough Sets, Genetic Algorithm. These techniques are considered more flexible and less computationally demanding than the other Artificial Intelligence techniques. According to [15], guiding principle of Soft Computing is *"to exploit the tolerance for imprecision, uncertainty, partial truth, and approximation to achieve tractability, robustness, low solution cost and better rapport with reality"*.

Fuzzy sets provide a natural framework for the process in dealing with uncertainty. Artificial Neural Networks are widely used for modeling complex functions, and provide learning and generalization capabilities. Genetic algorithms are efficient search and optimization tools. Rough sets help in granular and knowledge discovery [6].

Apart from the above soft computing tools, Natural language processing and Swarm Intelligence are relatively new areas where research is going for smooth information retrieval. NLP is an area of AI to gather the knowledge related to human behavior and understanding and then map these understandings to develop new tools and techniques to make computer systems to act accordingly. Swarm intelligence is the study of collective behavior of individuals working in a group in a decentralized but self-organized environment.

The main objective of this paper is to provide an outline of Information Retrieval and how various Fuzzy techniques can be utilized to solve the different problems associated with the efficient retrieval of information.

4 Fuzzy Methodologies for Information Retrieval

Fuzzy logic has been extensively used in the domain of Information retrieval from the early days of Networking. It has become more relevant in today's scenario because of its inherent ability to solve the complex problems faced by exponentially increasing size of web and web information. While the last decade witnessed the progress in the field of Information Retrieval in form of various retrieval mechanisms introduced for a better and effective query-document mapping targeted for static and offline data repository like company databases, institutional databases etc., It has turned its core focus towards online/offline, dynamic and fast growing contents. The challenges we have discussed in Sect. 2.3 clearly demonstrate that the today's information retrieval system needs to evolve and follow different strategies to cope up with these newly and primitive challenges. The concept of globalization has forced the corporate as well as academic world to go global and so is their data and thus prompting them to convert their offline static data into online dynamic one.

Keeping these things in mind, this paper intends to find out different methodologies of Information Retrieval paradigm currently in practice. This paper has mainly taken into account the advances introduced in the literature in this field in the last 20 years. Based on the data available we can divide last 20 years into 3 parts (i) Time ranging from 1991 to 2000, (ii) time ranging from 2001 to 2005, (iii) Time ranging from 2006 to 2012. These methodologies have been discussed in subsequent sections.

4.1 Fuzzy Concept Network

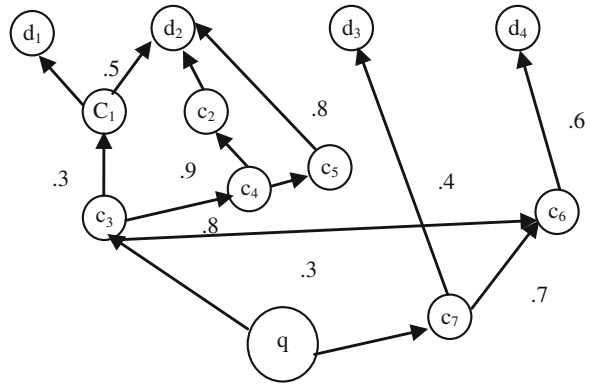
The term "Concept" refers to as a process of identification of the important domain terms of the document. It might be considered as an extension of indexing process. One of the earliest work describing the "concept" was proposed in [16] where the author tried to extract pertinent noun groups from the documents.

This "concept" was further extended and proposed as Fuzzy Concept Network in [17] for information retrieval from a knowledge base. A Fuzzy Concept Network consists of nodes and directed links (Fig. 2). These nodes represent concepts frames while the directed links represent relationships between concepts semantically. The nodes can represent a concept or document. Real value μ is associated with the directed links, where $\mu \in [0,1]$. This real value μ reflects the strength of the semantic association between the nodes.

With the help of this model and multi valued fuzzy logic (Modus Ponens and Modus Tollens), authors presented FIRST: Fuzzy Information Retrieval System.

A fuzzy relations between concept network was proposed in [18], as links between the concepts. Basically four types of relationships were proposed namely, Fuzzy Positive Association, Fuzzy Negative Association, Fuzzy Generalization and Fuzzy Specialization. Each member of a relation has a numeric value between 1 and 0 associated with it that express the strength of the relationships.

Fig. 2 A fuzzy concept network



This model not only takes the directly linked concepts but traverses one or more link to find related concepts. Besides above four relationships, another relationship name “context relationship” was proposed between a concept representing a dialogue context and the dependent relationship between two concepts.

Extended Fuzzy Concept Network was proposed in [19], consisting of nodes and directed links. Each directed link connects two concepts or connects from a concept c_i to a document d_j , based on the following condition:

$$c_i \xrightarrow{(\mu,A)} X \tag{3}$$

where, $A \in \{P, N, G, S\}$ and $X \in \{c_j, d_j\}$ and $\mu \in [0, 1]$.

A total of 8 Combinations of directed links are possible.

Every directed link in an extended fuzzy concept network is labeled with a pair of values (μ,FR) , where μ denotes the degree of relevance and FR denotes the fuzzy relationships between the concepts and concept/document.

A more comprehensive and flexible approach was presented in [20] as *Multi-relationship Fuzzy Concept Network* where instead of using single fuzzy association between concepts and concepts/documents, a multi-relationship fuzzy association was presented. For this approach, Eq. 3 is modified as (μ,A) is replaced with the quadruple $(\langle \mu_P,P \rangle, \langle \mu_N,N \rangle, \langle \mu_G,G \rangle, \langle \mu_S,S \rangle)$, where $\mu_P, \mu_N, \mu_G, \mu_S$ describes the fuzzy association between these concepts thus giving a very high level of membership among index terms and the query keyword. A pictorial representation of the evolution of fuzzy concept network is given in Fig. 3.

4.2 Formal Concept Analysis (FCA)

Formal concept analysis has been introduced by Wille [21] in 1982 for analyzing and structuring a domain of interest. Formal concept analysis is a formal technique for data analysis and knowledge representation. It defines formal contexts to

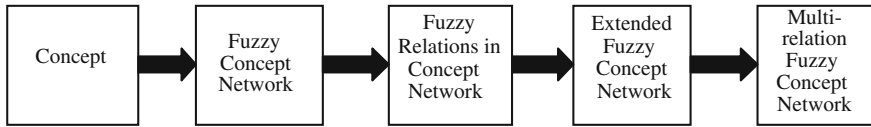


Fig. 3 Evolution of fuzzy concept network

represent relationships between objects and attributes in a domain. From here, FCA can then generate formal concepts and interpret the corresponding concept lattice making information retrieval much easier. Subsequent part gives some basics of the formal concept.

Formal Context: A formal context is a triple (O,A,R) , where O and A are two sets of elements called *object* and *attributes*, respectively, and R is a binary relation between O and A .

Formal Concept: Given a context (O,A,R) , Let E,I be two sets such that $E \subseteq O$ and $I \subseteq A$. Then consider the dual set E' and I' , i.e. the sets defined by the attributes applying to all the objects belonging to E and the objects having all the attributes belonging to I , respectively, i.e.:

$$E' = \{a \in A | oRa \forall o \in E\}$$

$$I' = \{o \in O | oRa \forall a \in I\}$$

A formal concept of the context (O,A,R) is a pair (E,I) such that $E \subseteq O$ and $I \subseteq A$ and the following conditions hold:

$$E' = I, \quad I' = E.$$

The set E and I represents the concept extensional and intensional components, respectively and are referred to as *extent* and the *intent* of the concept, respectively.

The extension covers all objects belonging to concept, while the intension comprises all attributes valid for all those objects [22]. The term ‘object’ and ‘attributes’ also referred as ‘documents’ and ‘terms’. This mapping makes it easy to use concepts of FCA in information retrieval.

During the year 2001–2006, Formal concept analysis was extensively used to improve retrieval of information. Formal concept analysis was used to develop domain Ontologies with the help of similarity graph. Formica [23] successfully explained the concept of mapping the concept lattice to similarity graph. The author successfully established that the similar attributes are more important than the common objects helping greater ontology integration (Fig. 4).

4.3 Fuzzy Formal Concept Analysis (FFCA)

Fuzzy formal concepts analysis is a generalization of Formal Concept Analysis for modeling uncertainty information [24]. FFCA can support ontology construction when some information is more relevant than other data or semantic web search

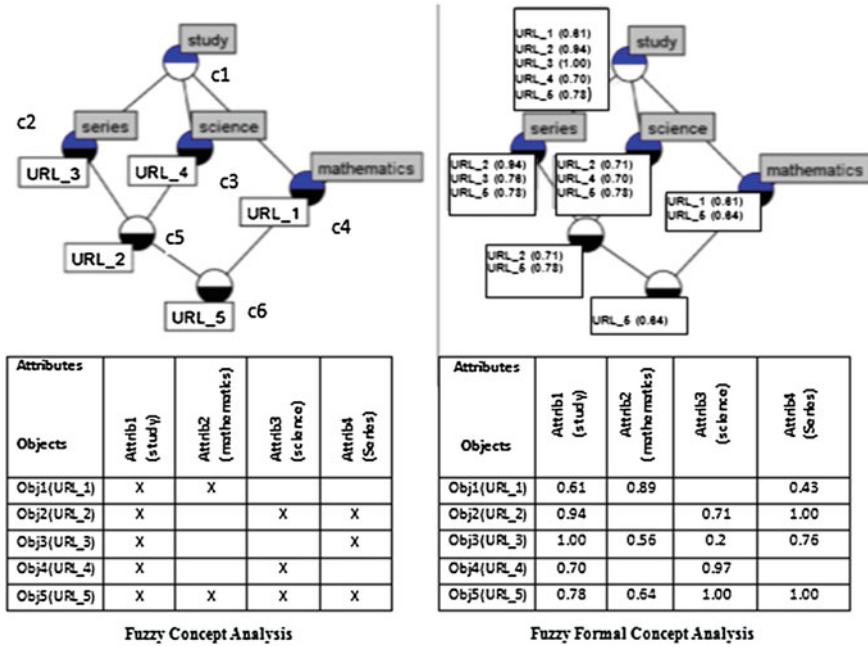


Fig. 4 FCA versus FFCA

when the user is not sure about what he/she is looking for. FFCA provides a mathematical framework which can support the construction of formal ontologies in the presence of uncertain data for the development of the semantic web. FFCA was proposed by Tho et al. [25], where they combine fuzzy logic and FCA to propose FFCA where uncertainty information is directly represented by a real number of membership value in the range of [0,1]. After the induction of FFCA, the research domain was diverted towards FFCA.

In one the most recent work with respect to FFCA, Maio et al. [26], used the FFCA for a better ontology based retrieval mechanism supporting data organization and visualization and proposed a better navigation model.

The major problem with this approach is the size of the concept lattice. As the attributes and the objects grows, the size of the lattice also grows almost exponentially, specially while dealing with the enormous web data. Although some work is going on like by Singh et al. [27] but this is area of research is still in its infant state.

4.4 Operators

While the Formal Concept Analysis and Fuzzy Formal Concept Analysis approach uses the concept of lattice for better mapping of keyword with index terms, an operator uses the mathematical concept of mapping values from one vector space

Fig. 5 Earliest fuzzy operators

$$(x, \dots, y) \rightarrow \text{MIN}(x, \dots, y) \quad (x, \dots, y) \rightarrow \text{MAX}(x, \dots, y)$$

to another. Unarguably, MAX and MIN operators were the two earliest fuzzy operators proposed by Zadeh [28]. While the MIN operator, which maps a group of given arguments to its minimum value, used for “ANDING” of terms, MAX operator, which maps a group of given arguments to their maximum value, was used for “ORING” of the terms. With the passage of time, these operators were considered inappropriate as the level of complexity were increasing. After these operators many other fuzzy operators were introduced for effective information retrieval.

$$(x, \dots, y) \rightarrow \text{MIN}(x, \dots, y)$$

Kim et al. [11], investigated the various behavioral aspects of T-operators and concluded that Zadeh’s basic fuzzy operators MAX and MIN are inappropriate for a model of Information retrieval system.

Fuzzy operators can be classified into two broad categories namely (i) T-operators and (ii) Averaging operators. While t-norms and t-conorm dominated the era ranging from 1973 to 1988, averaging operators dominated the research in fuzzy Information Retrieval afterwards after its introduction by Yager [29] in 1988. Introduction of averaging operator helped in inclusion of more effective multi criteria decision making in the domain of Information retrieval. Table 1 list the summary of some of the operators useful in Fuzzy Information Retrieval.

Smolikova et al. [30] compared the various aggregation operators for selecting applicants for a Ph.D programs and concluded that the various averaging operators have their own domain of discourse and they work best when they are used in the environment for which they are developed.

A detailed theory of T-operators is given in [31] and theory of weighted averaging operators is given in [21, 32].

4.5 Rough Set

Rough set theory [33] is an extension of set theory for data analysis in the presence of inexact, uncertain or vague information. The combination of rough set theory and FCA provides an interesting framework for semantic web development, for the definition of hybrid similarity measures for ontology mapping, alignment and integration etc. [34]. Although this area of research is sometimes used in conjunction with FCA, it independently have very little relevance towards Information Retrieval approach, so we will not discuss this topic in detail.

Table 1 Fuzzy operators

Sr. No	T-Operators	Year	Averaging operators	Year
1	Zadeh [28]	1973	Ordered weighted averaging operator [29]	1988
2	Lukasiewicz logic [35]	1976	Infinite-one operator [36]	1990
3	Probabilistic operators [37]	1980	Quasi-arithmetic means [38]	1999
4	Yager [39]	1980	Ordered weighted geometric operator [40]	2000
5	Dombi [41]	1980	Ordered weighted geometric averaging operator [42]	2002
6	Weber [43]	1983	Induced ordered weighted geometric operator [44]	2004
7	Hamacher [43]	1983	Geometric mean averaging operator [45]	2005
8	Dubois [46]	1986	Weighted power mean averaging operator [47]	2007

5 Conclusion

Information retrieval is one of the most important domain of web intelligence. Most of the million dollar E-Commerce organization and social media sites heavily depend on the effective and efficient information retrieval for their revenue generation. In other words this domain is responsible for trillions of dollars in world economics. Fuzzy is playing an important role in making this task simpler. The inherent power of Fuzzy logic with a high level of flexibility makes it possible to design and develop today's sophisticated Information Retrieval system which can effectively match with users ever growing demands. This paper is an endeavor to explore and discuss the contribution of Fuzzy logic for solving information retrieval problems that are getting complex with increase in the number of WebPages and searches online. Not only Search Engine's like Google or yahoo but a common search functionality of a e-commerce portal like www.amazon.com, www.slideworld.com or social networking websites like face book keeps on improving their systems as per user perspective to increase its future usage with aim—"Improved and Intelligent Information Retrieval".

References

1. Zhong, N., Jiming, L., Yao, Y.Y., Ohsuga, S.: Web intelligence. In: Proceedings of 24th Annual International Computer Software and Application Conference, COMPSAC (2000)
2. <http://www.worldwidewebsite.com/>. Retrieved on 18 Jan 2013
3. <http://www.internetworldstats.com/stats.htm>. Retrieved on 18 Jan 2013
4. Impact of Internet Technologies: Search. Mckinsey & Company, New York, July (2011)

5. Arotaritei, D., Mitra, S.: Web mining: a survey in the fuzzy framework. *Fuzzy Sets Syst.* **148**, 5–19 (2004)
6. Pal, S.K., Talwar, V., Mitra, P.: Web mining in soft computing framework: relevance, state of the art and future directions. *IEEE Trans. Neural Netw.* **13**(5), 1163–1177 (2002)
7. Mitra, S., Pal, S.K., Mitra, P.: Data mining in soft computing framework: a survey. *IEEE Trans. Neural Netw.* **13**(1) 3–14 (2002)
8. Kobayashi, M., Takeda, K.: Information retrieval on the web. *ACM Comput. Surv.* **32**(2) (2000)
9. Domenech, J., de la Ossa, B., Sahuquillo, J., Gil, J.A., Pont, A.: A taxonomy of web prediction algorithm. *Expert Syst. Appl.* **39**, 8496–8502 (2012)
10. Ferrandez, A.: Lexical and syntactical knowledge for information retrieval. *Inf. Process. Manage.* **47**, 692–705 (2011)
11. Kim M.H., Lee, J.H., Lee, Y.J.: Analysis of fuzzy operators for high quality information retrieval. *Inf. Process. Lett.* **46**, 251–256 (1993)
12. Klir, G.J., Yuan, B.: *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Upper Saddle River, NJ (1995)
13. Lucarella, D.: Uncertainty in information retrieval: an approach based on fuzzy sets. In: proceedings of 9th Annual International Phoenix Conference on Computer and Communication, pp. 809–814. Arizona, USA (1990)
14. Salton, S.: *Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computers*. Addison-Wesley, Reading, MA (1989)
15. Zadeh, L.A.: What is soft computing. *Soft Comput.* **1**(1), 1 (1997)
16. Bruandet, M.F.: Outline of a knowledge base model for an intelligent information retrieval system. *Inf. Process. Manage.* **25**(1), 89–115 (1989)
17. Lucarella, D., Morara, R.: First: fuzzy information retrieval system. *J. Inf. Sci.* **17**, 81–91 (1991)
18. Kracker, M.: A fuzzy concept network model and its applications. In: Proceedings 1st IEEE Conference Fuzzy System, pp. 761–768 (1992)
19. Chen, S., Hong, Y.J.: Fuzzy Query processing for document retrieval based on extended fuzzy concept network. *IEEE Trans. Syst. Man Cybern. Part-B* **29**, 96–104 (1999)
20. Chen, S.M., Horng, Y., Lee, C.H.: Fuzzy information retrieval based on multi relationship fuzzy concept networks. *Fuzzy Sets Syst.* **140**, 183–205 (2003)
21. Herrera, F., Viedma, E.H., Chiclana, F.: A study of the origin and uses of the ordered weighted geometric operator in multicriteria decision making. *Int. J. Intell. Syst.* **18**, 689–707 (2003)
22. Ganter, B., Wille, R.: *Formal Concept Analysis: Mathematical Foundation*. Springer, Berlin (1999)
23. Formica, A.: Ontology based concept similarity in formal concept analysis. *Inf. Sci.* **176**, 2624–2641 (2006)
24. Belohlavek, R., Outrata, J., Vychodil, V.: Fast factorization by similarity of fuzzy concept lattice with hedges. *Int. J. Found. Comput. Sci.* **19**(2), 255–269 (2008)
25. Tho, Q.T., Hui, S.C., Fong, A.C.M., Cao, T.H.: Automatic fuzzy ontology generation for semantic web. *IEEE Trans. Knowl. Data Eng.* **18**(6), 842–856 (2006)
26. Maio, C.D., Fenza, G., Loia, V., Senatore, S.: Hierarchical web resources retrieval by exploiting fuzzy formal concept analysis. *Inf. Process. Manage.* **48**, 399–418 (2012)
27. Singh, P.K., Choudhary, A.K.: A method for decomposition of fuzzy formal context. In: International Conference on Modeling Optimization and Computing (ICMOS)-2012 in Procedia Engineering, vol. 38 pp. 1852–1857 (2012)
28. Zadeh, L.A.: Outline of a new approach to the analysis of complex systems and decision process. *IEEE Trans. Syst. Man Cybern.* **3**, 28–44 (1973)
29. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Trans. Syst. Man Cybern.* **18**(1), 183–190 (1988)
30. Smolikova, R., Wachowiak, M.P.: Aggregation operators for selection problem. *Fuzzy Sets Syst.* **131**, 23–34 (2002)

31. Gupta, M.M., Qi, J.: Theory of T-norms and fuzzy inference methods. *Fuzzy Sets Syst.* **40**, 431–450 (1991)
32. Yager, R.R., Kacprzyk, J., Beliakov, G.: Recent developments in the ordered weighted averaging operators: theory and practice. *Stud. Fuzziness Soft Comput.* **205** (2011)
33. Pawlak, Z.: Rough sets. *Int. J. Inf. Comput. Sci.* **11**(5), 341–356 (1982)
34. Zhao, Y., Halang, W., Wang, X.: Rough ontology mapping in e-business integration. *Stud. Comput. Intell.* **37**, 75–93 (2007)
35. Giles, R.: Lukasiewicz logic and fuzzy set theory. *Int. J. Man Mach. Stud.* **8**, 313–327 (1976)
36. Smith, M.E.: Aspects of the P-norm model of information retrieval: synthetic query generation, efficiency and theoretical properties. Ph.D. Dissertation, Cornell University (1990)
37. Bandler, W., Kohout, L.: Fuzzy power set and fuzzy implication operators. *Fuzzy Sets Syst.* **4**, 13–30 (1980)
38. Marichal, J.L.: Aggregation operators for multicriteria decision aid. Ph.D. Dissertation, University De Liege (1999)
39. Yager, R.R.: On a general class of fuzzy operators. *Fuzzy Sets Syst.* **4**, 235–242 (1980)
40. Chiclana, F., Herrera, F., Herrera-Viedma, E.: The ordered weighted geometric operator: properties and application. In: *Proceedings of the 8th International Conference on Information Processing and Management of Uncertainty in Knowledge Based Systems*, pp. 985–991. Madrid, Spain (2000)
41. Dombi, J.: A general class of fuzzy connectives. *Fuzzy Sets Syst.* **4**, 235–242 (1980)
42. Xu, Z.S., Da, Q.L.: The ordered weighted geometric averaging operators. *Int. J. Intell. Syst.* **17**, 709–716 (2002)
43. Weber, S.: A general concept of fuzzy connectives, negation and implications based on t-norms and t-conorms. *Fuzzy Sets Syst.* **11**, 115–134 (1983)
44. Chiclana, F., Herrera-Viedma, E., Herrera, F., Alonso, S.: Induced ordered weighted geometric operators and their use in the aggregation of multiplicative preference relations. *Int. J. Intell. Syst.* **19**, 233–255 (2004)
45. Chen, S.J., Chen, S.M.: Fuzzy information retrieval based on geometric mean averaging operators. *Int. J. Comput. Math. Appl.* **49**, 1213–1231 (2005)
46. Dubois, D., Prade, H.: New results about properties and semantics of fuzzy set-theoretic operators. *Fuzzy Sets*, Plenum Press, New York 59–75 (1986)
47. Hong, W.S., Chen, S.J., Wang, L.H., Chen, S.M.: A new approach for fuzzy information retrieval based on weighted power mean averaging operators. *Comput. Math. Appl.* **53**, 1800–1819 (2007)
48. Birkoff, G.: *Lattice Theory*. American Mathematical Society, Providence RI (1967)

Squaring Back off Based Media Access Control for Vehicular Ad-hoc Networks

Kamal Kant Sharma, Mukul Aggarwal and Neha Yadav

Abstract Inter-vehicle communications have great potential to induce great interest in research and industry. Vehicular ad hoc networks (VANETs) may significantly improve passenger safety and comfort. The deployment of VANETs is a challenging task in weakly interconnected and in highly overloaded networks both. A good back off technique can reduce a large number of collisions in the MAC layer I VANET. This will reduce the collision probability and hence increases the utilization of network resources. A uniform random distribution has been employed to choose the back off value in the Binary Exponential Back off (BEB) technique used in the IEEE 802.11 MAC protocol. This random choosing VANETs leads to unnecessary idle times and reduced throughput. This paper proposes a new back off technique called “Squaring Back off (SB)” in which the differences between the consecutive contention window sizes are reduced to a negligible value. Here, the value of the back off timer is based on the size of the contention window. The size of the contention window is varied in accordance with the result of the previous transmission. A successful transmission reduces the size of the contention window whereas a failure leads to the increase in the contention window size. Simulation results indicate that the proposed technique provides better throughputs and less idle times than the logarithmic and Fibonacci based techniques when used in a mobile ad-hoc environment. Squaring back off based media access control can prove very useful for vehicular ad-hoc networks.

Keywords VANET · Contention window · IEEE 802.11 · Medium access control · Vehicular mobile ad-hoc networks

K. K. Sharma (✉) · M. Aggarwal · N. Yadav
Krishna Institute of Engineering and Technology, Ghaziabad, UP, India
e-mail: knmietkamal@gmail.com

M. Aggarwal
e-mail: mukul.digital@gmail.com

N. Yadav
e-mail: nehayadav1508@gmail.com

1 Introduction

Wireless technology first came into existence in 1901, when G Marconi successfully transmitted radio signals across the Atlantic Ocean [1]. Since their emergence, wireless networks have become increasingly popular in the computer research enabling the enhancements in mobility, portability and affordability for most of the systems seen in use today. VANETs are inherited by MANETs. A Mobile Adhoc Network (MANET) is a collection of mobile nodes dynamically forming a temporary network without the use of any existing network infrastructure [2]. MANETs utilize multi-hop radio relaying to find the connectivity between the different mobile nodes. MANETs have received significant attention in recent years due to their easiness to set up and usage in many domains. Such networks can be very useful in situations where there is limited time and resources. MANETs find their usage in military applications where setting up a fixed infrastructure for communication among a group of soldiers in enemy territories may not be possible.

A **Vehicular Ad-Hoc Network** or **VANET** is a technology that uses moving cars as nodes in a network to create a mobile network. VANET turns every participating car into a wireless router or node, allowing cars approximately 100–300 m of each other to connect and, in turn, create a network with a wide range [1]. The shared media used by wireless networks helps a node to transmit a packet. Access to this media is controlled by the Media Access Control (MAC) protocol. MAC protocol of IEEE 802.11 uses Distributed Coordination Function (DCF) for contention based distributed access to the channel [3].

The Back off mechanism is a basic part of a MAC protocol. Since only one transmitting node uses the channel at any given time, the MAC protocol must suspend other nodes while the media is busy. In order to decide the length of node suspension, a back off timer is installed in the MAC protocol. The choice of back off mechanism should consider generating back off timers which allow adequate time for current transmissions to finish and avoid unintended idle time that leads to redundant delay in the network.

IEEE 802.11 is the most popular WLAN standard that defines the specification for the physical and MAC layers. In the IEEE 802.11 standard MAC protocol, the truncated binary exponential back off (BEB) technique is used [4]. Here, the initial contention window (CW) size is set to a random value between (0, CW min). Each time a collision occurs, the CW doubles its size up to a maximum of CW max. Thus at high load, the CW size is high and therefore the resolution power of the system is high. At low loads, small CW ensures low access delay. The term 'truncated' has been specified since after a certain number of increases, the exponentiation stops off suddenly at that level. BEB has the disadvantage of the channel capture effect and the instability.

An efficient back off algorithm should meet at least three requirements. A back off algorithm should maximize the total throughput of the network, minimize the delay of transmission, and finally, maintain a fair usage of the network among the

transmitting nodes. The Back off algorithms has been classified into two main categories; static and dynamic back off algorithms.

In static back off algorithms, the back off period will be of the form: Back off Timer = I, an integer. The value of I can be carefully chosen depending on many factors; such as the number of nodes in the network, having a fixed value can work under a certain scenario for a specific network topology. In the case of MANETs, the major challenges would be mobility and dynamic topology, i.e. positions of nodes within the network area.

In dynamic back off algorithms, back off periods are changed depending on many factors. The most common factor used is the result of last attempt of transmission by the node requesting channel access. In general, dynamic back off algorithms deploy a customized version of the general formula. The input of the formula is the current size of CW and the result of this formula is the new size of CW (CW new).

CW new is limited between a maximum value and a minimum value referred to as CW max and CW min, respectively. CW new is used to randomly choose the value of back off timer according to formula:

$CW_{new} = \text{Min}(f(CW), CW_{max})$, after successful transmission

$\text{Max}(g(CW), CW_{min})$, after a collision

Back off Timer = b, where b is a random integer and $CW_{min} < b < CW_{max}$.

Two important aspects of the back off algorithm that need to be taken care are as follows. Firstly, the increment behavior needs to be examined. The method used by the back off mechanism to increase CW size directly affects the balance between reducing the number of attempts to access the channel and reducing channel idle time. Successful collision avoidance will only be possible if adequate time is allowed between any two consecutive attempts to access the channel.

On the other hand, a back off algorithm should avoid unnecessarily long back off periods. Imposing a long back off period on a node is directly related to network idle time since the traffic flowing over the network is often unpredictable. Secondly, the decrement behavior after successful transmissions is also a major factor that needs to be explored. The back off algorithm has to decide the reaction of a successful transmission since this decision affects the chances of nodes winning the next contention over the network. Balance should be maintained between extremely long and extremely short new values of CW. Moreover, resetting the counters to an initial value after a successful transmission has been proved undesirable; a node that has successfully transmitted a message has a small window size afterwards. Therefore, this node generates smaller back off values leading to a higher possibility of winning the next contention over the channel.

The standard BEB technique implemented in IEEE 802.11 network protocols fail to achieve the best network throughput and have caused long delays over the network [5]. Fibonacci Increment Back off (FIB) and Logarithmic approaches achieves Higher Throughput than BEB but suffer from unintended idles times [6, 7].

In this paper, we present a new back off technique, referred to as the SB technique that can overcome the limitations of the existing back off techniques. The SB technique provides better throughput and less idle times as per the simulation results. The rest of the paper is organized as follows. Section 2 presents the new SB algorithm. Section 3 presents the simulation model. Section 4 then analyses the performance results. Lastly, Sect. 5 presents the conclusion.

2 The Proposed Technique

Back off algorithms have been used for the collision avoidance and to increase the utilization of network resources [8]. Changing the size of the CW has a greater impact on the performance of the algorithm employed. During a failure to transmit, increasing the CW rapidly increases large CW to even further larger sizes. Reaching such large window sizes decreases the expected wait time for a given node to access to the shared medium. Moreover, a large window size tends to contribute to increasing channel idle times, leading to a major waste in the shared channel bandwidth. Therefore, we have proposed a new technique to avoid the above mentioned limitations.

In our proposed algorithm, we have used Square series as the new size of CW, leading to reducing the increment factor when more transmission failures take place and hence introducing smaller increment on large window sizes. The squaring can be done as follows:

$$SQ(n) = n \times n; \quad n = 2, 3, 4 \quad (1)$$

The squaring series can be obtained by calculating the ratio between every two successive terms in the above formula as follows:

$$SQ(n) = SQ(n)/SQ(n - 1); \quad n = 2, 3, 4 \quad (2)$$

The incremental behavior of the proposed approach with the other two approaches has been discussed here. The three different increment formulas are Squaring, logarithmic and Fibonacci. Figure 1 shows the behavior of the three increment formulas. The size of CW, measured in time slots, is plotted against number of iterations. The iteration number is the number of consecutive transmission failures. As the number of iterations (Contention failures) increase, the squaring increment is the largest of all the other increments. The contention window is increased by larger sizes leading to higher throughput in squaring back off technique.

Algorithm: Squaring back off Nomenclature: {BO = back off time, CW = contention window, DIFS = DCF inter frame space, DCF = distributed coordination function}

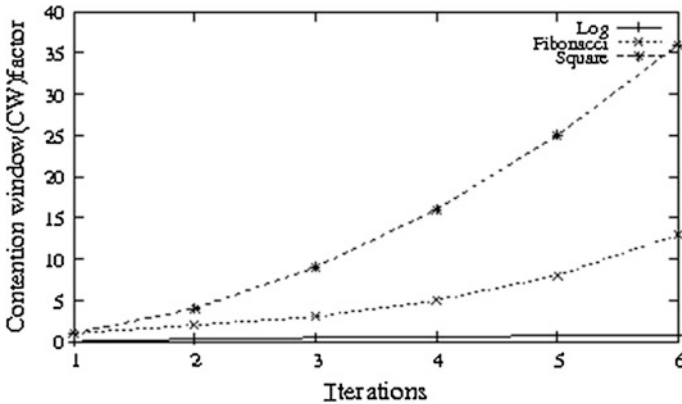


Fig. 1 Contention window factor versus iterations squaring back off algorithm

Step 0: Set BO to initial value

Step 1: While Bo \neq 0 do

 For each time slot

 If (channel is idle) then

 BO = BO - 1;

 If (channel is idle for more than DIFS) then

 Send;

 If (Send failure) then

 CW = Next Square series number;

 BO (i + 1) = square (i); (1 < i < CW - 1) Else

 CW = Initial Value;

 BO = 0; GO to step 1;

Step 2: Stop.

Figure 2 illustrates the properties of the newly proposed squaring series and the existing Fibonacci series. After a certain number of terms, the Fibonacci series ratio converges to a limit of 1.618 whereas the square series is formed by the ratio which almost reaches a value of approximately one which leads to reduced idle times.

Contention window in case of first failure can reach up to 4 for the square series leading to increased throughput whereas it is only 2 for the Fibonacci series that is only half of the value of that of square series. In square series, after a certain number of terms the ratio tends to decrease to approximately 1 whereas in Fibonacci series, the ratio tends to converge into a limit of $(1 + \sqrt{5})/2 \approx 1.6$. This leads to unnecessary idle times in Fibonacci back off technique that has been avoided in the squaring back off technique.

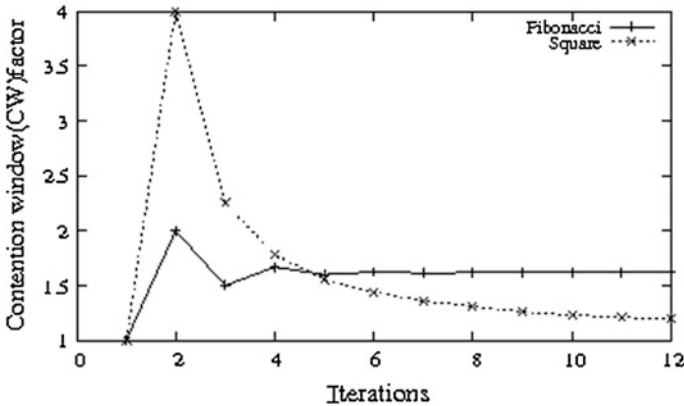


Fig. 2 Ratio of successive terms versus iterations

Table 1 Simulation parameters

Parameter	Value
Nodes	20, 40, ..., 100
Area	1,000 * 1,000 m
Speed	2, 4, ..., 20 m/s
Packet size	512 bytes
Simulation time	800 s
Transmission range	200 m
Pause time	0 s
Probability of broadcast	0.50

3 Simulation

The proposed algorithm has been evaluated using the C++ language with the help of a practical traffic model. The standard MAC protocol has been changed to implement the modified back off algorithm for vehicular ad hoc network. The routing protocol used is Ad-hoc On-Demand Distance Vector (AODV) protocol by taking different values of probability of broadcasting the messages. Every node in the network has the same probability of broadcasting the messages.

By making the variations of the parameters in the network such as the total number of nodes and the speed, results have been verified thoroughly. Simulations have been carried out for the networks having a total number of nodes varying between 20 and 100 mobile nodes and the speed values ranging from 2 to 18 m/s. Other simulation parameters include the area of 1,000 * 1,000 m, nodes transmission range of 200 m, simulation time of 800 s and the traffic generated is the constant bit rate (CBR) traffic. All the simulation parameters are given in Table 1.

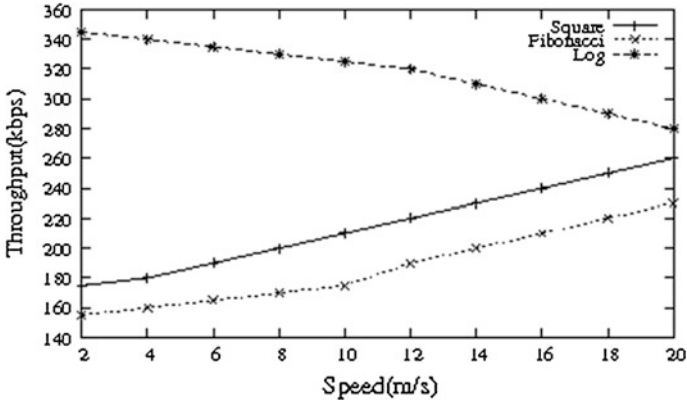


Fig. 3 Throughput versus speed (for 20 nodes)

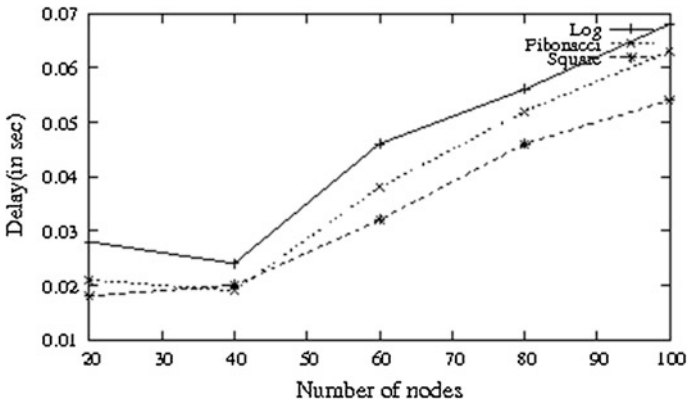


Fig. 4 Delay versus number of nodes (speed of 10 m/s)

4 Results

As shown in the Figs. 3, 4 and 5, the newly proposed squaring algorithm has improved the total throughput of the network. Also the channel idle times have been reduced to make the best use of the bandwidth provided. When the number of nodes is increased, the contention is higher to gain access to the channel. Due to the reduced increment in the window size, a larger size of data was successfully received by the nodes.

Both the square series and the Fibonacci series improve the throughput with the increase in speed, whereas the logarithmic series has reduced throughput (see Fig. 3). As the speed is increased, even though there are some problems occurring, the data packets reach the destination at a faster rate allowing the next transmission

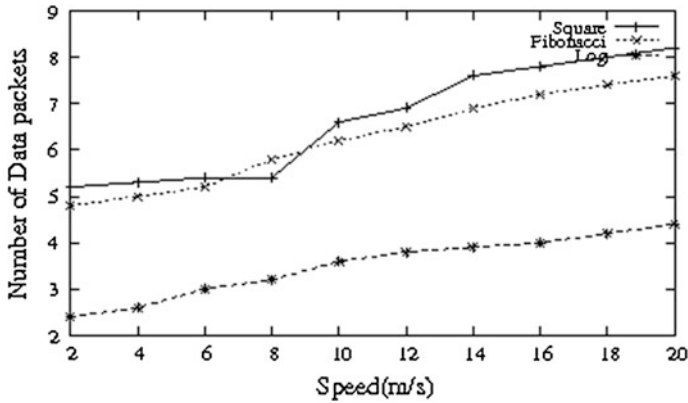


Fig. 5 Number of data packets versus speed (for 100 nodes)

to occur. As the number of transmissions is increased, the number of collisions occurring is also more. Once the collision occurs, the contention window size increases frequently. Thus, the square series has better throughput than the Fibonacci series.

The square series has the least delay when compared to the other approaches. When the number of nodes is increasing, the average time taken for the data packets to reach from the source to the destination keeps on increasing (see Fig. 4). The square series consumes the least time of all approaches.

According to the Fig. 5, the number of data packets lost is comparably less in the squaring series and thus the number of packets received is more as the speed increases.

5 Conclusion

The Binary Exponential Back off (BEB) algorithm used by the IEEE 802.11 MAC protocol uses uniform random distribution to choose the back off value. In this paper, we have proposed a new back off approach called “Squaring Back off” algorithm to reduce the increment factor for large contention window sizes. Results from simulations have demonstrated that the proposed algorithm has increased the total throughput and reduced the average idle times, thereby making the best use of the bandwidth and other network resources.

References

1. Gabriel, S., Melhem, R., Mosse, D.: BLAM: An energy aware MAC layer enhancement for wireless adhoc networks. In: Proceedings of WCNC04, Atlanta (2004)
2. Ko, Y., Vaidya, N.H.: Location-aided routing (LAR) in mobile ad-hoc networks. *Wirel. Netw.* **6**(4), 307–321, (2000)
3. Deng, J., Varshney, P.K., Haas, Z.J.: A new backoff algorithm for the IEEE 802.11 distributed coordination function. In: Proceedings of CNDS 04, San Diego, CA, 18–21 Jan 2004
4. IEEE Computer Society LAN MAN Standards Committee.: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications. In: ANSI/IEEE Std. 802.11, 1999 edn. The Institute of Electrical and Electronic Engineers, New York (1999)
5. Manaseer, S., Masadeh, M.: Pessimistic backoff for mobile ad hoc networks. In: ICIT2009 (2009)
6. Manaseer, S., Ould-Khaoua, M., Mackenzie, L.M.: Fibonacci backoff algorithm for mobile ad hoc network. In: PGNET (2006)
7. Hu, J., Raymond, C.D.: A statistics based design of MAC protocols with distributed collision resolution for ad-hoc networks. In: Proceedings of MobiWac05, Hawaii, June 2005
8. Bani Yassein, M., Ould-Khaoua, M., Papanastasiou, S.: On the performance of probabilistic flooding in mobile ad hoc networks. In: Proceedings of International Workshop on Performance Modelling in Wired, Wireless, Mobile Networking and Computing in Conjunction with 11th International Conference on Parallel and Distributed Systems (ICPADS-2005), IEEE Computer Society Press, Fukuoka, 20–22 Jul 2005
9. Bononi, L., Conti, M., Donatiello, L.: A distributed mechanism for power saving in IEEE 802.11 wireless LANs. *ACM MONET J.* **6**(3), 211–222 (2001)
10. Divecha, B., Abraham, A., Grosan, C., Sanyal, S.: Impact of node mobility on MANET routing protocols models. *J. Digit. Inf. Manage.* **5**(1), 19 (2007)
11. Taifour, M., Nait-Abdesselam, F., Simplot-Ryl, D.: Neighbourhood backoff algorithm for optimizing bandwidth in single hop wireless ad-hoc networks. In: Proceedings of MobiWac05, Hawaii, June 2005
12. Romaszko, S., Blondia, C.: Enhancements of the IEEE 802.11, a MAC protocol for ad hoc network with history of power adjustment. In: Proceedings of MobiWac05, Hawaii, June 2005
13. Aad, I., Ni, Q., Barakat, C., Turletti, T.: Enhancing IEEE 802.11 MAC in congested environments. In: Proceedings of IEEE ASWN, Boston, Aug 2004
14. Bianchi, G., Fatta, L., Oliveri, M.: Performance evaluation and enhancement of the CSMA/CA MAC protocol for 802.11 wire-less LANs. In: Proceedings of IEEE PIMRC, Taipei, Taiwan, pp. 407–411, Oct 1996

A Study on Expressiveness of a Class of Array Token Petri Nets

T. Kamaraj, D. Lalitha and D. G. Thomas

Abstract Adjunct Array Token Petri Net structure (AATPNS) to generate rectangular pictures has been defined in Lalitha et al (Indian J. Math. Math. Sci. 8(1):11–19, 2012) [7]. AATPNS with inhibitor arcs generated context free and context sensitive Kolam Array languages and Tabled 0L/1L languages. In this paper we study the expressiveness of this model by comparing with some other interesting array generating grammar devices like Pure 2D context free grammars with regular control, Regional tile rewriting Grammars, Prusa Grammars and also comparing with local languages.

Keywords Array token Petri nets · Adjunction · Pure 2D grammars · Regional tile grammars · Prusa grammars · Local languages

1 Introduction

Since seventies, the study of two dimensional languages generated by Grammars or recognized by Automata have been found in the theory of formal languages with the insight of computations in picture processing and pattern recognition [5, 12, 13]. In the quest of syntactic techniques for generation of digital picture patterns, a number of 2D Grammars have been proposed. Siromoney Matrix grammars [17],

T. Kamaraj (✉) · D. Lalitha

Department of Mathematics, Sathyabama University, Chennai 600119, India
e-mail: Kamaraj_mx@yahoo.co.in

D. Lalitha

e-mail: lalkrish_24@yahoo.co.in

D. G. Thomas

Department of Mathematics, Madras Christian College, Chennai 600059, India
e-mail: dgthomasccc@yahoo.com

Kolam Array Grammars (KAG) [15, 16], Tabled 0L/1L grammars (TOLG/TILG) [14] are some of the classical formalisms. Pure 2D context free grammars with regular control (RP2DCFG) [18], Prusa grammars (PG) [11], Regional Tile Rewriting grammars [10] are some of the recent and more expressive grammars. Tiling systems [3, 5] is a recognizing device for a ground level class of array languages REC, which involves the projection of the languages belonging to the class of local languages (LOC). Mutual relationship between these new formalisms and also with LOC is analysed in [2].

Recently another picture generating mechanism, Array token Petri Net structure ATPNS [8], has been evolved from string generating Petri nets [1, 6]. Petri Net [9] is one of the formal models used for analyzing systems that are concurrent, distributed and parallel. Tokens are the elements used to simulate the dynamism of the Petri Net systems. The language generated by the Petri net is set of all feasible transition sequence in that net. In ATPNS model, the authors have used arrays as tokens in some of the places as initial configuration called marking and catenation rules as labels of transitions. The language generated is the set of all arrays created at final places of the net. This model along with a control feature called inhibitor arcs generate the same family of languages as generated by KAG, TOLG and P2DCFG with regular control. To increase the generative capacity of this model, adjunction rules are introduced and Adjunct Array Token Petri net systems (AATPNS) [7] is defined. This new model generates Table 1L languages and strictly included ATPNS family of Languages.

Since AATPNS has been only compared with classical formalisms, we try to study the comparison of this model with recent generating devices and also with LOC for the expressiveness with respect to its generating capacity.

This paper is organized in the following manner. In Sect. 2, basic definitions of various array grammars, Petri Nets and notions of Petri nets pertaining to arrays have been recalled. In Sect. 3, we recall the definition of adjunct array token Petri nets, in more generalized form and provide some illustrative examples. In Sect. 4, we compare this model with various classes of picture languages generated by recent grammars and also with class LOC, for the understanding of generative capacity of this model.

2 Preliminaries

The following definitions and notations are mainly from [5, 8, 10, 18].

2.1 Array Grammars

Definition 1 Let J^{**} denotes the set of all arrays (pictures) over the elements of a finite set J and J^{++} denotes set of all non empty arrays over J . For $l, m \geq 0$, $J^{(l, m)}$

represents the set of arrays of size (l, m) . An array (picture) language is a subset of J^{**} . If $p \in J^{**}$, then $p(i, j)$ denotes a pixel in the place of i th row and j th column of p . $|p|_{col}$ denotes the number of columns of p and $|p|_{row}$ denotes the number of rows of p . $A \oplus B$ denotes a column catenation of the array A with array B which is defined when A and B have same number of rows. $A \circ B$ denotes row catenation of A with B provided the number of columns of A and B are same. If $x \in J^{**}$ then $(x)^n$ (resp. $(x)_m$) denotes horizontal (resp. vertical) juxtaposition of m copies of x .

Pure 2D context free grammars (P2DCFG) which make use of only terminal symbols have been recently studied [18]. Pure 2D context-free grammars, unlike Siromoney matrix grammars [17], admit rewriting any row/column of pictures with no priority of columns and rows. Row/column sub-arrays of pictures are rewritten in parallel by equal length strings and by using only terminal symbols, as in a pure string grammar.

Definition 2 A pure 2D context-free grammar (P2DCFG) is a 4-tuple $G = (\Sigma, P_c, P_r, \mathcal{M}_0)$, where Σ is a set of symbols, $P_c = \{t_{ci}/1 \leq i \leq m\}$, $P_r = \{t_{rj}/1 \leq j \leq n\}$.

Each t_{ci} ($1 \leq i \leq m$), called a column table, is a set of context free rules of the form $a \rightarrow \alpha$, $a \in \Sigma$, $\alpha \in \Sigma^*$ such that any two rules of the form $a \rightarrow \alpha$, $b \rightarrow \beta$ in t_{ci} , have $|\alpha| = |\beta|$ where $|\alpha|$ denotes the length of α .

Each t_{rj} ($1 \leq j \leq n$), called a row table, is a set of context free rules of the form $c \rightarrow \gamma^T$, $c \in \Sigma$, $\gamma \in \Sigma^*$ such that any two rules of the form $c \rightarrow \gamma^T$, $d \rightarrow \delta^T$ in t_{rj} , have $|\gamma| = |\delta|$.

$M_0 \subseteq \Sigma^{**} - \{\lambda\}$ is a finite set of axiom arrays.

Derivations are defined as follows. For any two arrays M_1, M_2 , $M_1 \Rightarrow M_2$ denotes that M_2 is obtained from M_1 by either rewriting a column of M_1 by rules of a column table t_{ci} in P_c or a row of M_1 by rules of a row table t_{rj} in P_r . \Rightarrow^* is the reflexive transitive closure of \Rightarrow .

The picture language $L(G)$ generated by G is the set of rectangular picture arrays $\{M/M_0 \Rightarrow^* M \in \Sigma^{**}, \text{ for some } M_0 \in \mathcal{M}_0\}$.

The family of picture array languages generated by pure 2D context-free grammars is denoted by P2DCFL.

Definition 3 A pure 2D context-free grammar with a regular control (RP2DCFG) is $G_c = (G, Lab(G), C)$, where G is a pure 2D context-free grammar, $Lab(G)$ is a set of labels of the tables of G and $C \subseteq Lab(G)^*$ is a regular string language. The words of $Lab(G)^*$ are called control words of G . Derivations $M_1 \xRightarrow{w} M_2$ in G_c are done as in G , except that if $w \in Lab(G)^*$ and $w = l_1 l_2 \dots l_m$, then the tables of rules with labels l_1, l_2, \dots, l_m are successively applied starting from M_1 to yield M_2 . The picture array language generated by G_c consists of all picture arrays obtained from the axiom array of G_c with derivations controlled as described above. (R)P2DCFL denotes the family of all picture array languages generated by pure 2D context-free grammars with a regular control.

The family P2DCFL is strictly included in family RP2DCFL [18].

Tiling Systems, a recognizing device for the class REC, uses Local languages as projections. A local language is defined by a set of 2×2 arrays (tiles).

Definition 4 Let J be a finite alphabet, θ be a finite set of tiles over $J \cup \{\#\}$. The local language defined by $L = \{p \in J^{++} \mid [[p]] \subseteq \Theta\}$ where $[[p]]$ denotes the set of all tiles contained in p surrounded by a special symbol $\# \notin J$. The family of all local languages is denoted by LOC.

Prusa Grammar device admits parallel application of rules so that non terminal symbols can be substituted with rectangular subpictures simultaneously. This model has more generative power than context free KAG [11].

Definition 5 Prusa Grammar (PG) is a tuple (J, N, R, S) , where J is the finite set of terminal symbols, disjoint from the set N of non terminal symbols; $S \in N$ is the start symbol; and $R \subseteq N \times (N \cup J)^{++}$ is the set of rules.

Let $G = (J, N, R, S)$ be a PG. We define a picture language $L(G, A)$ over J for every $A \in N$. The definition is given by the following recursive descriptions:

1. if $A \rightarrow w$ is in R , and $w \in \Sigma^{++}$, then $w \in L(G, A)$;
2. let $A \rightarrow w$ be a production in R , $w = (N \cup \Sigma)^{(m,n)}$, for some $m, n \geq 1$, and $p_{i,j}$ with $1 \leq i \leq m, 1 \leq j \leq n$, be pictures such that:
 - (a) if $w(i, j) \in \Sigma$, then $p_{i,j} = w(i, j)$;
 - (b) if $w(i, j) \in N$, then $p_{i,j} \in L(G, w(i, j))$;
 - (c) if $P_k = p_{k,1} \oplus p_{k,2} \oplus \dots \oplus p_{k,n}$, for any $1 \leq i \leq m, 1 \leq j \leq n$, $|p_{i,j}|_{col} = |p_{i+1,j}|_{col}$ and $P = P_1 \ominus P_2 \ominus \dots \ominus P_m$; then $P \in L(G, A)$.

The set $L(G, A)$ contains exactly the pictures that can be obtained by applying a finite sequence of rules (1) and (2). The language $L(G)$ generated by grammar G is denoted as $L(G, S)$.

Tile Grammars (TG) [4] perform an isometric derivation process for which homogeneous subpictures are replaced with isometric pictures of the local language defined by the right part of the rules.

Definition 6 A tile grammar (TG) is a tuple (J, N, S, R) , where J is the terminal alphabet, N is a set of non terminal symbols, $S \in N$ is the starting symbol, R is a set of rules. Let $A \in N$. There are two kinds of rules:

1. fixed size: $A \rightarrow t$, where $t \in \Sigma$;
2. variable size: $A \rightarrow \omega$, ω is a set of tiles over $N \cup \{\#\}$.

At each step of the derivation, an A -homogeneous sub picture is replaced with an isometric picture of the local language defined by the right part α of a rule $A \rightarrow \alpha$, where α admits a strong homogeneous partition. The process terminates when all non terminals have been eliminated from the current picture.

Regional Tile Grammars (RTG) [10] are the Tile Grammars with specified set of tiling.

Definition 7 A homogeneous partition is regional (HR) iff distinct (not necessarily adjacent) subdomains have distinct labels. A picture p is regional if it admits a HR partition. A language is regional if all its pictures are so. A regional tile grammar (RTG) is a tile grammar (see Definition 6), in which every variable size rule $A \rightarrow \omega$ is such that $LOC(\omega)$ is a regional language.

2.2 Petri Nets

Definition 8 Petri Net is one of the mathematical modeling tools for the description of distributed systems involving concurrency and synchronization. It is a weighted directed bipartite graph consisting of two kinds of nodes called places (represented by circles) and transitions (represented by bars). Places represents conditions and transition represents events. The places from which a directed arc runs to a transition are called input places of the transition and the places to which directed arcs run from a transition are called output places. Places in Petri nets may contain a discrete number of marks called tokens. Any distribution of tokens over the places will represent a configuration of the net called a marking. In the abstract sense, a transition of a Petri net may fire if it is enabled; when there are sufficient tokens in all of its input places.

Definition 9 A Petri net structure is a four tuple $C = \langle Q, T, I, O \rangle$ where $Q = \{q_1, q_2, \dots, q_n\}$ is a finite set of places, $n \geq 0$, $T = \{t_1, t_2, \dots, t_m\}$ is a finite set of transitions $m \geq 0$, $Q \cap T = \phi$, $I: T \rightarrow Q^\infty$ is the input function from transitions to bags of places and $O: T \rightarrow Q^\infty$ is the output function from transitions to bags of places.

Definition 10 An inhibitor arc from a place q_1 to a transition t_k has a small circle in the place of an arrow in regular arcs. This means the transition t_k is enabled only if q_1 has no tokens in it. In other words a transition is enabled only if all its regular arc input places have required number of tokens and all its inhibitor arc (if exists) input places have zero tokens.

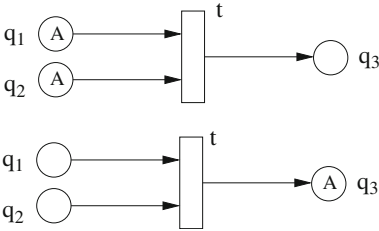
2.3 Array Token Petri Nets

In the array generating Petri Net structure, arrays over an alphabet J are used as tokens in some input places.

Definition 11 Row (resp. column) catenation rules in the form of $A \ominus B$ (resp. $A \oplus B$) can be associated with a transition t as a label, where A is a $m \times n$ array in the input place and B is an array language whose number of columns (resp. rows) depends on the number of columns (resp. rows) of A . Three types of transitions can be enabled and fired

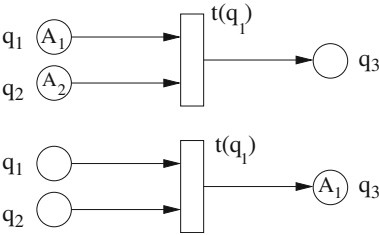
(1) When all the input places of transition t (without label) having the same array as tokens

- Each input place should have at least the required number of tokens (arrays)
- Firing t removes arrays from all its input places and moves the array to all its output places



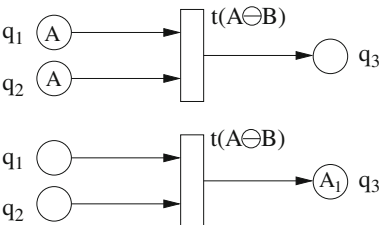
(2) When all the input places of transition t have the different arrays as tokens

- The label of t designates one of its input places which has sufficient number of same arrays as tokens
- Firing t removes arrays from all its input places and moves the array from the designated input place to all its output places.



(3) When all the input places of transition t (with row or column catenation rule as label) have the same array as tokens

- Each input place should have at least the required number of tokens (arrays)
- Firing t removes arrays from all its input places and creates the catenated array as per the catenation rule, in all its output places



In all the three types, firing of a transition t is enabled only if all the input places corresponding to inhibitor arcs (if exist) does not have any tokens in it.

Definition 12 An Array Token Petri net structure (ATPNS) is a five tuple $N = \langle J, C, M_0, \rho, F \rangle$ where J is a given alphabet, $C = \langle Q, T, I, O \rangle$ is a Petri net structure with tokens as arrays over J , $M_0 : Q \rightarrow J^{**}$, is the initial marking of the net, $\rho : T \rightarrow L$, a mapping from the set of transitions to set of labels of transitions and $F \subset Q$, is a finite set of final places.

Definition 13 If P is an ATPNS then the language generated by P is defined as $L(P) = \{X \in J^{**} / X \text{ is in the place } q \text{ for some } q \text{ in } F\}$. Starting with arrays (tokens) over a given alphabet as initial marking, all possible sequences of transitions are fired. Set of all arrays created in final places of F is called the language generated by Petri Net structure.

3 Adjunct Array Token Petri Net Structure

In this section, we recall the notions of adjunct array token Petri net structure [7] in generalized form and give some examples.

Definition 14 Adjunction is a generalization of catenation. In the row catenation $A \oplus B$, the array B is joined to A after the last row. But row adjunction can join the array B into array A after any row of A . Similarly column adjunction can join the array B into array A after any column of A . Let A be an $m \times n$ array in J^{**} called host array; $B \subset J^{**}$ be an array language whose members, called adjunct arrays have fixed number of rows. A row adjunct rule (RAR) joins an adjunct array B into a host array A in two ways : By post rule denoted by (A, B, ar_j) , array B is juxtaposed into Array A after j th row and by pre rule denoted by (A, B, br_j) , array B is juxtaposed into Array A before j th row. The number of columns of B is same as the number of columns of A . In the similar notion column adjunct rule (CAR) can also be defined in two ways : post rule (A, B, ac_j) and pre rule (A, B, bc_j) joining B into A , after j th column of A and before j th column of A respectively. It is obvious that a row catenation rule $A \oplus B$ in ATPNS is a post RAR rule (A, B, ar_m) and column catenation rule $A \oplus B$ is a post CAR rule (A, B, ac_n) . Transitions of a Petri net structure can also be labeled with row or column adjunct rules.

Definition 15 An Adjunct Array Token Petri Net Structure (AATPNS) is a five tuple $N = \langle J, C, M_0, \rho, F \rangle$ where J is a given alphabet, $C = \langle Q, T, I, O \rangle$ is a Petri net structure with tokens as arrays over J , $M_0 : Q \rightarrow J^{**}$, is the initial marking of the net, $\rho : T \rightarrow L$, a mapping from the set of transitions to set of labels where catenation rules of the labels are either RAR or CAR and $F \subset Q$, is a finite set of final places.

In AATPNS, the types of transitions which can be enabled and fired are similar to that of Definition 11 except the type (3) where labels of transitions may be RAR or CAR rules instead of row or column catenation rules.

When all the input places of a transition t (with RAR or CAR rule as label) have the same array as tokens.

Each input place should have at least the required number of tokens (arrays)

- Firing t removes arrays from all its input places and creates the array through adjunction as per the RAR or CAR rule, in all its output places In all the three types, firing of a transition t is enabled only if all the input places corresponding to inhibitor arcs (if exists) does not have any tokens in it.

Definition 16 If P is an AATPNS then the language generated by P is defined as $L(P) = \{X \in J^{**} / X \text{ is in the place } q \text{ for some } q \text{ in } F\}$. Starting with arrays (tokens) over a given alphabet as initial marking, all possible sequences of transitions are fired. Set of all arrays created in final places F is called the language generated by AATPNS Petri Net structure.

Example 1 Consider the Adjunct Array token Petri net structure $P_2 = \langle J, C, M_0, \rho, F \rangle$ where $J = \{0, 1\}$, $C = (Q, T, I, O)$, $Q = \{q_1, q_2\}$, $T = \{t_1, t_2\}$, $I(t_1) = \{q_1\}$, $I(t_2) = \{q_2\}$, $O(t_1) = \{q_2\}$, $O(t_2) = \{q_1\}$, M_0 is the initial marking: the array S is in q_1 and there is no array in q_2 , $\rho(t_1) = (A, B_1, ac_n)$ and $\rho(t_2) = (A, B_2, ar_m)$ and $F = \{q_1\}$.

Where $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $B_1 = (0)_m$ and $B_2 = (0)^{n-1}1$. Initially t_1 is the only enabled transition. Firing of t_1 adjoins a column of 0's after the last column of array S and puts the derived array in q_2 , making t_2 enabled. Firing t_2 adjoins a row of 0's ending with 1 after the last row of the array in q_2 and puts the derived array in q_1 . When the transitions t_1, t_2 fire the array that reaches the output place q_1 is shown as

$$\begin{matrix} 1 & 0 & \xrightarrow{t_1} & 1 & 0 & 0 & \xrightarrow{t_2} & 1 & 0 & 0 \\ 0 & 1 & & 0 & 1 & 0 & & 0 & 1 & 0 \end{matrix}$$

Firing the sequence $(t_1 t_2)^2$ generates the

$$\begin{matrix} & & & 1 & 0 & 0 & 0 \\ & & & 0 & 1 & 0 & 0 \\ \text{output array as} & & & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 1 \end{matrix}$$

The language generated by Petri net is set of square

pictures over $\{0, 1\}$ with 1's in the main diagonal and other elements are 0's (Fig. 1).

Example 2 Consider the Adjunct Array token Petri net structure $P_1 = \langle J, C, M_0, \rho, F \rangle$ where $J = \{a\}$, $C = (Q, T, I, O)$, $Q = \{q_1, q_2\}$, $T = \{t_1, t_2\}$, $I(t_1) = \{q_1\}$, $I(t_2) = \{q_2\}$, $O(t_1) = \{q_2\}$, $O(t_2) = \{q_1\}$, M_0 is the initial marking: the array S is in q_1 and there is no array in q_2 , $\rho(t_1) = (A, B_1, ac_n)$ and $\rho(t_2) = (A, B_2, ar_m)$ and $F = \{q_1\}$.

Where $S = a$, $B_1 = (a)_m$, $B_2 = (a)^n$. On firing the sequence $(t_1 t_2)^n$ $n \geq 0$ set of all square pictures over a is generated (Fig. 2).

Example 3 Consider the Adjunct Array token Petri net structure $P_3 = \langle J, C, M_0, \rho, F \rangle$ where $J = \{a\}$, the Petri net structure is $C = (Q, T, I, O)$ with

Fig. 1 Adjunct array token Petri net generating unit matrix pictures

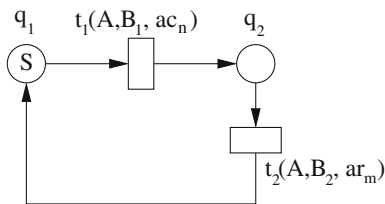
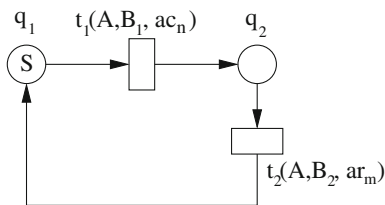


Fig. 2 AATPNS to generate square pictures of *a*'s



$Q = \{p_1, p_2, p_3, p_4, p_5, p_6, p_7\}$, $T = \{t_1, t_2, t_3, t_4, t_5, t_6\}$, $I(t_1) = \{p_1, p_2\}$, $I(t_2) = \{p_3\}$, $I(t_3) = \{p_4\}$, $I(t_4) = \{p_1, p_5\}$, $I(t_5) = \{p_5, p_6\}$, $I(t_6) = \{p_1, p_2\}$, $O(t_1) = \{p_3\}$, $O(t_2) = \{p_4\}$, $O(t_3) = \{p_2, p_5\}$, $O(t_4) = \{p_6\}$, $O(t_5) = \{p_1\}$, $O(t_6) = \{p_7, p_2\}$.

$\rho: T \rightarrow L$ is defined as follows: $\rho(t_1) = p_2$, $\rho(t_2) = (A, B_1, ac_n)$, $\rho(t_3) = (A, B_2, ar_m)$, $\rho(t_4) = \lambda$, $\rho(t_5) = \lambda$, $\rho(t_6) = \lambda$, $\rho(t_7) = \lambda$, $F = \{p_7\}$. The Petri net graph is given in Fig. 3. The arrays used are $S = \begin{matrix} a & a \\ a & a \end{matrix}$, $B_1 = (aa)_m$,

$$B_2 = \begin{pmatrix} a & \\ & a \end{pmatrix}^n.$$

To start with only t_1 is enabled. Firing of sequence of transitions $t_1t_2t_3$ results in a square of *a*'s of size 4×4 in p_2 and p_5 . At this stage both t_6 and t_4 are enabled. Firing the sequence $t_1t_2t_3t_6$ puts a square of size 4×4 in p_7 . Firing t_4 pushes the array to p_6 , emptying p_5 . In this position t_5 is enabled. Firing t_5 puts two copies of same array in p_1 . Since at this stage there are two tokens in p_1 , the sequence $t_1t_2t_3$ has to fire two times to empty p_1 . The firing of sequence $t_4t_5(t_1t_2t_3)^2t_6$ puts a square of *a*'s of size 8×8 in p_7 . The inhibitor input p_1 make sure that a square of size 6×6 does not reach p_7 . This AATPNS generates the language of squares of *a*'s of size 2^n , $n \geq 1$.

Example 4 The AATPNS $P_4 = \langle J, C, M_0, \rho, F \rangle$ with $J = \{a, b\}$, $F = \{p\}$ given in Fig. 4, where

$$S \in \left\{ \begin{matrix} a & b & b & a & b & b & b & b & b & a & a & b & a & a & a & a \\ b & b & b & b' & b & b & b & b' & a & a & a & a' & a & a & a & a \end{matrix} \right\},$$

$B_1 = (aa)_m$, $B_2 = (a)^n$, $B_3 = (bb)_m$, $B_4 = (b)^n$ generates the language of pictures composed by symmetrical *L* shaped strings of same character over the alphabet $J = \{a, b\}$.

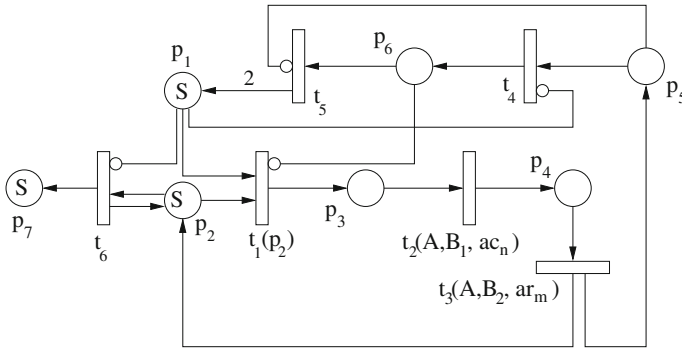


Fig. 3 AATPNS to generate square picture of a 's of size 2^n

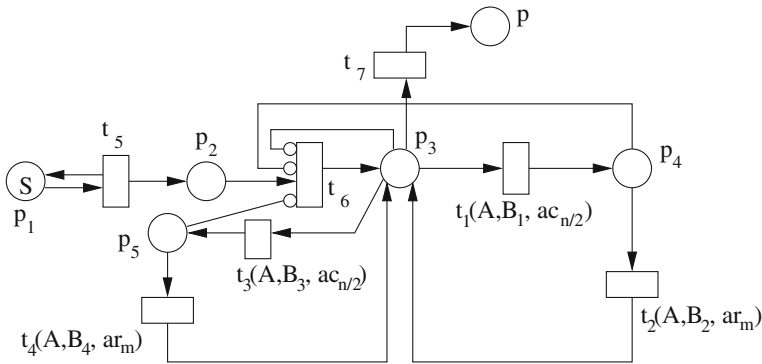


Fig. 4 AATPNS to generate symmetric L shaped strings

A typical picture in this language is

a	b	a	a	a	a	b	a
b	b	a	a	a	a	b	b
a	a	a	a	a	a	a	a
a	a	a	a	a	a	a	a

4 Comparative Results

In the following results $\mathcal{L}(X)$ denotes the family of all languages generated by X .

Theorem 1 $\mathcal{L}(RP2DCFG) \subset \mathcal{L}(AATPNS)$

Proof In [8] the authors have proved that $\mathcal{L}(RP2DCFG) \subseteq \mathcal{L}(ATPNS)$. By the result $\mathcal{L}(ATPNS) \subset \mathcal{L}(AATPNS)$ [7], we have $\mathcal{L}(RP2DCFG) \subseteq \mathcal{L}(AATPNS)$. For

Fig. 5 A picture in the language $L_{\{+b\}}$

```

a a a b a
a a a b a
b b b b b
a a a b a
a a a b a
    
```

strict inclusion the language generated by AATPNS in Example 3 cannot be generated by any RP2DCFG [2].

Theorem 2 $\mathcal{L}(AATPNS)$ and $\mathcal{L}(RTG)$ are incomparable but not disjoint.

Proof The language of squares over a , in Example 2, can be generated by an RTG $G = \langle J, N, S, R \rangle$, where $N = \{S, A, B, A', B', C\}$, $J = \{a\}$ and R consists of variable size rules:

$$\begin{aligned}
 S &\rightarrow \left[\begin{array}{ccccc} \# & \# & \# & \# & \# \\ \# & S & S & A & \# \\ \# & S & S & A & \# \\ \# & B & B & C & \# \\ \# & \# & \# & \# & \# \end{array} \right], & A &\rightarrow \left[\begin{array}{ccc} \# & \# & \# \\ \# & A & \# \\ \# & A' & \# \\ \# & \# & \# \end{array} \right], \\
 B &\rightarrow \left[\begin{array}{cccc} \# & \# & \# & \# \\ \# & B & B' & \# \\ \# & \# & \# & \# \end{array} \right]
 \end{aligned}$$

Fixed size rules as $S, A, A', B, B' \rightarrow a$
 Therefore $\mathcal{L}(AATPNS)$ and $\mathcal{L}(RTG)$ are non disjoint.

The language L_{+b} of pictures consists of a horizontal and a vertical string of b 's (not in the border) in the background of a 's can be generated by RTG [10]. A typical member of L_{+b} is given in Fig. 5. This language cannot be generated by any AATPNS, as the number of transitions in the net cannot depend on the size of the array. In an $m \times n$ array of a 's a column of b 's can be adjuncted in $n - 1$ ways and a row of b 's can be adjuncted in $m - 1$ ways. To insert both a column of b 's and a row of b 's the net requires $(m - 1)(n - 1)$ transitions with corresponding adjunction rules. Hence it is not feasible to generate these arrays using AATPNS.

The language in Example 4 cannot be generated by any RTG grammar G [2].

Theorem 3 $\mathcal{L}(AATPNS)$ and $\mathcal{L}(PG)$ are incomparable but not disjoint.

Proof The language of squares over a , in Example 2 can also be generated by a Prusa Grammar [11]. Again the language L_{+b} in the proof of Theorem 2 can also be generated by a PG [11]. Since $\mathcal{L}(PG) \subset \mathcal{L}(RTG)$ [10], the language in Example 4 cannot be generated by Prusa grammar also.

Theorem 4 $\mathcal{L}(AATPNS)$ and LOC are incomparable but not disjoint.

Proof The language of squares over a , in Example 2 is not a local language [5]. In [2] the authors have proved that the language L_{diag} of pictures containing arbitrary number of diagonals of 1 (no single 1 are admitted at corners) that are separated by at least one diagonal of 0's is in LOC . But L_{diag} cannot be generated by any AATPNS. The only 2×2 array in the language is $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. But there are four 3×3 arrays with the property belonging to the language. To generate four 3×3 arrays from the start array we need 8 transitions with different array languages involved in the labels of the transitions. Hence it is impossible to construct a net with finite number of transitions. Finally the language of square pictures over $\{0, 1\}$ with 1's in the main diagonal and other elements are 0's in LOC [5] and it can also be generated by an AATPNS given in Example 1.

5 Conclusion

In this work, we consider Adjunct array token Petri net structure, recently introduced by Lalitha et al. [7]. We compare this model with recent context free array grammars and the class LOC . The future works concern the study of hierarchy of AATPNS with the other existing generating devices. The relationship of AATPNS with context free and context sensitive controlled pure two dimensional grammars is to be investigated.

References

1. Baker, H.G.: Petri Net Languages Computation Structures Group Memo 68, Project MAC, MIT, Cambridge, Massachusetts (1972)
2. Bersani, M.M., Frigeri, A., Cherubini, A.: On some classes of 2D languages and their relations. In: IWCIA, pp. 222–234, (2011)
3. Cherubini, A., Crespi-Reghizzi, S., Pradella, M., Peitro, P.S.: Picture languages: Tiling systems versus tile rewriting grammars. *Theor. Comput. Sci.* **356**, 90–103 (2006)
4. Crespi-Reghizzi, S., Pradella, M.: Tile rewriting grammars and picture languages. *Theor. Comput. Sci.* **340**, 257–272 (2005)
5. Giammarresi, D., Restivo, A.: Two-dimensional languages. In: Rozenberg, G., Salomaa, A. (eds.) *Handbook of formal languages*. vol. 3, pp. 215–267. Springer Verlag, (1997)
6. Hack, M.: Petri net languages. Computation Structures Group Memo 124, Project MAC, MIT (1975)
7. Lalitha, D., Rangarajan, K., Thomas, D.G.: Adjunct array images using petri nets. *Indian J. Math. Math. Sci.* **8**(1), 11–19 (2012)
8. Lalitha, D., Rangarajan, K., Thomas, D.G.: Rectangular arrays and petri nets. In: IWCIA, pp. 166–180, (2012)
9. Peterson, J.L.: *Petri Net Theory and Modeling of Systems*. Prentice Hall Inc, Englewood Cliffs (1981)

10. Pradella, M., Cherubini, A., Crespi-Reghizzi, S.: A unifying approach to picture grammars. *Inf. Comput.* **209**, 1246–1267 (2011)
11. Prusa, D.: Two-dimensional Languages. Ph.D. Thesis, (2004)
12. Rosenfeld, A., Siromoney, R.: Picture languages—a survey. *Languages of Design* **1**(3), 229–245 (1993)
13. Siromoney, R.: Advances in array languages. In: Proceedings of the 3rd international workshop on graph grammars and their application to computer science. LNCS, Springer, vol. 291, pp. 549–563 (1987)
14. Siromoney, R., Siromoney, G.: Extended controlled table L -arrays. *Inf. Control* **35**, 119–138 (1977)
15. Siromoney, G., Siromoney, R., Kamala, K.: Array grammars and kolam. *Comput. Graph. Image Process.* **3**(1), 63–82 (1974)
16. Siromoney, G., Siromoney, R., Kamala, K.: Picture languages with array rewriting rules. *Inf. Control* **22**, 447–470 (1973)
17. Siromoney, G., Siromoney, R., Krithivasan, K.: Abstract families of matrices and picture languages. *Comput. Graph. Image Process.* **1**, 284–307 (1972)
18. Subramanian, K.G., Rosihan Ali, M., Geethalakshmi, M., Nagar, A.K.: Pure 2D picture grammars and languages. *Discrete Appl. Math.* **157**(16), 3401–3411 (2009)

Non-dominated Sorting Particle Swarm Optimization Based Fault Section Estimation in Radial Distribution Systems

Anoop Arya, Yogendra Kumar and Manisha Dubey

Abstract This paper addresses the fault section estimation as a real multi-objective optimization problem. Non-dominated sorting particle swarm optimization (NSPSO) based on the concept of NSGA-II, has been proposed to alleviate the problems associated with conventional multi objective evolutionary algorithms. An analytical fault analysis and iterative procedure to get the multiple estimates of fault location and NSPSO based optimization algorithm to further nail down the exact fault location has been presented. The techniques fully consider loads, laterals and customer trouble calls in radial distribution systems, take into account for all types of fault. Due to the presence of various conflicting objective functions, the fault location task is a multi-objective, optimization problem. In the proposed technique, the multi-objective nature of the fault section estimation problem is retained using non-dominated sorting approach. As a result, the proposed methodology is generalized enough to be applicable to any radial distribution systems. The applicability of the proposed methodology has been demonstrated through detail simulation studies on standard test systems. Results are used to reduce the possible number of potential fault location which helps and equips the operators to locate the fault accurately.

Keywords Radial distribution systems · Fault section estimation · NSPSO · NSGA-II · Particle swarm optimization

A. Arya (✉) · Y. Kumar · M. Dubey
Department of Electrical Engineering, MANIT, Bhopal, India
e-mail: anooparya.nitb@gmail.com

1 Introduction

Fault location in distribution systems, however, because of their specific topological and operational characteristics, still presents challenges. Still today, Fault location in distribution systems is often performed through visual inspection, field methods and brute force methods [1]. These techniques are not feasible on underground systems and require a long time in large distribution networks. The goal of power system fault analysis is to provide enough information to utility staff to be able to understand the reasons for the interruption better, and provide as quick as possible an action to restore the power delivery. A fault in a power system can trigger cascaded events, leading to major outage if the fault is not isolated quickly and accurately. The requirement to improve service quality, service continuity indexes and to reduce loss of revenue, forces the utilities to rapidly locate faults in their distribution network. Moreover, the exact location of the fault has to be determined, so that the maintenance team can repair the faulted components and restore the system.

Due to the heterogeneity of these feeders, unbalances due to the untransposed lines, presence of laterals, dynamic characteristics of the loads, an efficient method is required to solve fault section estimation problem.

Traditionally, fault diagnosis is performed off line by experienced engineers. Faulted section of the feeder is located generally using information of customer trouble calls. For estimation of fault section, the maintenance crews rely mainly on phone calls by customers and trial and error methods [2, 3]. But in this process, it takes several hours to identify the exact location of fault. However, software tools for fault location have emerged in recent years. To improve the accuracy and speed of fault location the information is stored in a database and intelligent systems in a control centre and can be accessed for diagnosis of a fault event. Data recorded by recorders at substation, customer phone calls location and status of reclosers are used [3].

Researchers have done considerable work in the area of fault diagnosis particular to radial distribution systems. These methods use various algorithmic approaches, where the fault location is iteratively calculated by updating the fault current.

Analytic Calculation-Based Methods (ACBM) identifies fault location through analytical calculation using information of current and voltage values available from Supervisory Control and Data Acquisition (SCADA) and Fault Recorders [1]. In this class, one of the earliest fault-location algorithms belongs to Srinivasan et al. [4], in which the authors used the concept of simplified distributed parameters for fault location. The presence of load taps beyond the fault is treated by consolidating those load models with that of the remote end load. Taking the loads and their variable impedance behaviour into account was the main contribution which caused noticeable error reduction of the fault-location algorithm.

Other example of the most comprehensive works in the distribution fault location area belongs to Aggarwal et al. [5]. The presented algorithm utilizes

superimposed components to identify the fault location. According to the algorithm, the admittance seen to the left and right side of the assumed fault point are calculated. Then, using the calculated admittance; the superimposed fault current is estimated. Das et al. [6, 7], further developed the algorithm of Srinivasan et al. [4]. His method makes use of pre-fault data to estimate the load's value when the fault occurs and then uses the simplified distributed parameter model for the construction of the fault-location algorithm. Their presented results show good accuracy within the computer simulation environment.

Reasoning-Based Methods (RBM) reasons the fault section by using logical rules or connotative illation, including some techniques, such as Expert System (ES) [8], Artificial Neural Networks (ANN) [9], Fuzzy Logic (FL) [10] etc.

Fuzzy Logic aims to settle the incomplete and uncertain information problem with different degrees of success. These methods are satisfactory and efficient, and some have been used in practical systems. However, there are also some obvious shortcomings, which are weak in solving the problems of fault diagnosis in large-scale power systems. The greatest inconvenience of Fuzzy Logic approach lies in the choice of the membership functions, usually defined based on empirical data. This way, it is necessary to develop efficient algorithms for estimating membership functions in a coherent manner. There is no doubt that fault location using ANN techniques is very fast and gives result without taking any time. But it is seen that the accuracy is very poor in comparison to other techniques. So in ANN based fault location techniques, more research is required to improve the accuracy.

The third technique is Optimization-Based Methods (OBM) which formulates the FSE problem as a 0–1 integer optimization problem. It can then be solved by using a global optimization method such as Genetic Algorithm (GA) [11], Tabu Search (TS) [12] and PSO [13] etc.

More research work is required to make Genetic Algorithm more efficient and to improve the accuracy. To improve the performance of genetic algorithm technique, the elite-preserving operator, which favours the elites of a population by giving them the opportunity to be directly carried over to the next generation, should be used. It has been proved that GAs converges to the global optimal solution in the presence of elitism. Zhengyou et al. [13] proposed Binary Particle Swarm Optimization (BPSO) which takes the failure of protective relays or Circuit Breakers into account. Numerical results reveal that BPSO is superior to GA for the convergence speed and accurate estimation results.

In most of the methods, loads and laterals have not been considered during fault section estimation. None of the methods take into account the location of customer calls as one of the objectives, which is very practical and prevailing component of fault section estimation techniques. Moreover, effect of variation in the fault resistance has not been presented by most of the authors.

To alleviate the problems mentioned in the existing literature, a Non-dominated sorting particle swarm optimization (NSPSO) developed by Li [14] based on NSGA-II developed by Deb et al. [15] is presented for solving the fault section estimation problem in this paper. In this technique, the multi-objective nature of the fault section estimation problem is retained without the need of any tuneable

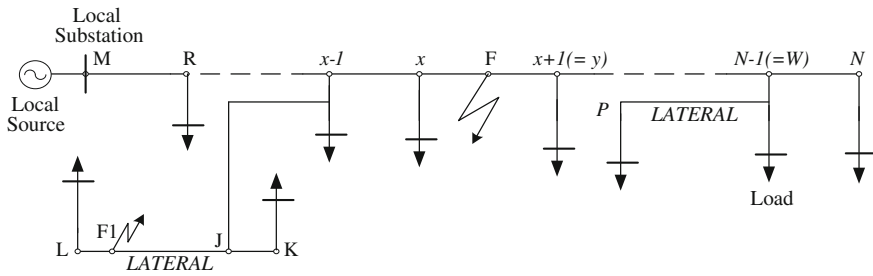


Fig. 1 Single line diagram of a distribution system experiencing a fault [7]

weights or parameters. As a result, the proposed methodology is generalized enough to be applicable to any power distribution network. The applicability of the proposed methodology has been demonstrated through detail simulation studies on different test systems.

2 Overview of the Proposed Fault Location Method

A typical distribution system, which may include three-phase, two-phase, as well as one-phase laterals and loads, is depicted in Fig. 1. A fault section estimation method has been developed in which fault analysis is done by utilizing voltage and current measurements obtained at the local substation based on direct short-circuit analysis. The multiple estimate of the location of fault, faulted phase and type of fault is obtained by existing method described in [6]. The selected system consists of an equivalent source, the line between nodes M and N and two laterals as shown in Fig. 1. Loads are tapped at several nodes and conductors of different types are used on this circuit. Throughout the description, the term “node” will be adopted to represent a single-phase connection point, while a “bus” may contain one, two, or three nodes depending on the number of phases it contains. For example, a three-phase bus will have three nodes.

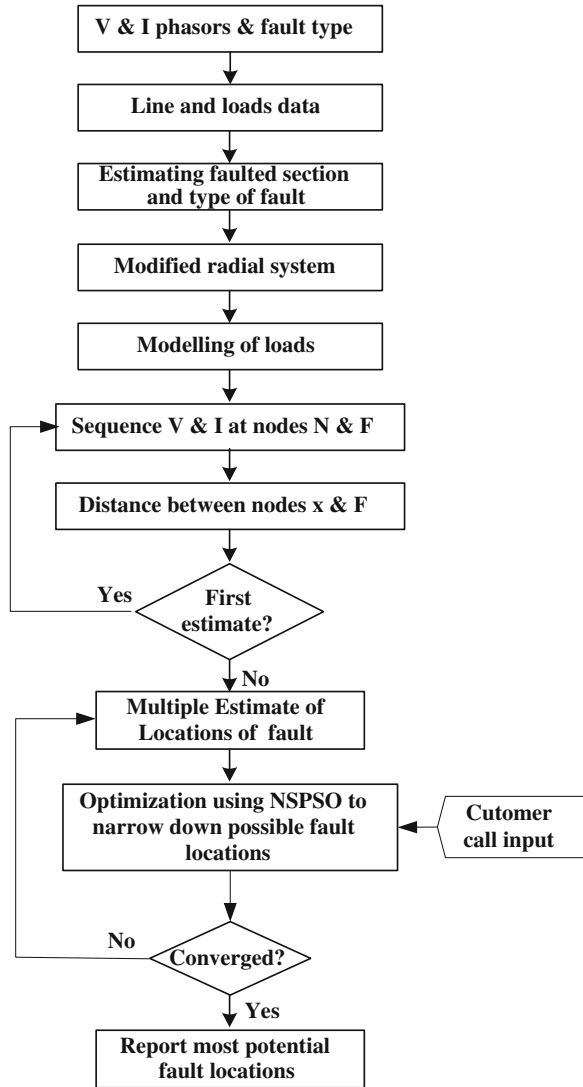
The block diagram shown in Fig. 2 gives an overview of the proposed technique. A detailed description of the technique is described in the next section.

The technique consists of seven major steps as described in [6, 7].

2.1 Pre-fault and Post-fault Data Acquisition

When a fault is detected, the fundamental frequency components of the pre-fault voltage and current phasors at node M are saved. The fundamental frequency components of voltage and current phasors at node M during the fault are

Fig. 2 Block diagram for estimating the location of the fault



estimated and the fault type is determined after a pre-set time has elapsed. These actions are taken on-line and are depicted in the Fig. 3. The pre-fault and fault data, along with line and load parameters, are used in an off-line mode to estimate the location of the fault. Necessary line and load parameters are obtained from a database.

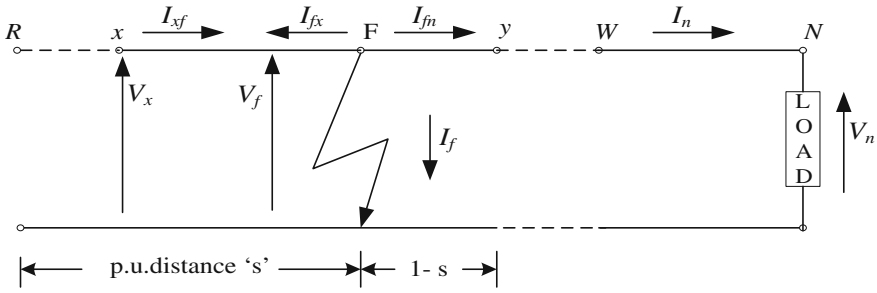


Fig. 3 Voltages and currents at nodes F and N during fault

2.2 Estimation of the Faulted Section

The sequence voltages and currents at node M , before and during the fault, are calculated from the estimated phasors. A preliminary estimate of the location of the fault is made, say between nodes x and $x + l$ ($=y$) as shown in Fig. 1. Line parameters, the type of fault and the phasors of the sequence voltages and currents are used to obtain this estimate. Impedance, measured at the terminal, could point to multiple locations in the system, if it has laterals. Steps given in Sects. 2.3–2.6 are applied using each of the apparent locations.

2.3 Modification in Radial Distribution System

All laterals between node M and the apparent location of the fault are ignored and the loads on a lateral are considered to be present at the node to which the lateral is connected. For example, load at nodes K and L are lumped with the load at node $x - l$.

2.4 Modelling of Loads

All loads up to node x are considered independently and the loads beyond the fault, node F , are assumed to be consolidated with the load at the remote node, N . Non-linear models of loads are used to take into account their dependency on voltage. The constants of load models are computed from the pre-fault load voltages and currents.

2.5 Estimating Sequence Voltages and Currents at the Fault and at Remote Node

Using the voltage-dependent load model, determined in Sect. 2.4, sequence voltages and currents, at node x during the fault, are computed taking into account the load currents. The sequence voltages at the remote end are taken calculated as a function of the distance of the fault from node x . The sequence voltages and currents at the fault node, F , are also obtained as functions of distance of the fault from node x and the impedance of the consolidated load at the remote node.

2.6 Estimating the Distance of Fault from Node x

The relationships between the sequence voltages and currents at the fault are used to estimate the distance of the fault from node x . The first estimate of the distance is obtained using the pre-fault sequence admittance of the consolidated load at node N . Subsequently, the value of the admittance is updated using the net values of the sequence voltages at node N and voltage dependent load model determined in Sect. 2.4. The procedure is repeated until a converged solution is obtained.

3 Voltages and Currents at the Fault and Remote End

The sequence currents at node F for the network shown in Fig. 3, and the sequence voltages at nodes F and N during the fault are estimated using the sequence voltages and currents at node x . The sequence voltages and currents at node R during the fault are calculated by using the voltages and currents at M and the load models. The sequence components of load currents are estimated. The currents at node R flowing towards the fault, are then calculated. The sequence voltages and currents at node x during the fault are calculated using procedure given in [7] for each sequence component and for all sections up to node x . The sequence voltages and currents at node F during the fault are estimated by assuming that all loads beyond node x are consolidated into a single load at N , as shown in Fig. 3.

The voltages and currents at nodes F and x are related by

$$\begin{bmatrix} V_f \\ I_{fx} \end{bmatrix} = \begin{bmatrix} 1 & -sB_{xy} \\ sC_{xy} & -1 \end{bmatrix} \begin{bmatrix} V_x \\ I_{xf} \end{bmatrix} \tag{1}$$

where, ‘ s ’ is the per unit distance to node F from node x , B_{xy} and C_{xy} are the constants of section between the nodes x and $x + l(=y)$.

3.1 Converting Multiple Estimates to Single Estimate

The proposed fault estimation technique could provide multiple estimates if the distribution system has laterals. The number of estimates, for a fault, depends on the system configuration and the location of the fault. It is, therefore, necessary to convert the multiple estimates to a single estimate.

To further nail down the most potential fault location, Non-dominated sorting particle swarm optimization (NSPSO) developed by Li [14], have been applied in the following section. The proposed algorithm estimates the location of shunt faults on radial distribution systems including customer calls as one of the objectives.

4 Problem Formulation of Fault Section Estimation

Fault location problem is formulated as multi objective function problem. These objective functions [16] will use the data collected, such as current and voltage measurements and customer phone calls to create a list of most potential fault locations and verify them with the available multiple estimates from the iterative approach as described in Sects. 2 and 3.

4.1 Objective Functions

(a) Minimization of difference of measured and calculated fault current

$$\min f_1 = \sum_{n=1}^N \sum_{l=1}^{NF} (I_M - I_C) \quad (2)$$

The first objective function compares the calculated fault currents with the measured fault current value. Note that the pre-fault load current must be added to fault currents calculated by the fault analysis routine. Locations where the calculated fault current is within a specified range of the measured current are listed as potential fault locations. It is assumed that the difference between the calculated and measured values were within 1–2 %.

(b) Minimization of distance between customer fault location and predicted fault location

$$\min f_2 = \sum_{n=1}^N \sum_{m=1}^{NC} \sum_{l=1}^{NF} L_{CC-F} \quad (3)$$

The second objective function minimizes the distance between the location of customer phone call and existing fault to further reduce the list of potential fault locations.

(c) Minimization of pre fault voltage and voltage during fault

$$\min f_3 = \sum_{n=1}^N \sum_{l=1}^{NF} (V_r - V_f) \tag{4}$$

where,

- f_1, f_2, f_3 the defined fitness function,
- N number of buses
- NF number of faults
- NC number of customer calls
- I_M measured fault current
- I_C calculated fault current as obtained from Eq. (1)
- L_{CC-F} distance between customer call location and predicted fault locations as obtained from customer call input file
- V_r pre fault voltage as obtained from buffer memory
- V_f voltage during fault as obtained from Eq. (1)

The minima of single objective function converted from these multiple objectives using non-dominated sorting concept [15], correspond to the most likely fault locations. The optimization algorithm and fitness function evaluation are described in detail in the following section. By decreasing the search area for the maintenance crew the time required to find the exact location of the fault can be highly reduced.

Also if estimation is started with only the critical components rather than all the components in the circuit, much smaller number of potential fault locations can be obtained. The critical components are defined as the ones that are the most likely to fail, based on the historical observations by the distribution utility.

5 Particle Swarm Optimization

Kennedy and Eberhart [17] proposed an approach called particle swarm optimization (PSO) in 1995. In PSO, individuals, referred to as particles, are “flown” through hyper dimensional search space. Changes to the position of particles within the search space are based on the social-psychological tendency of individuals to emulate the success of other individuals. The changes to a particle within the swarm are therefore influenced by the experience, or knowledge, of its neighbours. The concept of PSO is simple and is easy to implement. In PSO, in

each iteration, each agent is updated with reference to two “best” values: *pbest* is the best solution (in terms of fitness) the individual particle has achieved so far, while *gbest* is the best obtained globally so far by any particle in the population. Each agent seeks to modify its position using the current positions, current velocities, the distance between the current position and *pbest*, and the distance between the current position and *gbest*. Almost all modifications vary in some way the velocity update equation as given in Eq. (5):

$$v_i^{k+1} = w_i v_i^k + c_1 rand_1 * (pbest_i - x_i^k) + c_2 rand_2 * (gbest - x_i^k) \quad (5)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (6)$$

In Eq. (5), c_1 and c_2 are positive constants, defined as acceleration coefficients, c_1 and c_2 , are often equal to 2, though other settings are used in different papers, typically $c_1 = c_2$ and in the range [1.5–2.5]; w_i is the inertia weight factor and it is used to control the balance between exploration and exploitation of velocity; $rand_1$ and $rand_2$ are two random functions in the range of [0-1]; x_i represents the i th particle and $pbest_i$ is the best previous position of x_i ; $gbest_i$ is the best particle among the entire population; v_i is the rate of the position change (velocity) for particle x_i . Velocity changes in the Eq. (5) comprise three parts, i.e. the *momentum* part, the *cognitive* part, and the *social* part. For number of particles, the typical range is 20–40 [18]. This combination provides a velocity getting closer to *pbest* and *gbest*. Every particle’s current position is then evolved according to the Eq. (6), which produces a better position in the solution space.

5.1 Non-dominated Sorting Particle Swarm Optimization (NSPSO)

NSPSO algorithm developed by Li [14] is based on the same non-dominated sorting concept used in NSGA-II [15]. Instead of comparing solely on a particle’s *pbest* with its potential offspring; the entire population of N particles’ *pbest* and N of these particles’ offspring are first combined to form a temporary population of $2N$ particles. After this, the non-dominated sorting concept is applied, where the entire population is sorted into various non-domination fronts. The first front being completely a non-dominant set in the current population and the second front being dominated by the individuals in the first front only and the front goes so on. Each individual in each front is assigned fitness values or based on front in which they belong to. Individuals in the first front are given a fitness value of 1 and individuals in second are assigned a fitness value of 2 and so on. In addition to the fitness value, two parameters called crowding distance and niche count [14] is calculated for each individual to ensure the best distribution of the non-dominated solutions. The crowding distance is a measure of how close an individual is to its neighbours and Niche count provides an estimate of the extent of crowding near a

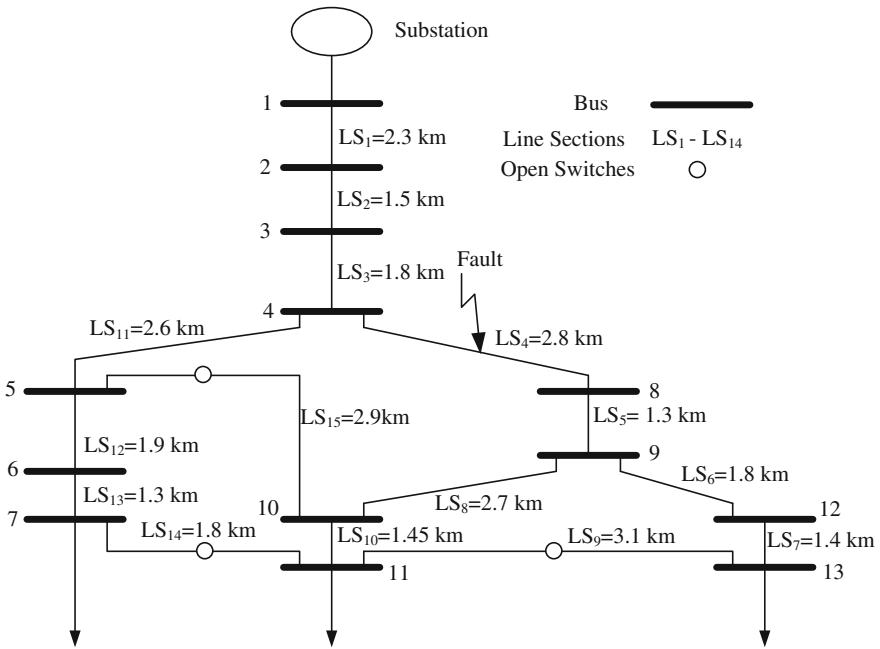


Fig. 4 Radial distribution system under test [16]

solution. The $gbest_i$ for the i th particle x_i is selected randomly from the top part of the first front (the particle which has the highest crowding distance). N particles are selected based on fitness and the crowding distance to play the role of $pbest$.

6 Implementation of NSPSO in FSE Problem

In this paper, the input to the algorithms is given as binary string. For example, line section 4 (LS_4) between the buses 4 and 8 as shown in Fig. 4 is coded as [0 0 1 0 0]. Similarly other sections are coded. The bits in the string will increase according to the increase in number of buses for different test systems. The voltage, current and impedances are calculated with respected to these nodes and line sections. Initial population is generated randomly. This is the simplest method, in which no knowledge about the network is required.

NSPSO have already been discussed in detail in [14] hence it is not repeated here.

6.1 Non Dominated Sorting Particle Swarm Optimization (NSPSO) Algorithm for FSE

Algorithmic Steps for implementation of NSPSO based Fault Section Estimation:

- Step 1:* The information available to the algorithm are, (i) system data, (ii) pre fault configuration (iii) post fault configuration. (iv) customer calls input file.
- Step 2:* Generate initial swarm R_k , initial velocity V_k , maximum velocity V_{max} and initial position x_k randomly and code the Line section as binary string, for ex. LS_5 is coded as [0 0 1 0 1].
- Step 3:* Evaluate fitness values using objective functions from Eqs. (2), (3), and (4) and initialize $pbest$ and $gbest$ values.
- Step 4:* Calculate new velocity V_{k+1} and update particle's position using Eqs. (5) and (6).
- Step 5:* Calculate new fitness value in the next iteration and if the new fitness value is better than the current one, update $pbest$ otherwise go to step 4. Determine $gbest$.
- Step 6:* $R_k = x_k \cup pbest_k$ (combine the current solution and all personal best),
- Step 7:* $F_i = \text{non dom sort}(R_k)$ i.e. implement the non-dominated sorting on R_k
- Step 8:* $pbest_{k+1} = \Phi$ and $i = 1$; until $|pbest_{k+1}| + |F_i| \leq N$ (until the $pbest$ set is filled);
- a) $i = i + 1$;
 - b) calculate the niche count and crowding distance for each particle in F_i ;
 - c) $pbest_{k+1} = pbest_{k+1} \cup F_i$;
- Step 9:* Sort F_i in descending order.
- Step 10:* Select randomly $gbest$ for each particle from a specified top part (e.g. top 5 %) of the first front i.e. F_1 .
- Step 11:* $|pbest_{k+1}| = |pbest_{k+1}| \cup F_i$ ($N - |pbest_{k+1}|$) (choose the first $N - |pbest_{k+1}|$ elements of F_i).
- Step 12:* Verify the customer calls from customer call input file in $pbest_{k+1}$. Check the radiality of the solutions in $pbest_{k+1}$ and modify them, if necessary.
- Step 13:* Update particle's position, x_{k+1} using Eqs. (5) and (6) and using the new $pbest$ and $gbest$.
- Step 14:* Choose the $gbest$ after calculating the performance metric values as described in [14] for best solution.
- Step 15:* Test for convergence, using if ($iter < iter_{max}$), stop otherwise go to Step 5.
- Step 16:* If algorithm converges write the fault location report.

The following inertia weight is used in the implementation of NSPSO

$$w_i = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \quad (7)$$

Table 1 Details of test system

Description	No. of buses	Systems nominal voltages (kV)	Total system load	
			kW	KVAR
Radial distribution system under test	13	11.00	2,652	2,652

where, w_{min} is the initial weight, w_{max} the final weight, $iter_{max}$ is the maximum iteration number, $iter$ is the current iteration number.

The values of w_{max} and w_{min} are 0.9 and 0.4 respectively. Other parameters of NSPSO algorithm are:

No. of iterations = 100, Acceleration Coefficients, $c_1 = 2.0$, $c_2 = 2.0$, Number of particles (or swarm size) $N_{par} = 20$

7 Results and Discussion

The proposed approach is applied on a radial distribution system under test as shown in Fig. 4 and described in Table 1. The approach described in Sect. 2 has been used for fault analysis and to find the multiple estimates of fault location. To further narrow down the most potential fault sections, NSPSO based optimization techniques have been used as described in Sects. 5.1, 6, and 6.1 respectively. The performance of NSPSO has been compared with NSGA-II [15] and another existing technique using Binary particle swarm optimization developed by Zhengyou et al. [13]. Zhengyou et al. used weighted sum approach for converting multi objective PSO into single objective PSO.

Different types of faults like LG, LLG, LL faults have been simulated at all nodes of the distribution network as shown in Fig. 4. Different fault resistance values were used in these simulations. The results indicate that the distances of the faults, estimated by the proposed technique i.e. NSPSO are substantially more accurate than the distances estimated by the NSGA-II and BPSO based algorithms. As shown in Fig. 4, every line section (part of line between two nodes or buses) has been assigned a number and distance of the buses is measured in kms from substation. This is done to locate the accurate distance of fault location from substation.

For system shown in Fig. 4, it is assumed that the fault has occurred between buses 4 and 8 and this location is LS_4 . The distance of line section LS_4 , where the fault has occurred, is 6.6–9.4 kms from substation.

Additionally, 1 and 2 % measurement error has been assumed for voltage and current measurements. As seen from the results shown in Table 2, even with these errors the algorithm is able to capture the exact potential fault location. From Table 2, it is clearly seen in the results that component LS_4 is faulty and the exact location of the fault is at a distance of 7.28 kms for LG and LLG faults and 7.56 kms from substation for LL faults. To develop the algorithm for optimization

Table 2 Impact of measurement errors on fault location estimates

Fault type	Fault resistance (ohm)	Fault location (p.u.)	Fault location estimate error (%)			
			With 1 % voltage error	With 2 % voltage error	With 1 % current and voltage error	With 2 % current and voltage error
LG	30	0.3	0.27	0.53	0.74	1.16
LL	1	0.6	0.51	1.04	0.97	2.02
LLG	[1, 1, 50]	0.7	0.57	1.18	1.39	2.76
LLL	[5, 5, 5]	0.9	0.59	1.23	1.78	3.52
LLLG	[1, 1, 1, 10]	0.9	0.62	1.19	1.83	3.58

Table 3 Fault location multiple estimates resulted without NSPSO

Faulted section (line section)	Fault type	Fault location (p.u.)	Fault distance from substation (kms)	Fault resistance (Ω)	FL estimate error (%)	Estimated fault resistance (Ω)
8-9 (LS_5)	LG	0.5	9.05	50	0.01	50.00
	LL	0.5	9.05	1	0.00	1.00
	LLG	0.6	9.18	[1, 1, 50]	0.01	[1.00, 1.00, 49.99]
	LLLG	0.7	9.31	[1, 1, 1, 50]	0.03	[0.99, 0.99, 0.99, 50.01]
3-4 (LS_3)	LG	0.3	4.34	30	0.01	30.01
	LL	0.5	4.70	5	0.01	4.99
	LLG	0.6	4.88	[0.5, 0.5, 50]	0.02	[0.50, 0.50, 49.99]
	LLL	0.8	4.24	[1, 3.5, 5]	0.02	[1.00, 3.49, 5.00]
9-10 (LS_8)	LG	0.6	11.32	50	0.00	50.00
4-8 (LS_4)	LG	0.6	7.28	10	0.00	10.00
	LL	0.7	7.56	1	0.01	0.99
	LLG	0.6	7.28	[1, 1, 50]	0.02	[1.00, 1.00, 49.99]
9-12 (LS_6)	LG	0.5	10.6	20	0.01	19.99
	LL	0.6	10.78	5	0.02	4.99
	LLL	0.6	10.78	[5, 5, 5]	0.03	[5.01, 5.01, 5.01]
12-13 (LS_7)	LG	0.7	12.48	50	0.01	50.00

technique NSPSO, MATLAB 7.11 [21] has been used and for simulating the different type of faults, PSCAD simulation software [22] has been used.

It is to be noted that, the algorithms NSPSO produces the most potential faulted section but the distance in p.u. from previous node or bus has been verified from the results obtained from iterative procedure i.e. Table 3. The distance in kms from substation to p.u distance has been calculated by adding all the lines from

Table 4 Potential fault location using NSPSO

Technique	Number of line sections	Potential faulted section	Distance of fault location from substation (kms)	Run time of algorithm (sec)
NSPSO [Proposed]	15	4–8 (LS ₄)	LG	7.28
			LL	7.56
			LLG	7.28
Elitist NSGA [16]	15	4–8 (LS ₄)	LG	7.35
			LL	7.77
			LLG	7.35
Binary PSO [14]	15	4–8 (LS ₄)	LG	8.71
			LL	8.93
			LLG	8.64

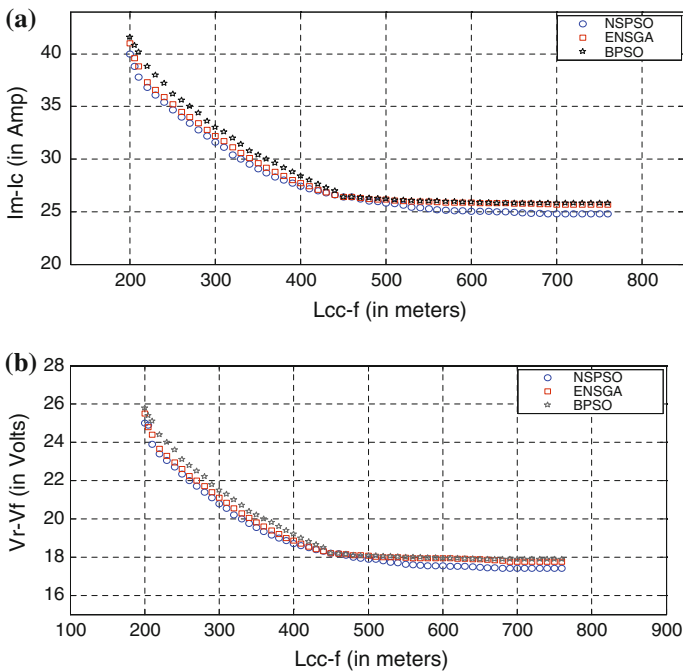


Fig. 5 a Comparison of Pareto fronts obtained from NSPSO, ENSGA and BPSO (Obj-1 v/s Obj-2). b Comparison of Pareto fronts obtained from NSPSO, ENSGA and BPSO (Obj-3 v/s Obj-2)

substation to location of fault using correct traversing path of the system. From Table 4, it is clear that all the three algorithms have given same result for faulted line section i.e. LS₄; it means that most potential and accurate location of fault for this test system is LS₄, whereas, the run time is lowest in NSPSO as compared to

NSGA-II and BPSO using weighted sum approach. There is noteworthy reduction in runtime of NSPSO from ENSGA and BPSO based techniques. It is almost 44.6 % lesser than ENSGA and 58.3 % lesser than BPSO. This shows a major contribution for quick and effective restoration of electric supply.

Figure 5a and b depict the comparison of Pareto-optimal sets obtained from NSPSO, ENSGA and BPSO respectively. It can be seen that the obtained solutions are well distributed on trade-off surface. It can also be seen that the NSPSO is more efficient than ENSGA looking into the optimality and diversification of solutions in the trade-off surface. The comparison of these results shows that the proposed method is well explored the search space.

8 Conclusion

This paper presents analytical fault analysis and iterative procedure to get the multiple estimates of fault location and NSPSO based optimization algorithm to further nail down the exact fault location. The proposed technique fully consider loads, laterals and customer trouble calls in radial distribution systems, take into account for all types of faults, and allow for different fault resistances values.

Simulation studies have demonstrated that proposed technique produces accurate and speedier results and are quite robust to voltage and current measurement errors. Speedy and precise fault location plays an important role in accelerating system restoration, reducing outage time and significantly improving system reliability. Compared with the genetic algorithm, NSPSO for fault-section estimation has got advantages that it uses smaller population size, more rapid convergence and diversity in Pareto optimal solution, which is the main goal of any multi objective optimization problem. The proposed algorithm is compared with two of the most complete methods published to date to confirm the superiority of the work in terms of accuracy.

References

1. IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines, pp. 1–36 (2005)
2. Liu, Y., Schulz, N.N.: Integrated outage information filter for the distribution system using intelligent method. *Power Eng Soc Summer Meeting IEEE* **4**, 2344–2349 (2000)
3. Nastac, L., Thatte, A.A.: A heuristic approach for predicting fault locations in distribution power systems. *Proceedings of the 38th North American power symposium NAPS 2006*, pp. 9–13, 17–19 Sept 2006
4. Srinivasan, K., St-Jacques, A.: A new fault location algorithm for radial transmission lines with loads. *IEEE Trans. Power Del.* **4**(3), 1676–1682 (1989)
5. Aggarwal, R.K., Aslan, Y., Johns, A.T.: New concept in fault location for overhead distribution systems using superimposed components. *IEE Proc. Gener. Transm. Distrib.* **144**(3), 309–316 (1997)

6. Das, R., Sachdev, M.S., Sidhu T.S.: A technique for estimating locations of shunt faults on distribution lines. In Proceedings of the IEEE Communications, Power, Computing Conference, vol. 1, pp. 6–11. IEEE
7. R. Das, Sachdev, M.S., Sidhu, T.S.: A fault locator for radial sub transmission and distribution lines. In: Proceedings IEEE Power Engineering. Society Summer Meeting, vol. 1. pp. 443–448 (2000)
8. Tsai, M.-S.: Development of an object oriented service restoration expert system with load variations. *IEEE Trans. Power Syst.* **23**(1), 219–225 (2008)
9. Coser, J, do Vale, D.T, Rolim, J.G.: Artificial neural network based method for fault location in distribution systems. *Proc. 13th Int. Conf. Intell. Syst. Appl. to Power Syst.*, pp.157-162. Nov (2005)
10. H. Chin, “Fault Section Diagnosis of Power System Using Fuzzy Logic”, *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 245-250, Feb (2003)
11. F.S. Wen Z. X. Han, “Fault section estimation in power systems using genetic algorithm”, *Elect. Power Syst. Res.*, vol. 34, no. 3, pp.165-172 (1995)
12. Wen, F.S., Chang, C.S.: Probabilistic-diagnosis theory for fault-section estimation and state identification of unobserved protective relays using Tabu-search method, *IEE Proc. Gener. Transm. Distrib.* **145**(6), 722–730 (1998)
13. Zhengyou He, Hsiao-Dong Chiang, Chaowen Li and Qingfeng Zeng, “Fault-Section estimation in Power Systems Based on Improved Opt. Model and Binary Particle Swarm Optimization”, *IEEE Power Energy Soc. General Meeting*, pp. 1-8, (2009)
14. X. Li, “A Non-Dominated Sorting Particle Swarm Optimizer for Multi objective Optimization”, in *Proc. Genetic & Evol. Comput. Conf.*, Chicago, USA, pp. 37-48. Aug (2003)
15. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, “A fast and elitist multi objective genetic algorithm: NSGA-II”, *IEEE Trans. Evol. Comput.* vol.6, no.2, pp. 182-197, April (2002)
16. Anoop Arya, Yogendra Kumar, Manisha Dubey, Radharaman Gupta, “Multi-Objective Fault Section Estimation in Distribution Systems using Elitist NSGA”, *Advances in Intelligent Systems and Computing*, Springer, vol.202, pp. 211-219, Dec. (2012)
17. James Kennedy and Eberhart. Russel. C., “Particle Swarm Optimization”, in *Proc. 4th IEEE Int. Conf. on Neural Netw.*, Piscataway, New Jersey, pp. 1942-1948 (1995)
18. Kennedy, James, Eberhart, Russel C.: *Swarm Intelligence*. Morgan Kaufmann Publishers, San Fransisco, California (2001)
19. Benabid, R., Boudour, M., Abido, M.A.: Optimal location and setting of SVC and TCSC devices using non-dominated sorting particle swarm optimization. *Elect. Power Syst. Res.* **79**, 1668–1677 (2009)
20. Kersting, W.H.: Radial distribution test feeders. Proceedings of the IEEE Power Engineering Society Winter Meeting, pp. 908–912 (2001)
21. MATLAB: The Language of Technical Computing. The Mathworks Inc., Massachusetts (2010)
22. User’s Guide of PSCAD/EMTDC, Manitoba HVDC Research Centre (2002)

Approximate Solution of Integral Equation Using Bernstein Polynomial Multiwavelets

S. Suman, Koushlendra K. Singh and R. K. Pandey

Abstract The aim of present article is to find the approximate solution of integral equation using Bernstein multiwavelets approximation. Bernstein polynomial multiwavelets are constructed using orthonormal Bernstein polynomials. These Bernstein polynomial multiwavelets approximate the solution of integral equation. Using orthogonality property of Bernstein polynomial multiwavelets operational matrix of integration is obtained which reduces the integral equation in the system of algebraic equation and can be solved easily. The examples of different profiles are illustrated.

Keywords Abel integral equation · Volterra integral equation · Bernstein polynomials multiwavelets

1 Introduction

Abel's integral equation [1] uses in many branches of science. Usually, physical quantities accessible to measurement are quite often related to physically important but experimentally inaccessible ones by Abel's integral equation. Some of the examples are: microscopy [2], seismology [3], radio astronomy [4], satellite

S. Suman (✉)

Indian Institute of Technology, Patna 800013, Bihar, India

e-mail: ece.math@gmail.com

K. K. Singh · R. K. Pandey

Indian Institute of Information Technology Design and Manufacturing, Jabalpur 482005

Madhya Pradesh, India

e-mail: koushlandras@iiitdmj.ac.in

R. K. Pandey

e-mail: rkpandey@iiitdmj.ac.in

photometry of airglows [5], electron emission [6], atomic scattering [7], radar ranging [8], optical fiber evaluation [9, 10]. But it is most extensively used in flame and plasma diagnostics [11–13] and X-ray radiography [14]. In flame and plasma diagnostics the Abel's integral equation relates the emission coefficient distribution function of optically thin cylindrically symmetric extended radiation source to the line-of-sight radiance measured in the laboratory. Obtaining the physically relevant quantity from the measured one therefore, requires, the inversion of the Abel's integral.

$$f(x) = \int_0^x \frac{\varphi(t)}{\sqrt{x-t}} dt, \quad 0 \leq x \leq 1 \quad (1)$$

where (x) , is the data function and $\varphi(t)$ is the unknown function. The exact solution is given by

$$\varphi(x) = \frac{1}{\pi} \int_0^x \frac{1}{\sqrt{x-t}} \frac{df(t)}{dt} dt, \quad 0 \leq x \leq 1 \quad (2)$$

Assuming without loss of generality, $f(0) = 0$ Abel integral equation is also known as Singular Volterra equation of first kind.

As the process of estimating the solution function $\varphi(t)$, if the data function $f(x)$ is given approximately and only at a discrete set of data points, is ill-posed since small perturbations in the data $f(x)$ might cause large errors in the computed solution (x) . In fact, two explicit analytic inversion formula were given by Abel [1], but their direct application amplifies the experimental noise inherent in the radiance data significantly.

Singular Volterra integral equation of second kind is given as

$$\varphi(t) = f(x) + \int_0^x \frac{\varphi(t)}{\sqrt{x-t}} dt, \quad 0 \leq x \leq 1 \quad (3)$$

where $\varphi(t)$ is the unknown function to be determined, $k(x, t)$, the kernel, is a continuous and square integrable function, $f(x)$ being the known function satisfying $f(a) = 0$.

2 Wavelet and Bernstein Polynomial Multi-wavelets and Function Approximation

The Wavelet Transform has emerged as a powerful mathematical tool in many areas of science and engineering, more so in the field of signal processing Wavelets transform [17] is a method to analysis a signal in time and frequency domain, it is effective for the analysis of time-varying non stationary. Wavelets are a class of functions constructed from dilation and translation of a single function

called the mother wavelet. When the dilation and translation parameters a and b vary continuously, the following family of continuous wavelets are obtained

$$\psi_{ab}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad a, b \in \mathbb{R}, \quad a \neq 0,$$

when the parameters a and b are restricted to discrete values as $a = 2^{-k}, b = n2^{-k}$ then, we have the following family of discrete wavelets $\psi_{kn}(t) = 2^{k/2} \psi(2^k t - n), k, n \in \mathbb{Z}$, where the function ψ , the mother wavelet, satisfies $\int_R \psi(t) dt = 0$.

2.1 Definition

The increasing sequence $\{V_k\}_{k \in \mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ with scaling function $\varphi \in V_0$ is called MRA if (i) $\bigcup_k V_k$ is dense in $L^2(\mathbb{R})$ and $\bigcap_k V_k = \{0\}$, (ii) $f(t) \in V_k$ iff $f(2^{-k}t) \in V_0$, (iii) $\{\varphi(t-n)\}_{n \in \mathbb{Z}}$ is a Riesz basis for V_0 .

Note that (iii) implies that the sequence $\{2^{k/2} \varphi(2^k t - n)\}_{n \in \mathbb{Z}}$ is an orthonormal basis for V_k . Let $\psi(t)$ be the mother wavelet, then $\psi(t) = \sum_{n \in \mathbb{Z}} a_n \varphi(2t - n)$ and $\{2^{k/2} \psi(2^k t - n)\}_{k, n \in \mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{R})$ under suitable conditions. Hence, we can say that wavelet theory is based on above refinement equation which defines the scaling function φ . The scaling function leads to multiresolution approximations (MRAs), wavelets, and fast decomposition and reconstruction algorithms. Multiwavelets are generalization of wavelets, which has been around since the early 1990s. Multiwavelets are useful in the field of signal processing as well as image processing. In this replace the scaling function φ by a function vector $\varphi(x) = (\varphi_1(x), \varphi_2(x), \varphi_3(x), \dots, \varphi_r(x))^T$ called multi-scaling function, and refinement equation by

$$\varphi(x) = \sqrt{m} \sum_k H_k \varphi(mx - k)$$

The recursion coefficient H_k are now $r \times r$ matrices.

Different variations of wavelet bases (orthogonal, biorthogonal, multiwavelets) have been presented and the design of the corresponding wavelet and scaling functions has been addressed [15]. Multiwavelets are generated by more than one scaling function. Multiwavelets lead to MRAs and fast algorithms just like scalar wavelets but they have some advantages in comparison to single wavelets: they can have short support coupled with high smoothness and high approximation order, and they can be both symmetric and orthogonal. The properties such as short support, orthogonality, symmetry and vanishing moments are known to be important in signal processing and numerical methods. A single wavelet cannot possess all these properties at the same time. On the other hand, a multiwavelet

system can have all of them simultaneously. This suggests that multiwavelets could perform better in various applications.

The general form of B-polynomials of m th-degree [16] are defined in the interval $[0, 1)$ as

$$B_{i,m}(x) = \binom{m}{i} x^i (1-x)^{m-i}, \quad \text{where } i = 0, \dots, m \tag{4}$$

Recursive definition can also be used to generate the B-polynomials over $[0, 1]$ so that the i th m th degree B-polynomial can be written as

$$B_{i,m}(x) = (1-x)B_{i,m-1}(x) + xB_{i-1,m-1}(x) \tag{5}$$

B-polynomial multi-wavelets $\psi_{mn}(t) = \psi(k, n, m, t)$ have four arguments: translation argument $n = 0, 1, 2 \dots 2^k - 1$, dilation argument k can assume any positive integer m is the order for B-polynomial and t is the normalized time. They are defined on the interval $[0, 1)$ as

$$\psi_{n,m}(t) = \begin{cases} 2^{\frac{k}{2}} WB_m(2^k t - n), & \text{for } \frac{n}{2^k} \leq t < \frac{n+1}{2^k}, \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

where $m = 0, 1 \dots M, n = 0, 1, 2 \dots 2^k - 1$. In the above equation coefficient $2^{\frac{k}{2}}$ is for orthonormality, the dilation parameter is $a = 2^{-k}$ and the translation parameter is $b = n2^{-k}$. Here WB_m is the orthonormal form of B-polynomials of order m .

Using Gram- Schmidt orthonormalization process on $B_{i,n}(x)$, we obtain a class of orthonormal polynomials from Bernstein polynomials. We call them orthonormal Bernstein polynomials of order n and denote them by $b_{0,n}, b_{1,n}, \dots, b_{n,n}$. We construct the Bernstein multiwavelet with help of these orthonormal Bernstein polynomials.

For $M = 3$ the four orthonormal polynomials are given by,

$$\begin{aligned} b_{0,3}(x) &= \sqrt{7} [(1-x)^3], \\ b_{1,3}(x) &= 2\sqrt{5} [3x(1-x)^2 - \frac{1}{2}(1-x)^3], \\ b_{2,3}(x) &= \frac{10\sqrt{3}}{3} [3x^2(1-x) - 3x(1-x)^2 + \frac{3}{10}(1-x)^3], \\ b_{3,3}(x) &= 4[x^3 - \frac{9}{2}x^2(1-x) + 3x(1-x)^2 - \frac{1}{4}(1-x)^3]. \end{aligned}$$

2.2 Function Approximation

A function $f(x)$ defined over $[0, 1)$ may be expanded as

$$f(x) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} C_{nm} \psi_{nm}(x), \tag{7}$$

Where $C_{nm} = (f(x), \psi_{nm}(x))$, in which (\cdot, \cdot) denotes the inner product. If the infinite series in Eq. (7) is truncated, then that equation can be written as

$$f(x) \approx \sum_{n=0}^{2^k-1} \sum_{m=0}^M C_{nm} \psi_{nm}(x) = C^T \psi(x) \tag{8}$$

Where C and $\psi(x)$ are $(2^k - 1)(M + 1) \times 1$ matrices given by

$$C = [c_{00}, c_{01}, \dots, c_{0M}, c_{10}, \dots, c_{1M}, \dots, c_{(2^k-1)0}, \dots, c_{(2^k-1)M}]^T, \tag{9}$$

$$\psi(x) = [\psi_{00}(x), \psi_{01}(x), \dots, \psi_{0M}(x), \psi_{10}(x), \dots, \psi_{1M}(x), \dots, \psi_{(2^k-1)0}(x), \dots, \psi_{(2^k-1)M}(x)]^T \tag{10}$$

3 Method of Solution

3.1 Solution of Singular Integral Equation

Using Eq. (8), we approximate $\varphi(x)$ and $f(x)$ as

$$\varphi(x) = A^T \psi(x) \tag{11}$$

$$f(x) = F^T \psi(x) \tag{12}$$

where the matrix F is known.

Then from Eqs. (1), (10) and (11) for the Abel integral equation we have

$$F^T \psi(x) = \int_0^x \frac{A^T \psi(x)}{\sqrt{x-t}} dt, \tag{13}$$

For the for Volterra integral equation of second kind,

Using Eqs. (3), (10) and (11), we have,

$$A^T \psi(x) = F^T \psi(x) + \int_0^x \frac{A^T \psi(x)}{\sqrt{x-t}} dt, \tag{14}$$

For simplification of the integral we use the following well known formula,

$$\int_0^x \frac{t^n}{\sqrt{x-t}} dt = \frac{\sqrt{x} x^{n+\frac{1}{2}} \Gamma(n+1)}{\Gamma(n+\frac{3}{2})},$$

Using this formula in the Eqs. (13), (14) we have

$$\int_0^x \frac{A^T \psi(x)}{\sqrt{x-t}} dt = P \psi(x), \tag{15}$$

From Eqs. (13), (15) we get,

In case of Volterra integral equation of first kind of Abel Kernel,

$$A^T = F^T P^{-1}, \tag{16}$$

In case of Volterra integral equation of second kind of Abel Kernel,

$$A^T = F^T (I - P)^{-1}, \tag{17}$$

Now, using (16) one can easily find the approximate solution of Eq. (1). The solution can be given as

$$\emptyset(x) = A^T \psi(x),$$

Now, using Eq. (17), one can find solution of equation of (3).

The solution can be given as

$$\emptyset(x) = A^T \psi(x), \tag{18}$$

4 Numerical Examples

Test Example 1: For $M = 3$ consider an Abel’s integral equation (Volterra integral equation of first kind)

$$\int_0^x \frac{f(t)}{\sqrt{x-t}} dt = \frac{2\sqrt{x}(105 - 56x^2 + 48x^3)}{105}, \quad 0 \leq x \leq 1.$$

Where the exact solution is $f(x) = x^3 - x^2 + 1, 0 \leq x \leq 1$.

Test Example 2: For $M = 3$ consider the following singular volterra integral equation $f(x) = x^2 + \frac{16x^{5/2}}{15} - \int_0^x \frac{f(t)}{\sqrt{x-t}} dt, 0 \leq x \leq 1$. having the exact solution $f(x) = x^2, 0 \leq x \leq 1$.

5 Results and Conclusions

The present article gives to develop an efficient and accurate method for solving singular Volterra integral equation Using Bernstein Polynomial Multiwavelets. The required solution and given function is approximated using Bernstein polynomials multiwavelet and then operational matrix of integration is obtained. This operational matrix leads to a system of linear equation in order to find the approximation coefficient of required solution. Two illustrative examples having different profiles are solved and their results are listed in the Table 1. For different values of t errors for both examples are calculated and values of errors are of the

Table 1 Error of Exm 1 and Exm 2

t	Error of Ex 1	Error of Ex 2
0	-1.045×10^{-5}	1.007×10^{-8}
0.1	2.321×10^{-6}	1.118×10^{-11}
0.2	2.587×10^{-6}	-1.147×10^{-8}
0.3	-1.774×10^{-6}	-2.448×10^{-5}
0.4	-2.891×10^{-6}	-3.909×10^{-5}
0.5	-2.25×10^{-3}	-5.282×10^{-5}
0.6	7.011×10^{-4}	6.386×10^{-5}
0.7	6.171×10^{-4}	5.169×10^{-5}
0.8	-4.893×10^{-4}	-1.068×10^{-5}
0.9	-6.051×10^{-4}	-4.461×10^{-5}
1	2.282×10^{-3}	2.854×10^{-5}

order of 10^{-11} to 10^{-5} for second example. In case of example 1 errors varies from 10^{-8} to 10^{-3} . Results justify that Bernstein polynomials multiwavelet is a better option to solve integral equation.

References

1. Abel, N.H.: Resolution d'un probleme de mecanique. *J. Reine. Angew. Math.* **1**, 153–157 (1826)
2. Mach, L.: *Wien Akad. Ber. Math. Phys. Klasse* **105**, 605 (1896)
3. Jakeman, A.J., Anderssen, R.S.: Abel type integral equations in stereology I. General discussion. *J. Microsc.* **105**, 121–133 (1975)
4. Healy, S.B., Haase, J., Lesne, O.: Abel transform inversion of radio occultation measurements made with a receiver inside the Earth's atmosphere. *Annal. Geophys.* **20**, 1253–1256 (2002)
5. Solomon, S.C., Hays, P.B., Abreu, V.J.: Tomographic inversion of satellite photometry. *Appl. Opt.* **23**, 3409–3414 (1984)
6. Kosarev, E.L.: Applications of integral equations of the first kind in experiment physics. *Comput. Phys. Commun.* **20**, 69–75 (1980)
7. Buck, U.: Inversion of molecular scattering data. *Rev. Mod. Phys.* **46**, 369–389 (1974)
8. Hellsten, H., Andersson, L.E.: An inverse method for the processing of synthetic aperture radar data. *Inverse Prob.* **3**, 111–124 (1987)
9. Anderssen, R.S., Calligaro, R.B.: Non destructive testing of optical-fiber preforms. *J. Aust. Math. Soc. (Ser B)* **23**, 127–135 (1981)
10. Glantschnig, W.J.: How accurately can one reconstruct an index profile from transverse measurement data. *IEEE J. Lightwave Technol.* **3**, 678–683 (1985)
11. Keren, E., Bar-Ziv, E., Glatt, I., Kafri, O.: Measurements of temperature distribution of flames by moiré deflectometry. *Appl. Opt.* **20**, 4263–4266 (1981)
12. Tallents, C.J., Burgess, M.D.J., Luther-Davies, B.: The determination of electron density profiles from refraction measurements obtained using holographic interferometry. *Opt. Commun.* **44**, 384–387 (1983)
13. Fleurier, C., Chapelle, J.: Inversion of Abel's integral equation application to plasma spectroscopy. *Comput. Phys. Commun.* **7**, 200–206 (1974)
14. Yousefi, S.A.: B-Polynomial multiwavelets approach for solution of Abel's integral equation. *Int. J. Comp. Math.* **87**, 310–316 (2010)

15. Yousefi, S.A.: Numerical solution of Abel's integral equation by using Legendrewavelets. *Appl. Math. Comput.* **175**, 574–580 (2006)
16. Bhattacharya, S., Mandal, B.N.: Use of Bernstein polynomials in numerical solutions of Volterra integral equations. *Appl. Math. Scien.* **2**(36), 1773–1787 (2008)
17. Mohammad, P., Ali, T., Morteza, M.G., Saboori, S.: Wavelet compression of ECG signal using SPIHT algorithm. *Int. J. Inf Com Eng.* **2**, 219 (2005)

A Novel Search Technique for Global Optimization

Kedar Nath Das and Tapan Kumar Singh

Abstract Solving non-linear optimization with more accuracy has become a challenge for the researchers. Evolutionary global search techniques today are treated as the alternate paradigm over the traditional methods for their simplicity and robust nature. However, if an evolutionary problem is computationally burdened both the human efforts and time will be wasted. In this paper a much simpler and more robust optimization algorithm called Drosophila Food-Search Optimization (DFO) Algorithm is proposed. This new technique is based on the food search behavior of the fruit fly called 'Drosophila'. In order to evaluate the efficiency and efficacy of the DFO-algorithm, a set of 20 unconstrained benchmark problems have been used. The numerical results confirms the supremacy of DFO over the algorithms namely Hybrid Ant Colony-Genetic Algorithm (GAAP), Level-Set evolution and Latin squares Algorithm (LEA), which are reported as the most efficient algorithms in the recent literature.

Keywords Redundant search (RS) · G-Protein-coupled-receptor (GPCR) · Ligand and modified quadratic approximation (mQA)

1 Introduction

In recent years, many complex structured are arisen in the field of sciences and engineering. Till date, for solving complex non-linear optimization problems, many a number of evolutionary methods have been designed. Few popular algorithms are Diversity Guided Evolutionary Programming (DGEP) [1],

K. N. Das (✉) · T. K. Singh

Department of Mathematics, National Institute of Technology, Silchar 788010,
Assam, India
e-mail: kedar.iitr@gmail.com

T. K. Singh

e-mail: tksingh1977.nits@gmail.com

Self-Adaptive DE (jDE) [2], Modified ABC (ABC/best) [3], Structural Optimization (SO) [4], Self Adaptive Hybridized GA (S-LXPM) [5], Swallow Swarm Optimization (SSO) [6], Evolutionary Programming, etc. However, most of them suffer either with computational burden or with fine tuning of many parameters.

Therefore, in order to increase further the efficiency, reliability and stability, a new evolutionary optimization algorithm called ‘*Drosophila* Food-Search Optimization (DFO)’ Algorithm introduced in this paper. DFO basically mimics the optimal food-searching behavior of fruit fly called *Drosophila melanogaster*. Its mechanism and behavioral description is presented later in the next section.

The rest part of this paper is organized as follows. The Sect. 2, introduces the food-searching behavior of *Drosophila melanogaster*. The components used in the proposed techniques are discussed in Sect. 3. Section 4, introduces the DFO Algorithm. In Sect. 5, numerical results and their analysis are presented. Finally, the conclusion is drawn in Sect. 6.

2 *Drosophila melanogaster*

Drosophila melanogaster is a nature inspired fly approximately 3 mm long and used as a model organism in sense organ and feeling, especially in vision and smelling [7, 8] (shown in Fig. 1). They can collect all the smell in air even though the food sources are out of 40 km. *Drosophila* depends on some hair-like structures, called sensilla, which are located in multiple parts of their body along with proboscis, wing margins, legs and ovipositor [8]. Each of these receptors consisting with Gustatory Receptor Neurons (GRNs) that help *Drosophila* to select foods with nutrient and differentiate between compound content with high sugar as well as with high salt.

Drosophila ingests food through its proboscis, which comprises with both internal and external sensilla. The external sensilla detect preferable food sources and the internal sensilla check the foods before it is allowed into the digestive system. The internal sensilla plays a vital role either as sensors for harmful substances that, if activated elicit a ‘regurgitate’ responses or alternatively, to verify digestible substances and promote sucking reflexes. It also consisting with highly well characterized olfactory receptors called *Drosophila* Olfactory receptors (DOR) [9]. The DOR gene encode a family of seven transmembrane G-Protein coupled receptors, whose function recognizes specific odorant molecules. The DORs are expressed in dendrites of olfactory receptor neurons (ORNs) housed in sensilla located on the antennae and the maxillary palps. The space between the dendrite and the inner surface of the bristle filled with lymph, a secretion from support cells that are associated with each taste sensillum.

2.1 Ligand

A ligand is a signal triggering molecule in response to protein binding. It forms a complex molecular substance when attached with bio-molecule. It alerts a

Fig. 1 Images of *Drosophila melanogaster*



conformational chemical changes, when binds with a receptor and the conformational state of a receptor protein determines its functional state. The strength of binding of ligand with receptor is called affinity. In general, high-affinity ligand binding means greater intermolecular force between the ligand and its receptor. High affinity ligand binding also implies that a relatively low concentration of a ligand is sufficient to occupy a ligand binding site and for low affinity binding, a relatively high concentration of ligand is required before the binding site is maximally occupied and the maximum physiological response to the ligand is achieved.

2.2 G-Protein-Coupled-Receptor

G-Protein-Coupled-Receptor (GPCR) is a large protein family of receptors, which recognizes a vital role of odorant molecules [10] from outside the cell and activate inside signal transduction pathway and gives the cellular responses [11]. It is also called seven-transmembrane domain receptors because it passes through the cell membrane seven times. It also consists with two principle signal transduction pathways namely cAMP (Cyclic Adenosine Mono Phosphate) signal pathway and the Phosphatidylinositol signal pathway. When ligand binds with a GPCR, a conformational changes occur i.e. it mechanically activates the G-Protein (a family of proteins involved in transmitting chemical signals originating from outside a cell into the inside of a cell), which detaches from the receptors. The receptors can now either activate another G-Protein or switch back to its inactive position.

3 Components Used in the Proposed Technique

3.1 The Proposed Redundant Search

In recent years, many new exploration and exploitations techniques namely NSDE [12, 13], LSRCMA [14], RCMA [15] have been introduced by researchers. However, for further enhancement of search capacity and to maintain the diversity in the population, a new search technique called Redundant Search (RS) is proposed in this section. The working step of Redundant Search is as follows:

- Step 1: Generate a population matrix $X_{i,j}$ of size P with dimension D.
- Step 2: Generate unequal random numbers $r1, r2 \in [1, D]$ and $r3, r4 \in [1, P]$.
- Step 3: Split each individual population to two different individuals by the concept of neighboring searching. Thus the two new populations $Y_{i,j}$ and $Z_{i,j}$ are being generated from the current solution. For each $i = 1, 2, \dots, P$, the following mechanism follows:

$$Y_{i,k} = X_{i,k} + |X_{r3,k} - X_{r4,k}| \tag{1}$$

$$Z_{i,k} = X_{i,k} - |X_{r3,k} - X_{r4,k}|, \text{ for } k = r1 \text{ and } r2. \tag{2}$$

for $j \neq r1$ and $j \neq r2$, $Y_{i,j} = X_{i,j}$ and $Z_{i,j} = X_{i,j}$

Step 4: The new individuals for the new population are given by

$$X'_{i,j} = \text{Min} \{ f(X_{i,j}), f(Y_{i,j}), f(Z_{i,j}) \} \tag{3}$$

for $i = 1, 2, \dots, P$ and $j = 1, 2, \dots, D$.

3.2 Modified Quadratic Approximation

Quadratic Approximation (QA) [16] is a method to generate one ‘child’ from three parents from the population. The working principle of QA is as follows.

1. Select the individual R_1 with the best fitness value and more two random individuals R_2 and R_3 such that at least two of them are distinct.
2. Find the point of minima (child) of the quadratic surface passing through R_1, R_2 and R_3 defined as:

$$\text{Child} = 0.5 \left(\frac{(R_2^2 - R_3^2)f(R_1) + (R_3^2 - R_1^2)f(R_2) + (R_1^2 - R_2^2)f(R_3)}{(R_2 - R_3)f(R_1) + (R_3 - R_1)f(R_2) + (R_1 - R_2)f(R_3)} \right) \tag{4}$$

where $f(R_1), f(R_2)$ and $f(R_3)$ are the fitness function values at R_1, R_2 and R_3 , respectively.

In this section, a modified Quadratic Approximation (mQA) is proposed. In mQA, R_1, R_2 and R_3 picks randomly with a condition that all are distinct. Rest mechanism of mQA remains the same as that of QA.

4 Drosophila Food-Search Optimization Algorithm

A new optimization algorithm called *Drosophila* Food-Search Optimization (DFO) algorithm is proposed in this section. BFO algorithm is mainly based on the food-search behavior of the fruit fly ‘*Drosophila*’.

The *Drosophila melanogaster* is superior with its rich history in genetic especially for sensory system including taste and olfaction. The olfactory receptors of *Drosophila* can sense all kinds of food sources whose scents blowing in the air. Initially, after getting the sources, the gustatory receptors (GRs) available in external sensilla of proboscis check the preferences of the food sources. Then they pass the signal to external sensilla to report whether it is allowed in digestive system or not. The internal sensilla consist of GPCR, which sense the molecules outside the cell and activate the signal pathway. Each molecule of a food sources act as a ligand and affinity of protein binding of GPCR with ligand depends on the concentration of the ligand. So, if the concentration is high, GPCR will activate and provide signal to SOG region of the brain via nervous system. When the signal input is more, then SOG region provides positive signal to the proboscis to take food and due to high prefer ability of food sources, the signal input is more. Then they start flying towards the food. The pseudo code of DFO is as follows:

begin

Gen = 1

 Generate initial population

 Evaluate fitness of each individual in the population

 While (termination criterion is not satisfied) **do**

 Gen = Gen + 1

 Apply Tournament selection (size 2)

 Apply Redundant Search

 Apply Modified Quadratic Approximation operator

end do

end begin

The flow mechanism of S-LXPM is presented in Fig. 2.

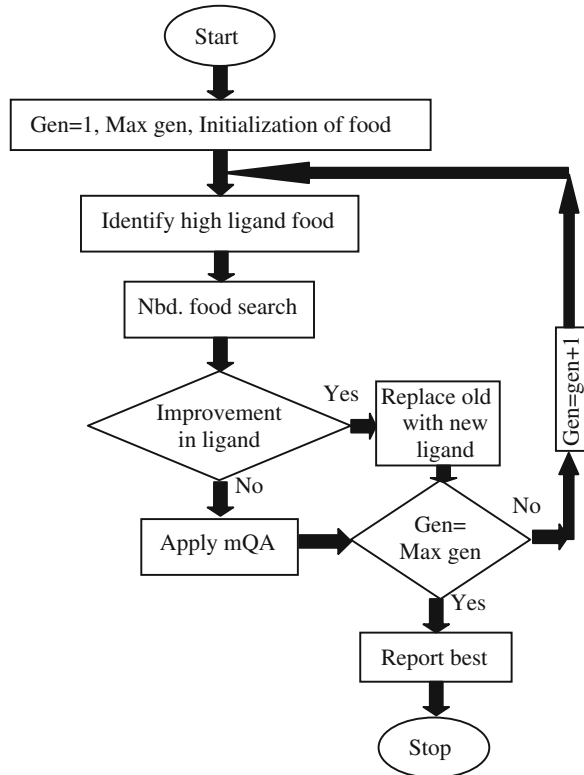
5 Experimental Analysis

5.1 Test Problems

In order to evaluate the performance of *Drosophila* Food-Search Optimization (DFO) algorithm, a set of 20 scalable benchmark problems are taken from the recent literature [16]. The problem set is listed in Table 1. The problems under consideration have the following characteristics.

1. Non-linear and Scalable
2. Bearing various difficulty levels
3. Mixture of unimodal and multimodal functions
4. Separable and non-separable function.

Fig. 2 Schematic representation of DFO



The basic idea to choose such a set is to testify the solving ability of the proposed technique on different varieties of problems.

For all 20 test functions, the result obtained by DFO is compared with Hybrid Ant Colony-Genetic Algorithm (GA-API) and Latin squares Algorithm (LEA) reported in [16]. For function F8, χ_{ij} , φ_{ij} are random integers in $[100, -100]$ and ω_j is a random number in $[\pi, -\pi]$.

5.2 Computational Setup

The proposed algorithm is implemented in C++ on a Pentium Dual Core, 2.00 GHz machine with 1 GB RAM under Windows 7 platform. A total of 50 runs are conducted for each test functions. To start a run, different seeds are used in order to generate the random numbers for the initial population. To stop a run, the criterion is either no improvement found in the objective function value for a total of consecutive 20 generations or a maximum of 500 generation is attained. For a fair comparison, the problem size and the population size are fixed same as reported in [16].

Table 1 A set of unconstrained benchmark problems

Sl. No	Function	Bounds	Dimension
1	$-\sum_{i=1}^n x_i \sin(\sqrt{ x_i })$	$[-500, 500]$	30
2	$10n + \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i)]$	$[-5.12, 5.12]$	30
3	$-20 \exp\left(-0.02 \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n} \sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	$[-32, 32]$	30
4	$1 + \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right)$	$[-600, 600]$	30
5	$\frac{\pi}{n} \left\{ 10 \sin^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2 [1 + 10 \sin^2(3\pi y_{i+1})] + (y_n - 1)^2 \right\} + \sum_{i=1}^n u(x_i, 10, 100, 4) \text{ where } y = 1 + \frac{1}{4}(x + 1) \text{ and}$	$[-50, 50]$	30
6	$0.1 \left\{ \sin^2(\pi x_1) + \sum_{i=1}^{n-1} (x_i - 1)^2 [1 + \sin^2(3\pi x_{i+1})] + (x_n - 1)^2 [1 + \sin^2(2\pi x_n)] \right\}$		30
7	$-\sum_{i=1}^n \sin(x_i) \cdot \sin^{20}\left(\frac{x_i}{\pi}\right)$	$[-50, 50]$ $[0, \pi]$	100

(continued)

$$u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m, & x_i > a \\ 0, & -a \leq x_i \leq a \\ k(-x_i - a)^m, & x_i < -a \end{cases}$$

Table 1 (continued)

Sl. No	Function	Bounds	Dimension
8	$\sum_{j=1}^n \left(\sum_{i=1}^n (\chi_{ij} \sin \omega_j + \varphi_{ij} \cos \omega_j) - \sum_{j=1}^n (\chi_{ij} \sin x_j + \varphi_{ij} \cos x_j) \right)^2$	$[-\pi, \pi]$	100
9	$\frac{1}{n} \sum_{i=1}^n (x^4 i - 16x^2 i + 5xi)$	$[-, 5]$	100
10	$\sum_{i=1}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right]$	$[-, 10]$	100
11	$\sum_{i=1}^n x_i^2$	$[-, 100]$	30
12	$\sum_{i=1}^n ix_i^4 + \text{random } [0, 1]$	$[-28, 1.28]$	30
13	$\sum_{i=1}^n x_i + \prod_{i=1}^n x_i $	$[-, 10]$	30
14	$\sum_{i=1}^n \left(\sum_{j=1}^j x_j^2 \right)^2$	$[-, 100]$	30
15	$\max_i \{ x_i , 1 \leq i \leq n \}$	$[-, 100]$	30
16	$4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4$	$[-, 5]$	2
17	$(x_2 - \frac{5.1}{4x_2^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos x_1 + 10$	$[-, 10]$	2
18	$[1 + (x_1 + x_2 + 1)^2(10 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2^2)(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	$\times [0, 15]$	2
19	$\sum_{i=1}^n \left[a_i - \frac{x_i(0x_i^2 + bx_i)}{b^2 + bix_i + 4x_i^2} \right]^2$	$[-, 5]$	4
20	$-\sum_{i=1}^4 c_i \exp \left(-\sum_{j=1}^6 a_{ij}(x_j - p_{ij})^2 \right)$	$[0, 1]$	6

Table 2 Comparison of time (sec) using DFO with others in [16]

Problem no	LEA	GAAPI	DFO
1	656.30	30.59	1.23
2	557.20	27.07	2.93
3	326.10	18.26	3.21
4	365.60	37.28	1.48
5	368.50	22.48	4.17
6	359.10	42.27	4.11
7	660.80	24.27	12.50
8	493.40	401.75	201.25
9	612.30	37.93	10.54
10	443.80	35.59	5.74
11	240.20	35.64	2.60
12	242.40	17.00	1.29
13	240.80	37.56	1.943
14	241.30	39.58	13.37
15	242.20	30.70	1.91
16	30.80	23.83	0.046
17	30.60	27.87	0.039
18	33.50	27.20	0.035
19	101.30	27.97	0.134
20	66.20	29.06	0.096

5.3 Computational Results and Analysis

Each test function in Table 1 undergoes 50 independent runs. The average computational time for each problem by DFO is compared with LEA and GAAPI in Table 2. Similarly, the mean and standard deviation of the objective function values along with the mean of number of function evaluations is reported in Table 3 for all 20 benchmark problems.

From Table 2 it is observed that DFO takes very less computational time as compared to LEA and GAAPI for all 20 test functions. Therefore, DFO is more efficient.

From Table 3, it is found that for functions F1–F2, F4–F9, F11, F14, F16–F18 and F20, DFO gives better objective function values than GAAPI and LEA. For F10 and F12, GAAPI performs better than DFO and LEA with less average functions evaluation. But for functions F3, F13, F15 and F19 LEA performs better than DFO and GAAPI. Similarly, the standard deviation for DFO is less for most of the cases. Overall, out of 20 test problems, DFO provides much better mean function values in 14 problems and exerts less standard deviation for 16 problems. Hence DFO can be treated as an efficient algorithm over LEA and GAAPI, even in solving higher dimensional problems.

The last column of Table 3 concludes that DFO needs more number of function evaluations as compared to the rest two. However, for the functions with low dimensions like F16, F17, F18, F19 and F20; DFO can able to solve using a less number of function evaluations.

Table 3 Comparison of mean and standard deviation along with average function evaluation using DFO with others in [16]

Problem no.	Algorithm	Mean	Standard-deviation	Average number of fun evaluations	Optimum value
1	LEA	-12569.45	4.83E-04	287,365	-12,569.5
	GAAPI	-12,569.5	5.77E-07	26,510	
	DFO	-12,569.5	6.92E-09	65,610	
2	LEA	2.10E-08	3.3E-18	223,803	0
	GAAPI	1.02E-06	4.58E-09	24,714	
	DFO	2.95E-19	1.27E-18	140,812	
3	LEA	3.2E-16	3.0E-17	105,926	0
	GAAPI	0.00163	9.1E-07	19,592	
	DFO	1.26E-15	2.86E-16	149,800	
4	LEA	6.10E-06	2.5E-17	130,498	0
	GAAPI	2.2E-05	2.05E-8	29,647	
	DFO	3.94E-19	6.50E-19	65,912	
5	LEA	2.42E-06	2.27E-06	132,642	0
	GAAPI	3.65E-10	1.82E-11	60,297	
	DFO	2.58E-19	1.15E-18	149,800	
6	LEA	1.73E-04	1.21E-04	130,213	0
	GAAPI	2.23E-07	2.06E-03	26,895	
	DFO	5.32E-20	2.88E-20	149,800	
7	LEA	-93.01	0.02314	289,863	-99.2784
	GAAPI	-39.7847	0.100521	21,235	
	DFO	-94.01	1.32E-03	149,800	
8	LEA	1.63E-06	6.52E-07	189,427	0
	GAAPI	1.40E-07	2.33E-02	28,778	
	DFO	3.95E-08	5.49E-07	149,800	
9	LEA	-78.310	6.13E-03	243,895	-78.3324
	GAAPI	-78.3323	1.38E-05	28,701	
	DFO	-78.3323	1.30E-05	149,800	
10	LEA	0.5609	0.1078	168,910	0
	GAAPI	4.18E-05	5.03E-03	29,171	
	DFO	20.9571	15.717	149,800	
11	LEA	4.73E-16	6.22E-17	110,674	0
	GAAPI	6.67E-09	1.89E-04	29,199	
	DFO	1.45E-29	2.04E-29	149,800	
12	LEA	5.14E-03	4.43E-04	111,093	0
	GAAPI	1.28E-05	7.21E-07	4,149	
	DFO	6.84E-03	2.06E-03	92,876	
13	LEA	4.25E-19	4.24E-19	110,031	0
	GAAPI	0.001297	2.01E-02	30,714	
	DFO	2.31E-17	1.54E-17	149,800	
14	LEA	6.78E-18	5.43E-18	110,604	0
	GAAPI	0.000537	3.1932	31,792	
	DFO	1.53E-26	2.22E-26	149,800	

(continued)

Table 3 (continued)

Problem no.	Algorithm	Mean	Standard-deviation	Average number of fun evaluations	Optimum value
15	LEA	2.68E-16	6.26E-17	111,105	0
	GA-API	0.000298	4.01E-03	31,040	
	DFO	3.75E-16	3.06E-16	149,800	
16	LEA	-1.03108	3.36E-07	10,823	-1.03163
	GA-API	6.94E-10	1.40E-07	27,241	
	DFO	-1.03163	1.51E-15	14,800	
17	LEA	0.398	2.65E-05	10,538	0.398
	GA-API	10.22525	5.43E-07	29,625	
	DFO	0.397887	2.45E-12	14,800	
18	LEA	3.00003	6.25E-05	11,721	3
	GA-API	3.002442	2.85E-05	29,625	
	DFO	3	1.87E-15	14,800	
19	LEA	3.51E-04	7.36E-05	55,714	0.000308
	GA-API	0.051743	8.45E-07	28,654	
	DFO	5.73E-04	1.24E-08	26,800	
20	LEA	-3.301	7.83E-03	28,428	-3.32
	GA-API	-22.231	3.87E-07	29,302	
	DFO	-3.32991	1.69E-12	29,204	

6 Conclusion

This paper presents a new and novel algorithm for solving Global Optimization problem called *Drosophila* Food-Search Optimization algorithm that mimics the optimal food search mechanism of fruit fly namely '*Drosophila*'. The performance of this algorithm is evaluated over a test bed of 20 benchmark problems with a varying difficulty levels. The experimental result shows that DFO outperforms GA-API and LEA. Thus, DFO algorithm is very much flexible and efficient for solving functions with lower to higher dimensional problems. At the same time, it has a ability to overcome premature convergence to a greater extend by maintaining the diversity in the population. Further research may be carried out for solving complex real-life problems by using DFO. The upcoming paper cares to handle the constraint optimization problems again by using the food search behavior of *Drosophila*.

References

1. Alam, M.S., Islam, M., Yao, X., Murase, K.: Diversity guided evolutionary programming: a novel approach for continuous optimization. *Appl. Soft Comput.* **12**(6), 1693–1707(2012) (Elsevier)
2. Breast, J., Greiner, S., Boskovic, B., Mernik, M., Zumer, V.: Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. *Trans Evol. Comput. IEEE.* **10**, 646–657 (2006)

3. Gao, W., Liu, S.: *Comput. Oper. Res.* **39**, 687–697 (2012)
4. Santoshi, K., Arakawa, M., Yamazaki, K.: Differential evolution as the global optimization technique and its application to structural optimization. *Appl. Soft Comput.* **11**, 3792–3803 (2011)
5. Das, K.N., Singh, T.K.: Self adaptive hybridization of quadratic approximation with real coded genetic algorithm. In: *Proceedings of Seventh international conference of Bio-Inspired Computing: Theories and Application (BICTA2012). Advances in Intelligent Systems and Computing*, vol. 202, pp. 503–513. Springer, Heidelberg (2013)
6. Neshat, M., Sepidnam, G., Sargolzaei, M.: Swallow swarm optimization algorithm: a new method to optimize. *Neural Comput. Appl.* (2012). doi:[10.1007/s00521-012-0939-9](https://doi.org/10.1007/s00521-012-0939-9)
7. Dethier, V.G.: *The Hungary Fly*. Harvard University Press, Cambridge (1976)
8. Stocker, R.F.: The organization of the chemosensory system in *Drosophila melanogaster*: a review. *Cell Tissue Res.* **275**, 3–26 (1994)
9. Vosshal, L.B.: The molecular logic of olfaction in *Drosophila*. *Chemo Senses* **26**, 207–213 (2001)
10. Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., Carlson, J.R.: A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **416**, 327–338 (1999)
11. Lodish, H., Berk, A., Zipursky, L., Matsudaira, P., Baltimore, D., Darnell, J.: *Molecular cell biology*. *Cell. Signal.* 533–567 (2004)
12. Yang, Z., He, J., Yao, X.: Making a difference to differential evolution. In: Michalewicz, Z., Siaty, P. (eds.) *Advances in Meta-heuristics for Hard Optimization*, pp. 397–414. Springer, Berlin (2007)
13. Yang, Z., Tang, K., Yao, X.: Differential evolution for high-dimensional function optimization. In: *Proceedings of the IEEE Congress on Evolutionary Computation (CEC07)*, pp. 3523–3530. IEEE (2007)
14. Molina, D., Herrera, F., Lozano, M.: Adaptive local search parameters for real coded memetic algorithms. In: *Proceedings of the IEEE Congress on Evolutionary Computation (CEC2005)*, pp. 888–895. IEEE (2005)
15. Lozano, M., Herra, F., Krasnogor, N., Molina, D.: A real-coded memetic algorithms with crossover hill-climbing. *Evol. Comput.* **12**, 273–302 (2004)
16. Ciornei, I., Kyriakides, E.: Hybrid ant colony genetic algorithm (GA-API) for global continuous optimization. *IEEE Trans. Syst. Man Cybernetics: Part B Cybernetics* **42**, 234–244 (2012)
17. Deep, K., Das, K.N.: Performance improvement of real coded genetic algorithm with quadratic approximation based hybridization. *Int. J. Intell. Defense Support Syst.* **2**(4), 319–334 (2009)

Cryptanalysis of Geffe Generator Using Genetic Algorithm

Maiya Din, Ashok K. Bhateja and Ram Ratan

Abstract The use of basic crypto-primitives or building blocks has a vital role in the design of secure crypto algorithms. Such crypto primitives must be analyzed prior to be incorporated in crypto algorithm. In cryptanalysis of any crypto algorithm, a cryptanalyst generally deals with intercepted crypts without much auxiliary information available to recover plaintext or key information. As brute force attack utilizes all possible trials exhaustively, it has high computing time complexity due to huge search space and hence is practically infeasible to mount on secure crypto algorithms. The Geffe generator is a non-linear binary key sequence generator. It consists of three linear feedback shift registers and a non-linear combiner. In this paper, we attempt Geffe generator to find initial states of all three shift registers used. The initial states are the secret key bits that maintain the security of Geffe generator. To find secret key bits, one has to search huge key space exhaustively. We consider divide-and-conquer attack and genetic algorithm to reduce exhaustive searches significantly. Simulation results show that correct initial states of all shift registers can be obtained efficiently.

Keywords Geffe generator · Stream cipher · Linear feedback shift register · Genetic algorithm · Divide-and-conquer attack

M. Din (✉) · A. K. Bhateja · R. Ratan
Scientific Analysis Group, Defence Research and Development Organization, Delhi, India
e-mail: anuragimd@gmail.com

A. K. Bhateja
e-mail: akbhateja@gmail.com

R. Ratan
e-mail: ramratan_sag@hotmail.com

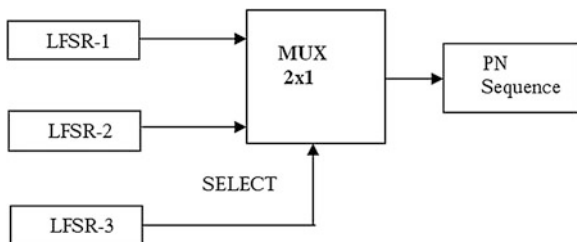
1 Introduction

Cryptology is the Art and Science of secure communications which has two major branches—cryptography and cryptanalysis. The first branch of cryptology is known as cryptography which deals with the design of cryptographic schemes/methods/systems for securing vital information. The security of information is achieved mainly through the confidentiality apart from authenticity and integrity. There are two branches of cryptography—symmetric key cryptography and asymmetric key cryptography. In symmetric key cryptography, the sender and receiver use same key for encryption as well as decryption. In asymmetric key cryptography, a pair of two keys is used, one for encryption and another for decryption. The key used for encryption is kept public and the key used for decryption is kept private. The symmetric key cryptography is known as secret key cryptography and asymmetric key cryptography is known as public key cryptography. The second branch of cryptology is known as cryptanalysis which deals with the analysis of crypto algorithms/systems to extract secret information by exploiting cryptographic weaknesses. There are active or passive cryptanalytic attacks that are applicable under certain situations. In passive attacks, an adversary or a cryptanalyst tries to extract information about key or message from given crypts by exploiting weaknesses in crypto algorithm. Such attack may be known plaintext attack, chosen plaintext attack, chosen ciphertext attack or known ciphertext attack.

Cryptanalysis also deals to evaluate cryptographic strength of crypto algorithms or crypto primitives prior to their usage in security applications. The strength of any cryptosystem depends on the characteristics of its individual components. The weaknesses of individual components can help a cryptanalyst to crack the cryptosystem. The crypto primitives and algorithms may be designed based on the concepts of block and/or stream ciphers. Nowadays, shift registers are being considered as a backbone of modern cryptography to design secure cryptosystems. A Linear Feedback shift register (LFSR) is known as a basic component of stream cipher. One may refer [1–4] for detail on such stream ciphers. This paper is concerned with the cryptanalysis of Geffe generator for identifying initial states of LFSRs.

Geffe generator [5] is a non-linear random binary key sequence generator which consists of three (LFSRs) and a nonlinear combiner. Here, we apply passive attack to identify initial states where only ciphertext is available. Various attacks have been reported in the literature on Geffe Generator such as correlation attack, algebraic attack and guessing attack [4, 6–10]. Most of the attacks mounted are exhaustive in nature and search all possible keys to find initial states of LFSRs. A divide-and-conquer attack [11, 12] is very useful to reduce exhaustive search if we divide the problem appropriately.

Genetic Algorithm (GA) is a heuristic approach which is based on the mechanisms of evolution and natural genetics. Evolutionary strategies draw inspiration from the natural search and selection processes leading to the survival of the fittest

Fig. 1 Geffe generator

individuals [13–16]. These methods have a high probability of locating the global solution optimally in the search space. In this paper, we propose a scheme which utilizes divide-and-conquer attack and GA for obtaining initial states of LFSRs of Geffe generator. Proposed scheme is adaptive and utilizes randomized key search in identifying correct initial states efficiently. The proposed scheme provides desired results in less time as discussed in simulation results. The method based on divide and conquer attack as well as genetic algorithm seems very usefull.

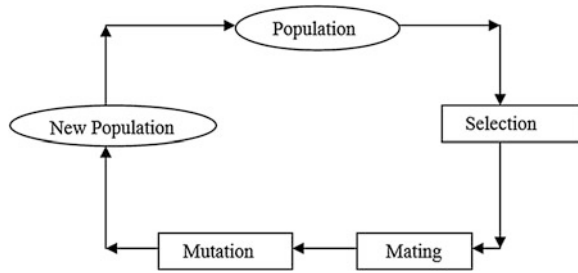
The paper is organized as follows: we describe Geffe generator in Sect. 2 and discuss Genetic Algorithm briefly in Sect. 3. We describe the proposed scheme for obtaining initial states of LFSRs of Geffe generator in Sect. 4. The initial states of LFSRs obtained as simulation results for Geffe generator are presented in Sect. 5. The paper is concluded in Sect. 6 followed by the references.

2 Geffe Generator

Geffe generator [4] is a clock controlled non-linear random binary key sequence generator utilizing three LFSRs of different length and one 2×1 multiplexer/combiner. The LFSR consists of a shift register with r cells and a linear feedback function (f). A linear combination of the contents of the sum of the cells is added modulo 2 to create a new bit according to a feedback polynomial f . The LFSR generate $\{s\}$ with period $2^r - 1$ when f is a primitive polynomial of degree r [4]. The structure of Geffe generator is shown in Fig. 1. The LFSR-3 is used as a clock controlled shift register. The Pseudorandom Number (PN) sequence bits S_i is generated by selecting the output of LFSR-1, X_i or LFSR-2, Y_i according to output of LFSR-3, Z_i . The generated PN sequence bit is selected as LFSR-1 output bit if LFSR-3 output bit is 0 otherwise LFSR-2 output bit is selected, i.e., $S_i = X_i$ when $Z_i = 0$ else $S_i = Y_i$.

Geffe generator is also defined as a nonlinear generator consisting of non-linear combing function $f(x_1, x_2, x_3) = x_1 \oplus x_1x_3 \oplus x_2x_3$ where x_1, x_2, x_3 are the output bits of LFSRs. The combining function f generates Z_i . The period and linear complexity are given as $(2^{N_1} - 1) \times (2^{N_2} - 1) \times (2^{N_3} - 1)$ and $N_1 + (N_1 + N_2) \times N_3$ respectively where, N_1, N_2 and N_3 are the length of LFSR-1, LFSR-2 and LFSR-3. The S_i is added under modulo 2 with binary plaintext bit-by-bit to generate ciphertext.

Fig. 2 Basic cycle of GA



3 Genetic Algorithms

Genetic Algorithm is a metaphoric abstraction of natural biological evolution with the basic concepts like natural selection, mating and mutation. These processes works similar to human genetics and are used to make the transition from current population to the next generation. The basic cycle of GA is shown below in Fig. 2.

Genetic Algorithms have the following basic components:

1. An initial population of random guesses of the solutions of a problem forms a set of chromosomes. A chromosome is typically encoded as a binary string with an allele value of 0 and 1 at each bit position.
2. A fitness function that assesses how effective each chromosome is at solving the problem.
3. A ‘survival of the fittest’ filter which is biased towards the selection of the better chromosomes; Fitter solutions survive while the weaker ones perish. A fitter string receives a higher number of offsprings. The chromosome with fitness rating higher than the average is attacked more than one offspring, while a chromosome with fitness value less than average is allocated less than one offspring. This is implemented through Roulette Wheel selection scheme.
4. A ‘crossover’ operator, which takes some pairs of chromosomes surviving the filter, and randomly; combines elements of each to produce ‘offspring’ with features of both ‘parent’ guesses.
5. A ‘mutation’ operator which randomly scrambles parts of some offspring, making them different from their parent chromosomes.

Steps (1)–(5) are used to generate new chromosomes which may be ‘filter’, i.e. better solutions, than those in the first, and the process is repeated over many generations. The end of process is usually dictated either by CPU time constraints, or by the emergence of an ‘evolutionary niche’ into which the chromosomes have settled. In either case, the fittest member of the final generation can be an optimal or near-optimal solution to the problem in hand.

4 Scheme of Finding Initial States

In this scheme, GA is considered and divide-and-conquer approach is applied for finding initial states. The generating polynomial (GP) of each LFSR in Geffe generator is known in our problem. Each chromosome is represented by binary bits consisting of corresponding initial bits of particular LFSR. The length of chromosome is N_i where N_i is the length of i th LFSR. The key space is 2^{N_i} for each LFSR. The fitness function to search key space is based on the linguistic characteristics of English text as monogram, bigram, trigram frequency features. The fitness function (FF) is given as $FF = \sum(f_j \times w_j)$, $j = 1-45$. We consider the most frequent features, 15 monograms, 15 bigrams and 15 trigrams, in deciphered text of length L . The value of FF for plain English text lies between 2.5 and 3.5 L .

We consider initial population of chromosomes randomly which are represented in binary form. We apply crossover, mutation and select chromosomes using Roulette Wheel Selection technique [11]. We take crossover at middle point of each pair of chromosomes and mutate bits of each chromosome randomly according to crossover and mutation probability [11]. We compute fitness value of each chromosome using above fitness function and obtain new population of chromosomes. We apply this scheme iteratively to obtain correct initial states. The time complexity in Brute force approach is $(2^{N_1} - 1) \times (2^{N_2} - 1) \times (2^{N_3} - 1)$ and it reduces to $(2^{N_1} - 1) + (2^{N_2} - 1) + (2^{N_3} - 1)$ when we consider divide-and-conquer attack. In our proposed scheme, we apply both divide-and-conquer attack and GA based adaptive search to reduce time complexity significantly and obtain initial states correctly.

5 Simulation Results

The scheme is implemented and run on 3.2 GHz machine. In our study, we have taken crypts of English text length $L = 400$ characters which are obtained by Geffe generator with different sets of LFSRs with varying length. We run the program for above crypts with crossover probability ($P_c = 0.95, 0.85, 0.75$) and mutation probability ($P_m = 0.055, 0.045, 0.035$). The simulation results are shown in Tables 1, 2 and 3 that show computational time taken for different values of P_c and P_m .

Tables 1, 2 and 3 show that the computational time decreases as the value of P_c and P_m decreases. This program provides correct results for $P_c = 0.75$ and $P_m = 0.035$ in minimum time. It is also observed that total computational time increases as length of LFSRs increases. The simulation results show that the proposed scheme is efficient and provides initial states of LFSRs correctly.

Table 1 Computational time in seconds (for $P_c = 0.95$, $P_m = 0.055$)

Length of LFSRs	LFSR-1	LFSR-2	LFSR-3	Total time
5, 7 and 11	101.24	166.12	263.21	530.57
7, 9 and 13	166.34	229.21	578.34	973.89
7, 11 and 17	167.01	256.79	1,963.51	2,387.31
9, 11 and 19	229.35	256.69	3,384.21	3,870.25

Table 2 Computational time in seconds (for $P_c = 0.85$, $P_m = 0.045$)

Length of LFSRs	LFSR-1	LFSR-2	LFSR-3	Total time
5, 7 and 11	98.43	156.56	245.01	500.04
7, 9 and 13	156.53	205.24	551.76	913.53
7, 11 and 17	156.55	245.12	1,913.57	2,315.24
9, 11 and 19	205.30	245.08	3,335.14	3,785.52

Table 3 Computational time in seconds (for $P_c = 0.75$, $P_m = 0.035$)

Length of LFSRs	LFSR-1	LFSR-2	LFSR-3	Total time
5, 7 and 11	91.24	151.13	239.25	481.62
7, 9 and 13	147.49	191.27	539.20	877.96
7, 11 and 17	149.05	236.23	1,880.35	2,265.63
9, 11 and 19	170.12	237.37	3,286.27	3,693.76

6 Conclusion

The Geffe generator has been crypt analyzed successfully for finding initial states of LFSRs. In the proposed scheme, a GAs based search approach and a divide-and-conquer attack have been used for identifying initial states of LFSRs used in Geffe generator. The simulation results shown indicates that the proposed scheme obtain initial states correctly in optimal time. The proposed scheme of finding initial states of Geffe generator can be applied to solve other non-linear random binary key sequence generators.

References

1. Stallings, W.: Cryptography and Network Security. Pearson, London (2012)
2. Menzes, A., Oorschot, P.V., Vanstone, S.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1996)
3. Geffe, P.R.: How to protect data with ciphers that are really hard to break. Electronics **46**(1), 99–101 (1973)
4. Klein, A.: Stream Ciphers. Springer, London (2013)
5. Wei, S.: On generalization of Geffe generator. Int. J. Comput. Sci. Netw. Sec. **6**(8A), 161–165 (2006)

6. Jenner, E.: Cryptanalysis of LFSR-based pseudorandom generators—a survey. http://www.erikzenner.name/docs/2004_survey_streamcipher.pdf (2004)
7. Golic, J.D.: Correlation properties of a general combiner with memory. *J. Cryptol.* **9**(2), 111–126 (1996)
8. Penzhorn, W.T., Bruwer, C.S.: New correlation attacks on stream ciphers. In: Proceedings of IEEE AFRICON 2002, pp. 203–208 (2002)
9. Meier, W., Staffelbach, O.: Fast correlation attacks on stream ciphers. In: Advances in Cryptology—EUROCRYPT ‘88, LNCS 330, pp. 301–314 (1988)
10. Kholosha, A.: Investigations in the design and analysis of key-stream generator. <http://alexandria.tue.nl/extra2/200410591.pdf>
11. Dawson, E., Clark, A.: Divide and conquer attack on certain classes of stream ciphers. *Cryptologia* **18**(1), 25–40 (1994)
12. Luengo, I.: Linear and non-linear sequences and applications to stream ciphers. *Recent Trends Crypt.—Contemp. Math.* **477**, 21–46 (2009)
13. Davis, L.: *The Handbook of Genetic Algorithms*. Van Nostrand Reinhold, New Jersey (1991)
14. Coley, D.A.: *An Introduction to Genetic Algorithms for Scientists and Engineers*. World Scientific, New Jersey (1999)
15. Haupt, R.L., Haupt, S.E.: *Practical Genetic Algorithms*. Wiley, New Jersey (1998)
16. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley, Boston (2003)

Genetic Algorithm Approach for Non-self OS Process Identification

Amit Kumar and Shishir Kumar

Abstract Computers have proved to be an inevitable part of our modern life. Every work of our modern life involve directly or indirectly the use of computers. A lot of our personal as well as confidential work related information is stored in computer systems. Therefore, it is an important task to secure this information and protect it. In this paper, the aim is to establish the sense in computer system that could differentiate between the self process (i.e. processes that are not harmful to our computer system) and the non-self process (i.e. processes that are harmful and dangerous to our computer system). A process coming in the system is identified whether the process is part of the stable system i.e. self process or is it a harmful process which can destabilize a system i.e. non-self process. This is done with the help of the detectors generated by the genetic algorithm. This technique would be used to classify the processes at process level into SELF (non-harmful) and NON-SELF (harmful or dangerous). This would help the system to sense the processes before the harmful processes do any harm to the system.

Keywords Artificial immune system · Genetic algorithm · Self · Non-self · Computer security · Detector

1 Introduction

Self awareness is one of the fundamental properties of life. Unusual program behavior often leads to data corruption, program crashes and security violations, despite these problems current computer systems have no general purpose

A. Kumar (✉) · S. Kumar
Jaypee University of Engineering and Technology, Guna, MP, India
e-mail: amitrathi10@yahoo.co.in

S. Kumar
e-mail: dr.shishir@yahoo.com

mechanism for detecting and responding to such anomalies [1]. Natural immune systems protect animals from dangerous pathogens, including bacteria, viruses, parasites and toxins. Their role in the body is analogous to that of the computer security systems in computing. Although there are many differences between living organisms and computers, the similarities are compelling and could point the way to improved computer security. The proposed solution to this is a computer that is able to recognize self. Biological and social systems of comparable and greater complexity have self and non-self identification ability which is crucial to their survival.

The analogy between computer security problems and biological processes was recognized as early as 1987. Later, Spafford [2] argued that computer viruses are a form of artificial life, and several authors investigated the analogy between epidemiology and the spread of computer viruses across networks. However, current methods for protecting computers against viruses and many other kinds of intrusions have largely failed to take advantage of what is known about how natural biological systems protect themselves from infection. Some initial work in this direction included a virus detection method based on T-cell censoring in the thymus and an integrated approach to virus detection incorporating ideas from various biological systems. However, these early efforts are generally regarded as novelties, and the principles they illustrate have yet to be widely adopted. Immunologists have traditionally described the problem solved by the immune system as that of distinguishing “self” from dangerous “other” (or “non-self”) and eliminating other. Self is taken to be the internal cells and molecules of the body, and non-self is any foreign material, particularly bacteria, parasites, and viruses. The problem of protecting computer systems from malicious intrusions [3] can similarly be viewed as the problem of distinguishing self from dangerous non-self [4]. In this case, non-self might be an unauthorized user, foreign code in the form of a computer virus or worm, unanticipated code in the form of a Trojan horse, or corrupted data.

A novel computational intelligence technique, inspired by immunology [5], has emerged, called Artificial Immune Systems [1, 6]. Artificial Immune Systems (AIS) are relatively young emerging techniques, which explore, derive and apply different biologically inspired immune mechanisms, aimed at computational problem solving [5, 7]. Several concepts from the immune have been extracted and applied for solution to real world science and engineering problems [4]. AIS also used to provide security to a computer system.

The approach is to design a tool for computer system that could distinguish a self process from the non-self process. Self process means part of the stable system and which do not harm the system [8–10]. These processes are generally user or system processes executing in a computer system. Non-self processes are the processes that use CPU resources inappropriately and are harmful to the system [8, 9]. These processes have properties of making themselves a part of the computer system, hide their original identities and replicate themselves into large number to extensively use system resources [11, 12]. Non-self means which makes the system unstable and harm it.

The aim of this paper is to describe a genetic approach (algorithm) to identify the self and non-self operating system's process. Genetic Algorithms (GAs) are based on the evolutionary ideas of natural selection and genetics. GA is always gives the optimized results [13, 14].

The relevance of the work done in this paper comes in computer security. The problem of computer security can be seen as the problem to distinguish between self and non-self. The main task is to find non-self processes, to give notification to user and pause or terminate the process. Our main aims to design such system for computer security that tries to draw an analogy between computers and biological systems. The biological equivalent for the similar task is the immune system. Our system like is modeled on the human body, is capable of recognizing and monitoring its own state of self.

2 Technology Used

The concept of Artificial Immune Systems is used in this paper with the inspiration of the natural immune system [1, 15]. The area of Artificial Immune Systems (AIS) deals with the study and development of computationally interesting abstractions of the immune system. Artificial Immune Systems are widely used for solving problems from mathematics, science, engineering, and information technology. Artificial Immune System is a part of biologically inspired computing and natural computation with interests in machine learning and it belong to the broad area of artificial intelligence [3, 15, 16]. In this paper, a approach is proposed based on the genetic algorithm, training and learning, pattern matching.

3 Proposed Idea

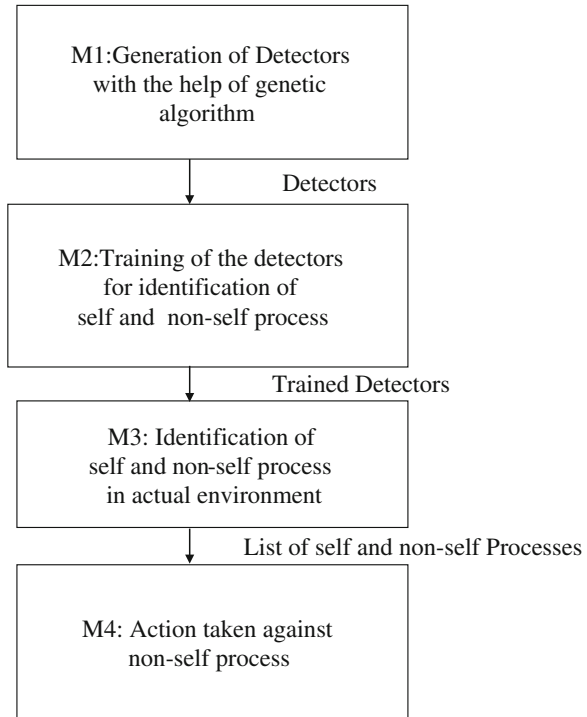
A modular design is proposed having four modules (as describe in Sect. 3.1). A fully developed system will be able to identify or isolate self and non-self processes of a computer system using the technology as described in Sect. 2.

3.1 Structure of Proposed Algorithm

The complete approach of process identification can be divided into four main modules as shown in Fig. 1. These four modules are:

- Generation of Detectors with the help of genetic algorithm.
- Training of the detectors for identification of self and non-self process.
- Identification of self and non-self process in actual environment.
- Action taken against non-self process.

Fig. 1 Structure of the proposed algorithm



The output of the first module are the detectors which are generated by using the genetic algorithm. These detectors are the input of second module. In second module we provide the training to these detectors. The output of second module are the trained detectors. Training is provided to identify the non-self operating system processes. Now these trained detectors become the input of third module. In third module the trained detectors work in actual environment. Third module provides a list of self and non-self operating system processes as output of this module. The output of third module will become the input of last module. In fourth module the user takes action on non-self processes. User can suspend, terminate or kill the non-self processes. User can resume the non-self processes and define this non-self process as self.

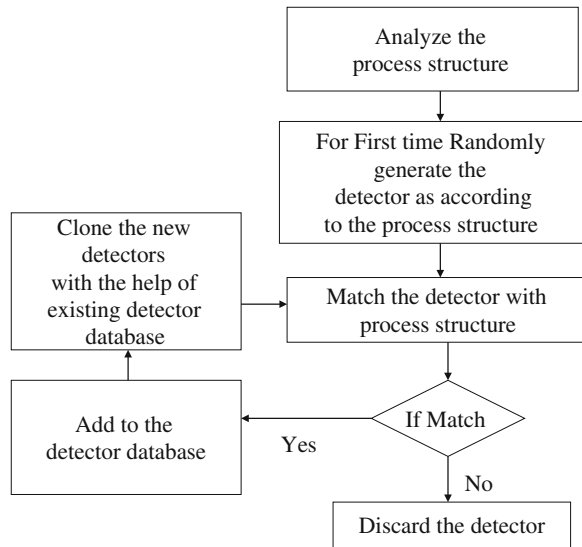
3.1.1 Module 1: Generation of Detectors

This module deals with generation of detectors which will be used to detect any instance of abnormal behavior by the system. The generation of detectors will be done using a genetic algorithm. This module will have two sub modules:

- **Detector Generator**

Detectors are generated with the help of genetic algorithm. A Genetic algorithm will generate detectors randomly. The detectors are not present in the data collected for the process but they are very close to the original sequences.

Fig. 2 Flow diagram for module 1



• Detector Verification

Each detector will be first matched with all the recorded sequences. The detector will only be accepted if it does not matches any of the recorded sequences, if it matches any of the sequences then it will be rejected. This is done to protect against auto-immunity i.e. the detector should not match with a sequence of normal behavior.

Figure 2 shows the working of first module. In this figure first we analyze the process and its structure. The main focus is on the process parameters. Parameters which are not null and have useful information of process like process id, file size, base address, memory usage, number of page faults, page file usage etc. Then the first set of detectors are randomly generated as according to the selected process parameters and process structure. These randomly generated detectors then match with the process structure for checking the range and size of parameters. If there is a match the randomly generated detectors are selected for next module. If there is miss match then these randomly detectors are discarded. To generate the new set of detectors we use the genetic operators.

Algorithm of module 1

- Step 1: Analyze the process structure of a process for important and relevant parameters.
- Step 2: Generate the detectors randomly for creating the first detector set according to the process structure.
- Step 3: The detectors generated in Step 2 are matched with the process structure.
 - Step 3.1: If the detector does not match with the process structure then it is discarded.

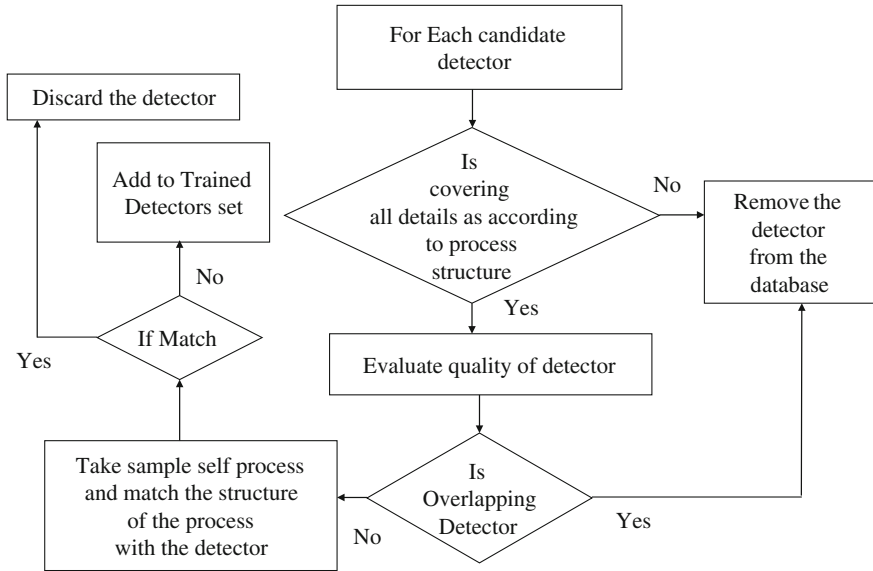


Fig. 3 Flow diagram for module 2

- Step 3.2: If the detector matches with the process structure then it is added to the detector database.
- Step 4: If the detector is added to the detector database then the new detectors are cloned to the existing detector database.
- Step 5: Go to Step 3 until sufficient detector database is generated.

3.1.2 Module 2: Training of the Detectors for Identification of Self and Non-self Process

- In this module, evaluate the detector set and necessary training is provided so that it covers all details as according to the process structure. A complete trained detector set would be generated at the end of this module.

Figure 3 shows the working of module 2. In this module fitness of every detectors generated in module 1 are checked. It may happen that some detectors have same parameters values, so these detectors are discarded. After checking fitness of detectors training is provided to detectors by using sample processes.

Algorithm of module 2

- Step 1: Take each of the candidate detectors one by one and check whether the detector is covering all the details as according to the process structure.

- Step 1.1: If the candidate detector does not cover all the details then it is removed from the detector database.
- Step 1.2: If the candidate detector covers all the details then it is evaluated for its quality.
- Step 2: For evaluating the quality of the detector, it is checked that if it is overlapping with another detector or not.
 - Step 2.1: If the detector overlaps, then it is removed from the detector database.
 - Step 2.2: If it doesn't overlap then go to next Step No 3.
- Step 3: Take Sample Self processes and match the process structure of these sample self process with the candidate non-overlapping detector.
 - Step 3.1: If it matches then the candidate detector is discarded.
 - Step 3.2: If it doesn't match then this candidate detector is added to the trained detector set.

3.1.3 Module 3: Identification of Self and Non-self Process in Actual Environment

- In this module, the system would be performed against a process in the actual environment. If the process is detected by the detector then it would be treated as a NON-SELF process and the User would be intimated about the result.

Figure 4 shows the working of the module 3. The trained detectors from module 2 are used to identify the non-self processes in actual environment. If any process is identified as non-self then we check that the child the non-self processes first.

Algorithm of module 3

- Step 1: Take the trained detectors.
- Step 2: Take a process from the actual current environment.
- Step 3: Check whether the process is identified by any of the trained detectors.
 - Step 3.1: If the process is not identified by the trained detectors then leave that process and go to Step 5.
 - Step 3.2: If the process is identified by the trained detectors then this process is added to a list of non-self processes and the user is informed to take an action.
- Step 4: The process in the non-self list is checked whether it has created child processes or not.
 - Step 4.1: If yes, then take that child process and proceed from Step No. 3
 - Step 4.2: If not, then go to Step No. 2.
- Step 5: Go to Step 2 until all the processes have been checked once.

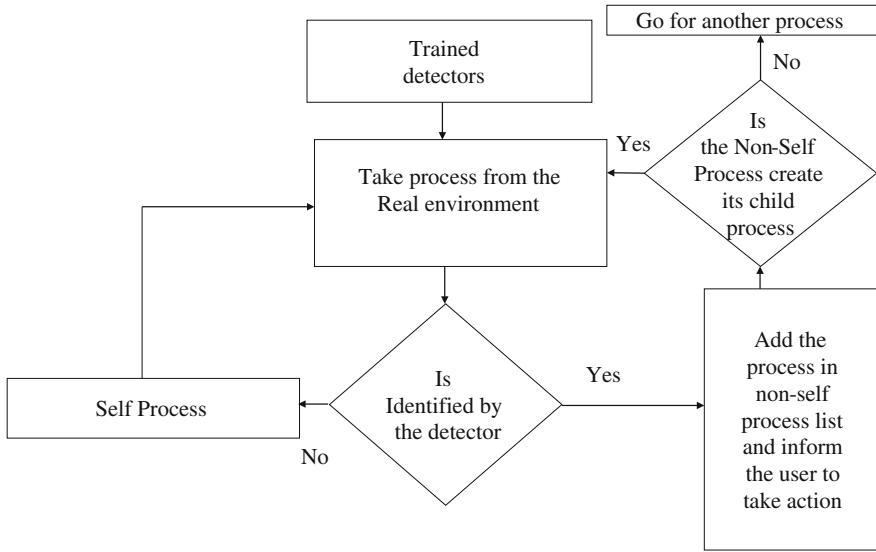


Fig. 4 Flow diagram module 3

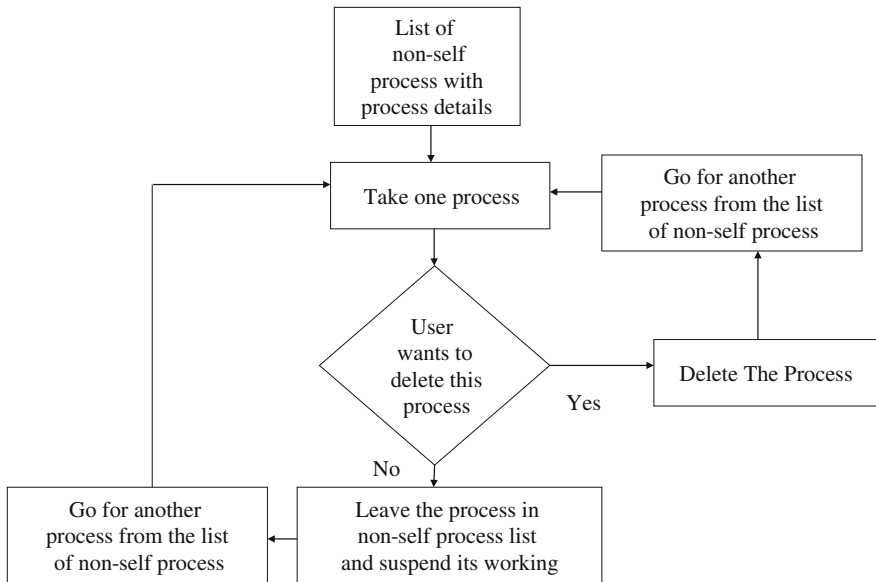


Fig. 5 Flow diagram module 4

3.1.4 Module 4: Action Taken Against the Non-self Processes

- In this last module, an action environment is provided to the user where he can select the action to be taken against the Non- Self processes. The Actions that a user can take includes:-
 - Deletion of the Non-Self Process.
 - Suspension of working of the Non-Self Process.

Figure 5 shows the user's action on non-self processes. User can delete or suspend the process.

Algorithm of module 4

- Step 1: Take one process from the list of non-self processes.
- Step 2: Ask the user whether he wants to delete this process or not.
 - Step 2.1: If the user wants to delete the non-self process then delete the process and go to Step No 3.
 - Step 2.2: If the user does not want to delete the non-self process under consideration then the process is left in the non-self processes list with its working being suspended.
- Step 3: Go for another process from the list of non-self processes.
- Step 4: Go to Step 2.
- Step 5: Repeat until the user is asked for his choice for all the non-self processes.

4 Comparison with Existing Approaches

There are various approaches to provide the security to the system on various levels. These approaches can be divided into two parts hardware based and software based security. For a hardware tool certain disk/data locks tool are available. For software tool there is firewall very good tool for network security in today's world of internet. Firewall is a filter that prevents fraud websites from accessing your computer and damaging the data. However, a firewall is not a great option for securing the servers on the internet because the main objective of a server is granting access to unknown users to connect to various web pages. Along with firewall, most of the users try installing a good anti-virus and security software to enhance the security level of your computer system. Users have to update the firewall and anti-virus on regularly bases. It may happen that some unwanted files (virus, worms etc.) can enter into the system and execute and make process. Then the system will be affected by it. To provide the security at the lowest level (process level) the approach proposed in the paper will be more effective than the other approaches as it works on processes parameters to identify the non-self (harmful) processes.

Table 1 Comparison of proposed approach with some anti-virus tools

Anti-virus	Scan time	Accuracy	Detection rate	Performance lag	Signature based detection	Regular updating required
AVG anti-virus	High	High	Normal	Yes	Yes	Yes
Norton anti-virus	High	High	High	Yes	Yes	Yes
Avast anti-virus	High	Medium	Normal	Yes	Yes	Yes
Microsoft security essentials	Average	High	High	Yes	Yes	Yes
Proposed approach	Low	Very High	Very High	Yes	No	No

Table 1 shows the comparison proposed approach with some existing anti-virus tool available in market. After analyzing it is found that the proposed approach is better than the others.

5 Conclusion and Future Work

This paper deals with the identification of self and non-self process to provide highest level of security for an computer system. The proposed technology can be applied to any type of computer system to provide the highest security. Since the work is done at the process level so higher level of security is provided.

In this paper, the work is divided into four modules, which are generation of detectors, training of the detectors, identification of the self and non-self processes, and the notification to the user. In the first module, a set of detectors are randomly generated according to the process structure of a self process. Then, in the second module, this detector set was trained to identify the non-self processes in the real environment and training the detector to identify the non-self processes. In the next module, this trained detector set was used in the real environment to identify the non-self processes and in the last module, the user is intimidated about the non-self processes and is given a choice of whether he wants to delete the process or suspend the working of the non-self process.

Since there are not many applications available on this concept, our system is relatively new. It is capable of identification of self and non-self processes and it notifies user. It monitors the specified process of the system for changes which might make the process harmful or due to which the process might be affecting the stability of the system. The system uses genetic algorithm. It does away with the idea of using virus databases for identification of viruses.

The plan is to design a tool based on the algorithms for generation of detectors, training and identifying self and non-self process and its subparts. After the full design of tool, the performance of the system will be analyzed.

References

1. Hofmeyr, S.A., Forrest, S.: Architecture of an artificial immune system. *Evol. Comput.* **8**(4), 443–473 (2000). doi:[10.1162/106365600568257](https://doi.org/10.1162/106365600568257)
2. Spafford, E.H.: Computer viruses as artificial life. *J. Artif. Life* 1(3), 249–265 (1994). (<http://spaf.cerias.purdue.edu/tech-reps/985.pdf>) (MIT Press)
3. Dasgupta, D. (ed.): *Artificial Immune Systems and their Applications*. Springer, Berlin (1999)
4. Lau, L.W.Y.: Computer immunology, The Department of Information Technology and Electrical Engineering, The University of Queensland. [http://download.adamas.ai/dbase/ebooks/VX_related/Computer%20Immunology\(Lau\).pdf](http://download.adamas.ai/dbase/ebooks/VX_related/Computer%20Immunology(Lau).pdf) (2002)
5. Farmer, J.D., Packard, N.H., Perelson, A.S.: The immune system, adaptation and machine learning. In: *Proceedings of the Fifth Annual International Conference in Physica D: Nonlinear Phenomena*, vol. 22, no. 1–3, pp. 187–204. Elsevier, Amsterdam (1986)
6. Bachmayer, S. (eds.): *Artificial immune system*. In: ch XI, pp. 209–230. Group Idea Publishing Hershey (2006) <http://www.cs.helsinki.fi/u/niklande/opetus/SemK07/paper/bachmayer.pdf>
7. Carter, J.H.: The immune system as a model for pattern recognition and classification. *J. Am. Med. Inf. Assoc.* **7**(3), 28–41(2000)
8. Percus, J.K., Percus, O.E., Perelson, A.S.: Predicting the size of the antibody-combining region from consideration of efficient self/nonself discrimination. In: *Proceedings of the National Academy of Science*, vol. 90, pp. 1691–1695. National Acad Sciences (1993)
9. Forrest, S., Hofmeyr, S.A., Somayaji, A.B., Longstaff, T.A.: A sense of self for UNIX processes, submitted to the IEEE Symposium on Security and Privacy (1996)
10. Forrest, S., Perelson, A.S.: Self non-self discrimination in a computer. In: *Proceedings of the IEEE Symposium on Research in Security and Privacy IEEE* (1994)
11. Niu, J.: Process Description and Control, CSC33200: Operating Systems, CS-CCNY, Fall. <http://www.sci.brooklyn.cuny.edu/~jniu/teaching/csc33200/csc33200.html> (2003)
12. Percus, J.K., Percus, O.E., Perelson, A.S.: Probability of self-nonself discrimination. In: *Theoretical and Experimental Insights into Immunology*, series vol. 66, pp. 63–70. NATO ASI (1992)
13. Eiben, A.E., Raue, P.E., Ruttkay, Z.: Genetic algorithms with multi-parent recombination. In: *PPSN III: Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel Problem Solving from Nature*, vol. 866, pp. 78–87. Springer, Berlin (1994)
14. Rojas, R.: *Genetic Algorithm from Neutral Networks: A Systematic Introduction*, pp. 427–448. Springer, Berlin (1996)
15. Dal, D., Abraham, S., Abraham, A., Sanyal, S, Sanglikar, M.: Evolution induced secondary immunity: an artificial immune system based intrusion detection system. *IEEE 7th Computer Information Systems and Industrial Management Applications*. doi: [10.1109/CISIM.2008.31](https://doi.org/10.1109/CISIM.2008.31) pp. 65–70. IEEE (2008)
16. Darmoul, S., Pierreval, H., Gabouj, S.H.: scheduling using artificial immune system metaphors: a review. In: *Proceedings of the International Conference on Service Systems and Service Management*, pp. 1150–1155. IEEE (2006)
17. Matloff, N.: Unix processes, Department of Computer Science University of California at Davis. <http://heather.cs.ucdavis.edu/~matloff/unix.html> (2004)

Gbest-Artificial Bee Colony Algorithm to Solve Load Flow Problem

N. K. Garg, Shimpi Singh Jadon, Harish Sharma and D. K. Palwalia

Abstract Load flow problem has a great significance to analyze the power system network due to its roll in planning and operation of network. Generally, Newton–Raphson (NR) method is used to analyze the load flow problems due to its efficiency and accuracy. But, NR method has some inherent drawbacks like inefficient for highly loaded network, assumption are required for initial values and abnormal operating conditions. To overcome the existing drawbacks of NR method, a recently developed swarm intelligence based algorithm, namely Gbest guided Artificial Bee Colony algorithm (GABC) is applied to solve the load flow problem for five bus network. The reported results of GABC are compared to the results of NR method and basic ABC algorithm, which show that the accuracy of unknown parameters such as voltage, angle and power produced by GABC is competitive method to the NR method and basic ABC algorithm based method.

Keywords Optimization • Swarm intelligence based algorithms • Load flow problem • Artificial bee colony

N. K. Garg (✉) · D. K. Palwalia
University College of Engineering, Rajasthan Technical University, Kota, India
e-mail: nk_garg@yahoo.com

D. K. Palwalia
e-mail: dheerajpalwalia@gmail.com

S. S. Jadon
ABV-Indian Institute of Information Technology, Gwalior, India
e-mail: shimpisingh2k6@gmail.com

H. Sharma
Harish Sharma Government Engineering College, Jhalawar, India
e-mail: harish.sharma0107@gmail.com

1 Introduction

```

k ← number of clusters
f ← number of dimensions
for n = 1 : k do
  for d = 1 : f do
    x1=rand();
    Position(n,d)=xmin(d)+x1*(xmax(d)-xmin(d));
  end for
end for

```

Load flow solutions are needed for planning and operation of power system. In planning and operation, the voltage profile, power transfer from one branch to another, reactive power and line losses etc. are analyzed. Load flow analysis is required to plan new network or extending the existing network by adding new generator sites, satisfying increase load demands and locating new transmission lines. The load flow equations, which are non linear in nature, are generally solved by Newton–Raphson, Gauss Siedel and Fast Decoupled methods [1]. The Newton–Raphson (NR) method is very popular and mostly used to solve load flow equations due to high convergence rate, but it has some limitations also for e.g. its performance is dependent on initial values of the network, difficult to determine the normal operating solutions and large complex power system [2].

Researchers are continuously working to improve the accuracy of load flow problem solutions [3–6]. Meta-heuristic search methods have also been applied to solve the load flow and parameter estimation problems [7–10]. Kim et al. [2] applied Genetic algorithm (GA) to solve the load flow problem. They generated, both normal solution and abnormal solution for the heavy load network using GA. Mohmood and Kubba [11] also applied GA to solve multiple load flow solution problem. In their proposed work, five busbars typical test system and 362-bus Iraqi National Grid are used to demonstrate the efficiency and performance of the proposed method. Wong and Li [12] proposed a new genetic-based algorithm, namely GALF for solving the load-flow optimization problem. The main objective of the proposed algorithm is to minimize the total mismatch in the nodal powers and voltages. Through extensive experiments, they claimed that GALF successfully determined both normal and abnormal solutions with mismatches in the vicinity of zero. El-Dib et al. [13] applied hybrid particle swarm optimization algorithm to solve the load flow problem. Further, Esmine et al. [14] proposed a new variant of PSO and applied it to solve loss minimization based optimal power flow problem. Dan Cristian et al. [15] applied PSO to obtain the optimal power flow and for transmission expansion planning of the network. Salomon et al. [16] also proposed a PSO based method for load flow problem. In their proposed methodology, the objective function is based on the minimization of power mismatches in the system buses. Esmine and Lambert-Torres [4] also used PSO to determine control variable settings, such as the number of shunts to be switched, for real power loss minimization in the transmission system.

In this paper, swarm intelligence based algorithms, namely artificial bee colony (ABC) [17] and Gbest-guided artificial bee colony (GABC) [18] algorithms have been used to solve load flow problem of 5 bus network. Through, simulation on 5 bus system, it is shown that GABC algorithm gives better result than the ABC and NR method.

The paper is further organized as follows: Sect. 2 formulate the load flow problem as an optimization problem. ABC algorithm and its advanced variant, namely GABC algorithm are described in Sect. 3. Section 4 shows the experimental results for the five bus network. At last, in Sect. 5, paper is concluded.

2 Formulation of Load Flow as an Optimization Problem

The load flow equations are simply the power balance equations at each bus, both active and reactive powers. The power balance equation expresses the fact that there are no power lose in any bus, which means that the input power to a bus equals the output power from that bus. Therefore, the objective of the load flow is to find the voltage magnitudes and angles of the different system buses that minimize the difference between the input power and the output power from the bus. So the load flow problem can be formulated as an optimization problem.

The load flow problem is solved by taking a single phase model and it is assumed to be operating under balanced condition. There are four quantities associated with each bus. These are voltage V , phase angle δ , real power P and reactive power Q . Buses in the power system are classified into three categories [1] as:

- Slack Bus: This bus is also known as swing bus and taken as reference bus. The magnitude of voltages and phase angles are specified.
- Generator Bus (PV): The real power and voltage magnitude are specified, while the phase angle and reactive power are unknown.
- Load bus (PQ): The active and reactive powers are specified. The magnitude and phase angle of the bus voltage are unknown.

Consider that there are total n number of nodes in a power system network. At any node i , the nodal active power P_i and reactive power Q_i are given by [19]:

$$P_i = E_i \sum_{j=1}^n (G_{ij}E_j - B_{ij}F_j) + F_i \sum_{j=1}^n (G_{ij}F_j + B_{ij}E_j), \forall i = 1, 2, 3, \dots, n \quad (1)$$

$$Q_i = F_i \sum_{j=1}^n (G_{ij}E_j - B_{ij}F_j) - E_i \sum_{j=1}^n (G_{ij}F_j + B_{ij}E_j), \forall i = 1, 2, 3, \dots, n \quad (2)$$

where G_{ij} and B_{ij} are the (i, j) element of the admittance matrix. The bus admittance matrix Y_{bus} is in order of $(n \times n)$, where n is the number of buses. E_i and F_i

are real and imaginary part of the voltage at node i . The magnitude of voltage V_i is calculated as,

$$V_i = \sqrt{E_i^2 + F_i^2} \quad (3)$$

The unknown variables in the network are,

- At PQ nodes, voltages are unknown and the differences in active and reactive powers are given by

$$\Delta P_i = |P_i^{sp} - P_i| \quad (4)$$

$$\Delta Q_i = |Q_i^{sp} - Q_i| \quad (5)$$

- At PV nodes, the reactive power, real and imaginary parts of the voltage are unknown but the magnitude of voltage is known, so the difference in voltages can be calculated as follows

$$\Delta V_i = |V_i^{sp} - V_i| \quad (6)$$

As a balanced network system requires the differences, shown in Eqs. (4), (5) and (6), to be zero, therefore to meet the required condition, the unknown variables at PQ and PV nodes are calculated and respective values are assigned in Eqs. (4) to (6). The load flow problem is formulated as an optimization problem and the designed objective function is described below [19]:

$$\text{Minimize } H = \sum_{i \in n_{PQ} + n_{PV}} |P_i^{sp} - P_i|^2 + \sum_{i \in n_{PQ}} |Q_i^{sp} - Q_i|^2 + \sum_{i \in n_{PV}} |V_i^{sp} - V_i|^2 \quad (7)$$

where n_{PQ} and n_{PV} are the total number of PQ and PV nodes respectively.

3 Artificial Bee Colony (ABC) Algorithm

In ABC, honey bees are classified into three groups namely employed bees, onlooker bees and scout bees. The number of employed bees are equal to the onlooker bees. The employed bees are the bees which searches the food source and gather the information about the quality of the food source. Onlooker bees stay in the hive and search the food sources on the basis of the information gathered by the employed bees. The scout bee, searches new food sources randomly in place of the abandoned foods sources. Similar to the other population-based algorithms, ABC solution search process is an iterative process. After, initialization of the ABC parameters and swarm, it requires the repetitive iterations of the three phases

namely employed bee phase, onlooker bee phase and scout bee phase. Each of the phase is described as follows:

3.1 Initialization of the Swarm

Initially, a uniformly distributed swarm of SN food sources where each food source x_i ($i = 1, 2, \dots, SN$) is a D -dimensional vector is generated. Here D is the number of variables in the optimization problem and x_i represent the i th food source in the swarm. Each food source is generated as follows:

$$x_{ij} = x_{min\ j} + rand[0, 1] (x_{max\ j} - x_{min\ j}) \quad (8)$$

here $x_{min\ j}$ and $x_{max\ j}$ are bounds of x_i in j th direction and $rand[0, 1]$ is a uniformly distributed random number in the range $[0, 1]$.

3.2 Employed Bee Phase

The position update equation for i th candidate in this phase is

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \quad (9)$$

here $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, D\}$ are randomly chosen indices. k must be different from i . ϕ_{ij} is a random number between $[-1, 1]$. After generating new position, a greedy selection is applied between the newly generated position and old one and the better one position is selected.

3.3 Onlooker Bees Phase

In this phase, the fitness information (nectar) of the new solutions (food sources) and their position information is shared by all the employed bees with the onlooker bees in the hive. Onlooker bees analyze the available information and selects a solution with a probability $prob_i$ related to its fitness, which can be calculated using following expression (there may be some other but must be a function of fitness):

$$prob_i(G) = \frac{0.9 \times fitness_i}{max\ fit} + 0.1, \quad (10)$$

here $fitness_i$ is the fitness value of the i th solution and $max\ fit$ is the maximum fitness amongst all the solutions. As in the case of employed bee, it produces a modification on the position in its memory and checks the fitness of the new

solution. If the fitness is higher than the previous one, the bee memorizes the new position and forgets the old one.

3.4 Scout Bees Phase

A food source is considered to be abandoned, if its position is not getting updated during a predetermined number of cycles. In this phase, the bee whose food source has been abandoned becomes scout bee and the abandoned food source is replaced by a randomly chosen food source within the search space. In ABC, predetermined number of cycles is a crucial control parameter which is called *limit* for abandonment.

Assume that the abandoned source is x_i . The scout bee replaces this food source by a randomly chosen food source which is generated as follows:

$$x_{ij} = x_{min\ j} + rand[0, 1](x_{max\ j} - x_{min\ j}), \text{ for } j \in \{1, 2, \dots, D\} \quad (11)$$

where $x_{min\ j}$ and $x_{max\ j}$ are bounds of x_i in j th direction.

3.5 Main Steps of the ABC Algorithm

Based on the above explanation, it is clear that the ABC search process contains three important control parameters: The number of food sources SN (equal to number of onlooker or employed bees), the value of *limit* and the maximum number of iterations. The pseudo-code of the ABC is shown in Algorithm 1 [20].

In 2010, Zhu and Kwong [18] proposed an improved ABC algorithm called GABC algorithm by incorporating the information of global best (gbest) solution into the solution search equation to improve the exploitation. GABC is inspired by PSO [21], which, in order to improve the exploitation, takes advantage of the information of the global best (gbest) solution to guide the search by candidate solutions. They modified the solution search equation of ABC as follows:

Algorithm 1 Artificial Bee Colony Algorithm:

Initialize the parameters;

while Termination criteria is not satisfied **do**

 Step 1: Employed bee phase for generating new food sources;

 Step 2: Onlooker bees phase for updating the food sources depending on their nectar amounts;

 Step 3: Scout bee phase for discovering the new food sources in place of abandoned food sources;

 Step 4: Memorize the best food source found so far;

end while

Output the best solution found so far.

Fig. 1 Five bus network

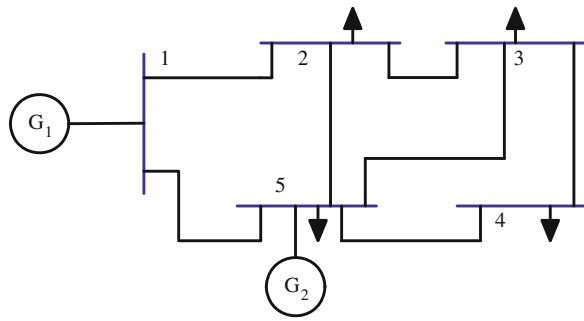


Table 1 Line impedance and line charging data for the system

Line bus to bus	Impedance	Line charging $Y/2$
1-2	$0.02 + j 0.10$	$j 0.030$
1-5	$0.05 + j 0.25$	$j 0.020$
2-3	$0.04 + j 0.20$	$j 0.025$
2-5	$0.05 + j 0.25$	$j 0.020$
3-4	$0.05 + j 0.25$	$j 0.020$
3-5	$0.08 + j 0.40$	$j 0.010$
4-5	$0.10 + j 0.50$	$j 0.075$

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) + \psi_{ij}(y_j - x_{ij}) \tag{12}$$

where the third term in the right-hand side of Eq. (12) is a new added term called gbest term, y_j is the j th element of the global best solution, ψ_{ij} is a uniform random number in $[0, C]$, where C is a non negative constant. According to Eq. (12), the gbest term can drive the new candidate solution towards the global best solution, therefore, the modified solution search equation described by Eq. (12) can increase the exploitation of ABC algorithm. Note that the parameter C in Eq. (12) plays an important role in balancing the exploration and exploitation of the candidate solution search. In this paper GABC is used to solve the load flow problem for five bus network system.

4 Results Analysis and Discussion

ABC and GABC algorithms are tested on five bus network which is shown in Fig. 1. Test network has 2 generators and 3 load buses. In this network, bus 1 is the slack bus, while bus 5 is the PV bus. Buses 2, 3 and 4 are PQ buses. A C++ program has been developed to analyze the ABC and GABC method over this problem. The line impedances and the line charging admittances are given in Table 1, which are used to calculate Y_{bus} matrix. Table 2 shows its initial values of bus voltages, angles, load connected to buses and powers to generators.

Table 2 Initial data for Bus voltages, power generated and load

Bus number	Bus voltage		Power generated		Load	
	Magnitude (pu)	Angle (deg)	P (MW)	Q (MVar)	P (MW)	P (MVar)
1	1.05	0	–	–	0	0
2	1	0	0	0	96	62
3	1	0	0	0	35	14
4	1	0	0	0	16	8
5	1.02	0	48	–	24	11

pu: per unit, deg: degree, MW: mega watt, MVar: mega volt ampere

Table 3 Result comparison amongst NR, ABC and GABC Method, pu: per unit, deg: degree

Bus Number	NR		ABC		GABC	
	Magnitude (pu)	Angle (deg)	Magnitude (pu)	Angle (deg)	Magnitude (pu)	Angle (deg)
1	1.05	0	1.05	0	1.05	0
2	0.9826	–5.0124	0.98554	–6.28181	0.984571	–5.01865
3	0.9777	–7.1322	0.991387	–11.2626	0.980781	–7.12979
4	0.9876	–7.3705	1.004046	–10.8028	0.991064	–7.30659
5	1.02	–3.2014	1.027102	–5.03639	1.02519	–3.24332

4.1 Parameter Setting

To solve the load flow problem using GABC and ABC, following experimental setting is adopted:

- Colony size $NP = 50$ [22, 23],
- $\phi_{ij} = rand[-1, 1]$,
- Number of food sources $SN = NP/2$,
- $limit = D \times SN$ [24, 25],
- The stopping criteria is either maximum number of function evaluations (which is set to be 200000) is reached or the acceptable error (1.0×10^{-05}) has been achieved,
- The number of simulations/run = 30,
- $C = 1.5$ [18].

4.2 Results Analysis

Initial values of the five bus network parameters are shown in Table 2. The Numerical results with parameter setting, mentioned in Sect. 4.1 are given in Table 3. Table 3 shows that the bus voltages for ABC and GABC have reached more close to accurate solutions than the NR method for the problem. GABC and ABC algorithms achieved the objective function error (refer Eq. (7)) of the order

of 10^{-5} , while the NR method achieved of 10^{-4} order. For the balanced condition of network, the power generated P should be 123 MW, while it is obtained 161 MW using ABC algorithm, 126.4 MW using GABC and 127.1 MW using NR method. Therefore, it is clear that the P generated by GABC algorithm is more close to the required power to balance the network. So, it can be analyzed from the results that the GABC can be considered a competitive method to solve the load flow problem.

5 Conclusion

This paper shows an efficient solution to the load flow problem of five bus system, using GABC. The reported results are analyzed and compared with the basic ABC and Newton–Raphson (NR) method. It is clear from the results analysis that the voltage produced for each bus by GABC is much similar to the reference value as compared to the basic ABC and NR method, while the active power produced by GABC is much close to the active power required to balance the network.

As the GABC method is independent from initial setting of parameters values, therefore, the proposed method can also be applied to solve the load flow problem for large, complex, and unbalanced network. Further, it is clear from results analysis that the accuracy for active power produced by GABC method can also be improved.

References

1. Saadat, H.: Power system analysis. WCB/McGraw-Hill, Singapore (1999)
2. Kim, H., Samann, N., Shin, D., Ko, B., Jang, G., Cha, J.: A new concept of power flow analysis. *J. Electr. Eng. Technol.* **2**(3), 312 (2007)
3. Teng, Jen-Hao: A direct approach for distribution system load flow solutions. *Power Delivery, IEEE Trans.* **18**(3), 882–887 (2003)
4. Esmín, A.A., Lambert-Torres, G.: Application of particle swarm optimization to optimal power systems. *Int. J. Innov. Comput. Inf. Control* **8**, 1705–1716 (2012)
5. Abido, M.A.: Optimal power flow using particle swarm optimization. *Int. J. Electr. Power Energy Syst.* **24**(7), 563–571 (2002)
6. Romero, A.A., Zini, H.C., Rattá, G., Dib, R.: A fuzzy number based methodology for harmonic load-flow calculation, considering uncertainties. *Latin Am. Appl. Res.* **38**(3), 205–212 (2008)
7. Wong, K.P., Li, A., Law, T.M.Y.: Advanced constrained genetic algorithm load flow method. In: *Generation, Transmission and Distribution, IEE Proceedings*, vol. 146, pp. 609–616. IET (1999)
8. Ting, T.O., Wong, K.P., Chung, C.Y.: Hybrid constrained genetic algorithm/particle swarm optimisation load flow algorithm. *IET Gener. Transm. Distrib.* **2**(6), 800–812 (2008)
9. Wong, K.P., Yuryevich, J., Li, A.: Evolutionary-programming-based load flow algorithm for systems containing unified power flow controllers. *IEE Proc. Gener., Transm. Distrib.* **150**(4), 441–446 (2003)

10. Mori, H., Sone, Y.: Tabu search based meter placement for topological observability in power system state estimation. In: Transmission and Distribution Conference, 1999 IEEE, vol. 1, pp. 172–177. IEEE (1999)
11. Mahmood, S.S., Kubba, H.A.: Genetic algorithm based load flow solution problem in electrical power systems. *J. Eng.* **15**, 1 (2009)
12. Wong, K.P., Li, A.: Solving the load-flow problem using genetic algorithm. In: Evolutionary Computation, 1995, IEEE International Conference on, vol. 1, p. 103. IEEE (1995)
13. El-Dib, A.A., Youssef, H.K., El-Metwally, M.M., Osman, Z.: Load flow solution using hybrid particle swarm optimization, ICEEC 2004. In: International Conference Electrical Electronic and Computer Engineering, pp. 742–746 (2004)
14. Esmiri, A.A., Lambert-Torres, G., Antonio, C., Zambroni de Souza, A.C.: A hybrid particle swarm optimization applied to loss power minimization. *Power Syst., IEEE Trans.* **20**(2), 859–866 (2005)
15. Dan Cristian, P., Barbulescu, C., Simo, A., Kilyeni, S., Solomonesc, F.: Load flow computation particle swarm optimization algorithm. In: Universities Power Engineering Conference (UPEC), 2012 47th International, pp. 1–6. IEEE (2012)
16. Salomon, C.P., Lambert-Torres, G., Martins, H.G., Ferreira, C., Costa, C.I.: Load flow computation via particle swarm optimization. In: Industry Applications (INDUSCON), 2010 9th IEEE/IAS International Conference on, pp. 1–6. IEEE (2010)
17. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Technical Report TR06, Erciyes University Press, Erciyes (2005)
18. Zhu, G., Kwong, S.: Gbest-guided artificial bee colony algorithm for numerical function optimization. *Appl. Math. Comput.* **217**(7), 3166–3173 (2010)
19. Wong, K.P., Li, A., Law, M.Y.: Development of constrained-genetic-algorithm load-flow method. In: Generation, Transmission and Distribution, IEE Proceedings, vol. 144, pp. 91–99. IET, (1997)
20. Karaboga, D., Akay, B.: A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
21. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Neural Networks, 1995. Proceedings., IEEE International Conference on, vol. 4, pp. 1942–1948. IEEE (1995)
22. Diwold, K., Aderhold, A., Scheidler, A., Middendorf, M.: Performance evaluation of artificial bee colony optimization and new selection schemes. *Memetic Comput.* 1–14 (2011)
23. El-Abd, M.: Performance assessment of foraging algorithms vs. evolutionary algorithms. *Inf. Sci.* **182**(1), 243–263 (2011)
24. Karaboga, D., Basturk, B.: Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. In: Foundations of Fuzzy Logic and Soft Computing, pp. 789–798 (2007)
25. Akay, B., Karaboga, D.: A modified artificial bee colony algorithm for real-parameter optimization. *Inf. Sci.* doi: [10.1016/j.ins.2010.07.015](https://doi.org/10.1016/j.ins.2010.07.015) (2010)

Neural Network Based Dynamic Performance of Induction Motor Drives

P. M. Menghal and A. Jaya Laxmi

Abstract In industries, more than 85 % of the motors are Induction Motors, because of the low maintenance and robustness. Maximum torque and efficiency is obtained by the speed control of induction motor. Using Artificial Intelligence (AI) techniques, particularly the neural networks, performance and operation of induction motor drives is improved. This paper presents dynamic simulation of induction motor drive using neuro controller. The integrated environment allows users to compare simulation results between conventional, Fuzzy and Neural Network controller (NNW). The performance of fuzzy logic and artificial neural network based controller's are compared with that of the conventional proportional integral controller. The dynamic modeling and simulation of Induction motor is done using MATLAB/SIMULINK and the dynamic performance of induction motor drive has been analyzed for artificial intelligent controller.

Keywords Neuro network (NNW) • PI controller • Fuzzy logic controller (FLC) • Sugeno fuzzy controller • Hebbian learning algorithm

P. M. Menghal (✉)

Faculty of Degree Engineering, Military College of Electronics and Mechanical Engineering, Secunderabad 500015, India
e-mail: prashant_menghal@yahoo.co.in

Department of EEE, Jawaharlal Nehru Technological University, Anantapur 515002, Andhra Pradesh, India

A. Jaya Laxmi

Department of EEE, Jawaharlal Nehru Technological University, College of Engineering, Kukatpally, Hyderabad 500085, Andhra Pradesh, India
e-mail: ajl1994@yahoo.co.in

1 Introduction

Three phase Induction Motor have wide applications as electrical machines. About half of the electrical energy generated in a developed country is ultimately consumed by electric motors, of which over 90 % are induction motors. For a relatively long period, induction motors have mainly been deployed in constant-speed motor drives for general purpose applications. The rapid development of power electronic devices and converter technologies in the past few decades, however, has made possible efficient speed control by varying the supply frequency, giving rise to various forms of adjustable-speed induction motor drives. In about the same period, there were also advances in control methods and Artificial Intelligence (AI) techniques. Artificial Intelligent techniques mean use of expert system, fuzzy logic, neural networks and genetic algorithm. Researchers soon realized that the performance of induction motor drives can be enhanced by adopting artificial-intelligence-based methods. The Artificial Intelligence (AI) techniques, such as Expert System (ES), Fuzzy Logic (FL), Artificial Neural Network (ANN), and Genetic Algorithm (GA) have recently been applied widely in control of induction motor drives. Among all the branches of AI, the NNW seems to have greater impact on power electronics and motor drives area that is evident by the publications in the literature. Since the 1990s, AI-based induction motor drives have received greater attention. Apart from the control techniques that exist, intelligent control methods, such as fuzzy logic control, neural network control, genetic algorithm, and expert system, proved to be superior. Artificial Intelligent Controller (AIC) could be the best controller for Induction Motor control [1–6]. Fuzzy controller conventionally is totally dependent to memberships and rules, which are based broadly on the intuition of the designer. This paper tends to show Neuro Controller has edge over fuzzy controller. Sugeno fuzzy controller is used to train the fuzzy system with two inputs and one output [7–11]. The performance of fuzzy logic and artificial neural network based controllers is compared with that of the conventional proportional integral controller.

2 Dynamic Modeling and Simulation of Induction Motor Drive

The induction motors dynamic behavior can be expressed by voltage and torque which are time varying. The differential equations that belong to dynamic analysis of induction motor are so sophisticated. Then with the change of variables the complexity of these equations decrease through movement from poly phase winding to two phase winding (q-d). In other words, the stator and rotor variables like voltage, current and flux linkages of an induction machine are transferred to another reference model which remains stationary [1–6].

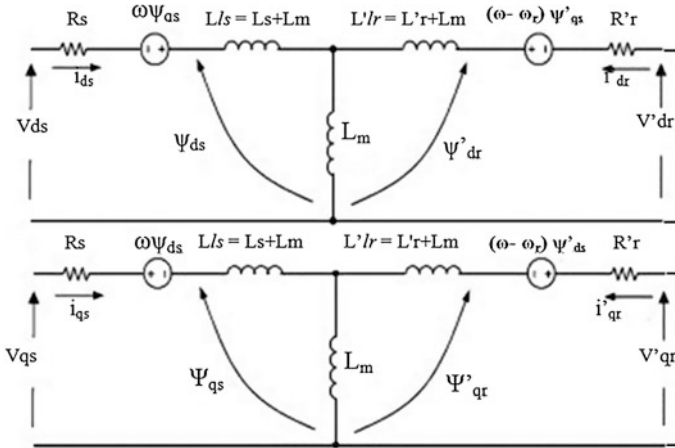


Fig. 1 d q model of induction motor

In Fig. 1 stator inductance is the sum of the stator leakage inductance and magnetizing inductance ($L_{ls} = L_s + L_m$), and the rotor inductance is the sum of the rotor leakage inductance and magnetizing inductance ($L_{lr} = L_r + L_m$). From the equivalent circuit of the induction motor in d-q frame, the model equations are derived. The flux linkages can be achieved as:

$$\frac{1}{\omega_b} \frac{d\psi_{qs}}{dt} = v_{qs} - \frac{\omega_e}{\omega_b} \psi_{ds} - R_s i_{qs} \tag{1}$$

$$\frac{1}{\omega_b} \frac{d\psi_{ds}}{dt} = v_{ds} - \frac{\omega_e}{\omega_b} \psi_{qs} - R_s i_{ds} \tag{2}$$

$$\frac{1}{\omega_b} \frac{d\psi_{qr}}{dt} = v_{qr} - \frac{(\omega_e - \omega_r)}{\omega_b} \psi_{dr} - R_s i_{qr} \tag{3}$$

$$\frac{1}{\omega_b} \frac{d\psi_{dr}}{dt} = v_{dr} + \frac{(\omega_e - \omega_r)}{\omega_b} \psi_{qr} - R_s i_{dr} \tag{4}$$

By substituting the values of flux linkages in the above equations, the following current equations are obtained as:

$$i_{qs} = \frac{(\psi_{qs} - \psi_{mq})}{X_{ls}} \tag{5}$$

$$i_{ds} = \frac{(\psi_{ds} - \psi_{md})}{X_{ls}} \tag{6}$$

$$i_{qr} = \frac{(\psi_{qr} - \psi_{mq})}{X_{ls}} \tag{7}$$

$$i_{dr} = \frac{(\psi_{dr} - \psi_{md})}{X_{ls}} \quad (8)$$

where ψ_{mq} and ψ_{md} are the flux linkages over L_m in the q and d axes. The flux equations are written as follows:

$$\psi_{mq} = X_{ml} \left(\frac{\psi_{qs}}{X_{ls}} + \frac{\psi_{qr}}{X_{lr}} \right) \quad (9)$$

$$\psi_{md} = X_{ml} \left(\frac{\psi_{ds}}{X_{ls}} + \frac{\psi_{dr}}{X_{lr}} \right) \quad (10)$$

$$X_{ml} = \frac{1}{\frac{1}{X_m} + \frac{1}{X_{ls}} + \frac{1}{X_r}} \quad (11)$$

In the above equations, the speed ω_r is related to the torque by the following mechanical dynamic equation as:

$$T_e = T_{load} + J \frac{d\omega_m}{dt} = T_{load} + \frac{J2}{p} \frac{d\omega_r}{dt} \quad (12)$$

then ω_r is achievable from above equation, where:

p: number of poles.

J: moment of inertia (kg/m^2).

In the previous section, dynamic model of an induction motor is expressed. The model constructed according to the equations has been simulated by using MATLAB/SIMULINK as shown in Fig. 2 in conventional mode of operation of induction motor. A 3 phase source is applied to conventional model of an induction motor and the equations are given by:

$$V_a = \sqrt{2}V_{rms} \sin(\omega t) \quad (13)$$

$$V_b = \sqrt{2}V_{rms} \sin\left(\omega t - \frac{2\pi}{3}\right) \quad (14)$$

$$V_c = \sqrt{2}V_{rms} \sin\left(\omega t + \frac{2\pi}{3}\right) \quad (15)$$

By using Parks Transformation, voltages are transformed to two phase in the d-q axes, and are applied to induction motor. In order to obtain the stator and rotor currents of induction motor in two phase, Inverse park transformation is applied in the last stage [4].

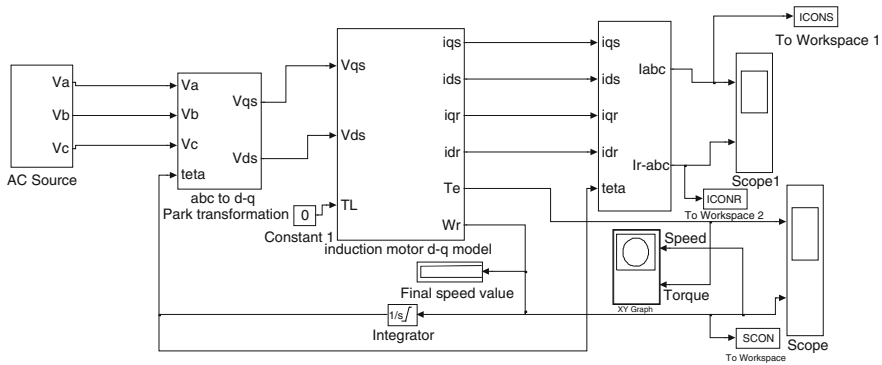


Fig. 2 Simulated induction motor model with conventional controller

3 Fuzzy Logic Controller

The speed of induction motor is adjusted by the fuzzy controller. The membership function of Δ_e , e and three scalar values of each triangle are applied into this controller. In Table 1, the fuzzy rules decision implemented into the controller are given. The conventional simulated induction motor model as shown in Fig. 2 is modified by adding Fuzzy controller and is shown in Fig. 3. Speed output terminal of induction motor is applied as an input to fuzzy controller, and in the initial start of induction motor the error is maximum, so according to fuzzy rules fuzzy controller produces a crisp value. Then this value will change the frequency of sine wave in the speed controller. The sine wave is then compared with triangular waveform to generate the firing signals of IGBTs in the PWM inverters. The frequency of these firing signals also gradually changes, thus increasing the frequency of applied voltage to Induction Motor [9].

As discussed earlier, the crisp value obtained from Fuzzy Logic Controller is used to change the frequency of gating signals of PWM inverter. Thus the output AC signals obtained will be variable frequency sine waves. The sine wave is generated with amplitude, phase and frequency which are supplied through a GUI. Then the clock signal which is sampling time of simulation is divided by crisp value which is obtained from Fuzzy Logic Controller (FLC). So by placing three sine waves with different phases, one can compare them with triangular waveform and generate necessary gating signals of PWM inverter. So at the first sampling point the speed is zero and error is maximum. Then whatever the speed rises, the error will decrease, and the crisp value obtained from FLC will increase. So, the frequency of sine wave will decrease which will cause IGBTs switched ON and OFF faster. It will increase the AC supply frequency, and the motor will speed up. Figure 3 shows the Fuzzy logic induction motor model.

Table 1 Modified fuzzy rule decision

		Δ_e				
		NB	NS	ZZ	PS	PB
e	PB	ZZ	NS	NS	NB	NB
	PS	PS	ZZ	NS	NS	NB
	ZZ	PS	PS	ZZ	NS	NS
	NS	PB	PS	PS	ZZ	NS
	NB	PB	PB	PS	PS	ZZ

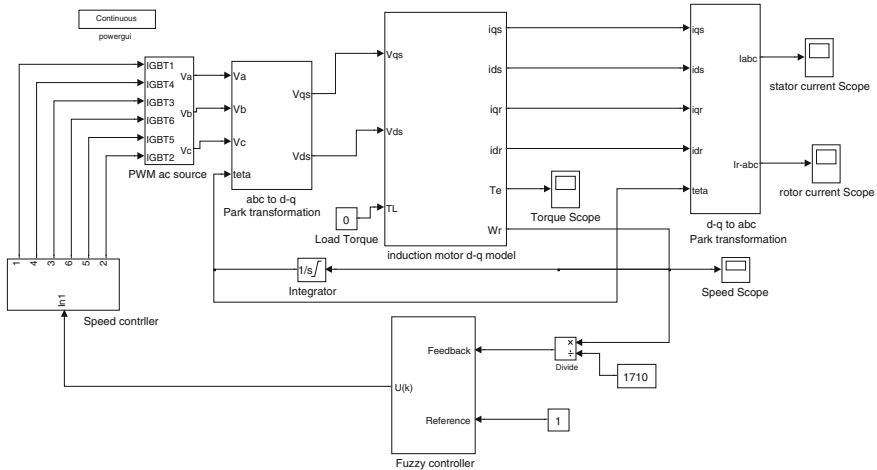


Fig. 3 Fuzzy control induction motor model

4 Adaptive Neuro Fuzzy Controller

In the designing of a controller, the main criterion is the controllability of torque in an induction motor with good transient and steady state responses. With certain drawbacks, PI controller is able to achieve these characteristics. The main drawbacks are (1) The gains cannot be increased beyond certain limit. (2) Non linearity is introduced, making the system more complex for analysis. The shortcomings of PI controller are overcome by artificial intelligent techniques. One such technique is the use of Fuzzy Logic in the design of controller either independently or in hybrid with PI controller. The draw-backs of Fuzzy Logic Control and Artificial Neural Network are replaced by Adaptive Neuro-Fuzzy Inference System (ANFIS). Adaptive neuro fuzzy combines the learning power of neural network with knowledge representation of fuzzy logic. A neuro fuzzy system is based on a fuzzy system which is trained by a learning algorithm derived from neural network theory. Depending on the applications, one can use either ANN or FIS, or combination of both. In this paper, the inputs will be $e(k)$ and $\Delta_e(k)$ [9, 12, 17]. A first-order Sugeno fuzzy model has rules which are as follows:

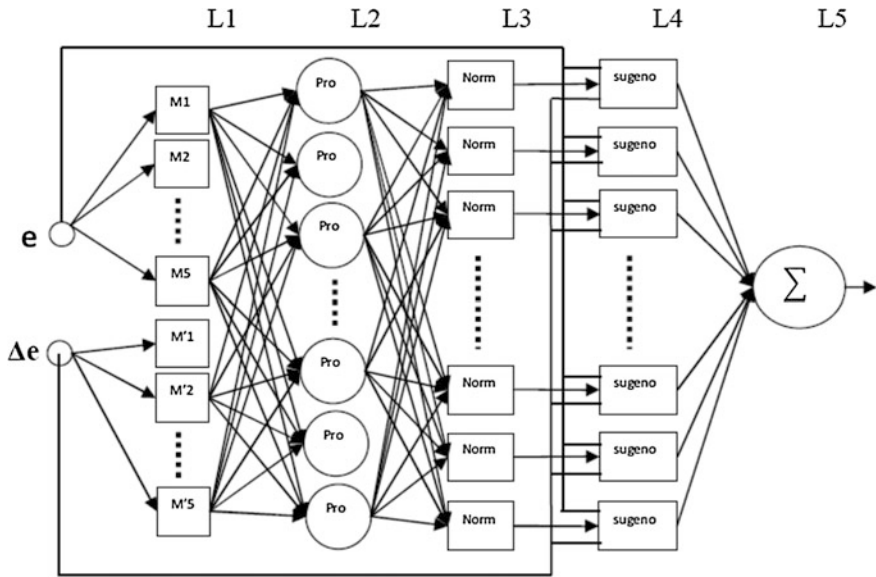


Fig. 4 ANFIS layout

- Rule1: If x is A_1 and y is B_1 , then $f_1 = p_1x + q_1y + r_1$
- Rule2: If x is A_2 and y is B_2 , then $f_2 = p_2x + q_2y + r_2$.

In the Sugeno model if-then rules are used, and output of each rule is linear combination of inputs plus a constant value. The learning algorithm applied to this model is Hebbian. This method is feed forward and unsupervised and the weights will be adjusted by the following formula:

$$w_i(\text{new}) = w_i(\text{old}) + x_i y \tag{16}$$

The ANFIS layout is shown in Fig. 4. It states that if the cross product of output and input is positive, then it results in increase of weight, otherwise decrease of weight.

In layer 2 of ANFIS layout, the triangular membership function is same as that of the fuzzy controller model. The output of layer 2 is given by:

$$O_2 = \mu_1, \mu_2, \mu_3 \tag{17}$$

Layer 3 indicates the pro (product) layer and its output is product of inputs, which is given by:

$$O_3 = \mu_i(e) \cdot \mu_j(\Delta e) \tag{18}$$

Layer 4 represent Norm and it calculates the ratio of i th firing strength to sum of all firing strengths. The obtained output is normalized firing strength, which is given by:

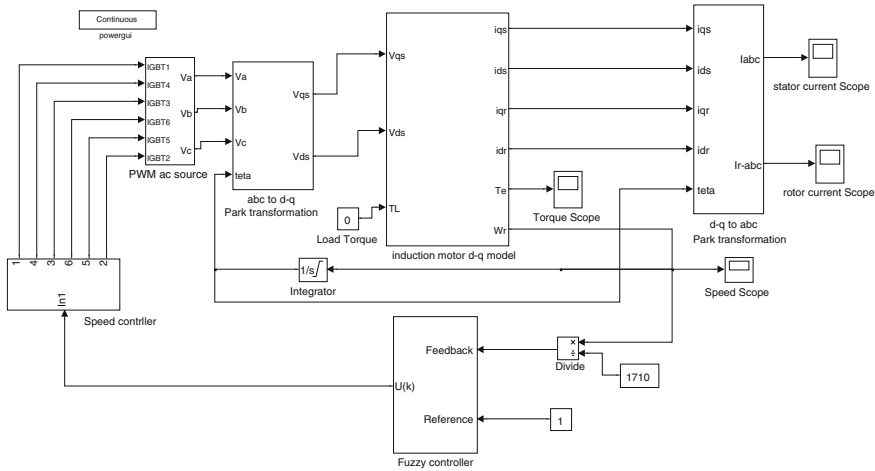


Fig. 5 Adaptive neuro-fuzzy controller simulation model

$$O_4 = \frac{w_i}{\sum w} \tag{19}$$

Layer 5 is an adaptive node with functionality as follows:

$$O_5 = w_i f_i = w_i(p_i(e) + q_i(\Delta e) + r_i) \tag{20}$$

where p_i , q_i , r_i are consequent parameters, which are initially are set to 0.48, 0.25 and 1 respectively. Then they are adaptively adjusted with Hebbian learning algorithm. Layer 6 calculates the output which is given by:

$$O_6 = \frac{\sum w_i f_i}{\sum w_i} \tag{21}$$

Figure 5 shows the overall structure of Adaptive Neuro-Fuzzy controller.

5 Neuro Controller

The most important feature of Artificial Neural Networks (ANN) is its ability to learn and improve its operation using neural network training [13, 14]. The objective of Neural Network Controller (NNC) is to develop a back propagation algorithm such that the output of the neural network speed observer can track the target one. The network structure of the NNC, indicates that the neural network has three layered network structure. The first is formed with five neuron inputs $\Delta(\omega_{ANN}(K + 1)), \Delta(\omega_{ANN}(K)), \omega_{ANN}, \omega_S(K - 1), \Delta(\omega_S(K - 2))$. The second layer consists of five neurons. The last one contains one neuron to give the command variation $\Delta(\omega_S(K))$. The aim of the proposed NNC is to compute the

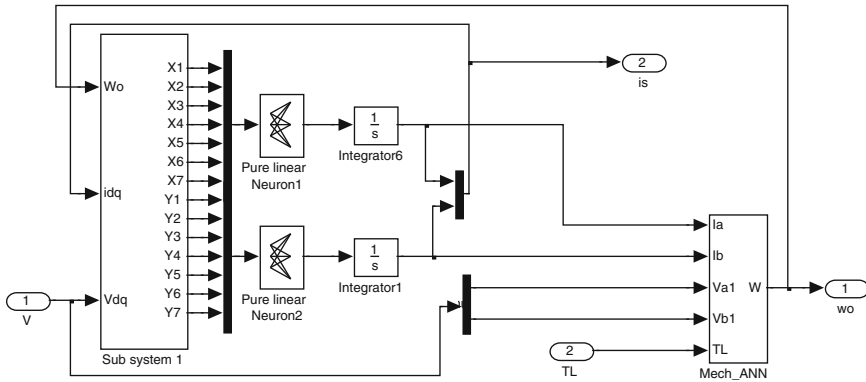


Fig. 6 Neural network controller

command variation based on the future output variation $\Delta(\omega_{ANN}(K + 1))$. Hence, with this structure, a predictive control with integrator has been realised. At time k , the neural network computes the command variation based on the output at time $(k + 1)$, while the later isn't defined at this time. In this case, it is assumed that $\omega_{ANN}(K + 1) \equiv \omega_{ANN}(K)$. The control law is deduced using the recurrent equation given by,

$$\omega_S(K) = \omega_S(K - 1) + G\Delta(\omega_S(K)).$$

It can be seen that the d axis and q axis voltage equations are coupled by the terms d_E and q_E . These terms are considered as disturbances and are cancelled by using the proposed decoupling method. If the decoupling method is implemented, the flux component equations become

$$\Phi_{dr} = G(s)v_{ds}$$

$$\Phi_{qr} = G(s)v_{qs}$$

Large values of η may accelerate the ANN learning and consequently fast convergence but may cause oscillations in the network output, whereas low values will cause slow convergence. Therefore, the value of η has to be chosen carefully to avoid instability. The proposed neural network controller is shown in Fig. 6.

6 Simulation Results and Discussion

Modeling and simulation of Induction motor in conventional, fuzzy and adaptive neuro fuzzy are done on MATLAB/SIMULINK. A complete simulation model and dynamic performance for inverter fed induction motor drive incorporating the proposed FLC, adaptive neuro fuzzy controller and Neuro controller has been

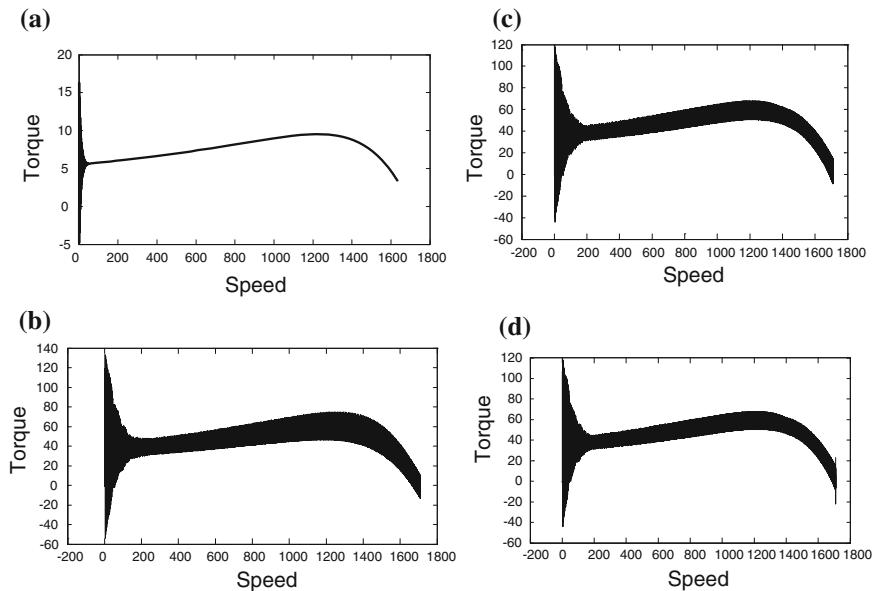


Fig. 7 Torque-speed characteristics. **a** Torque-speed characteristics with conventional controller. **b** Torque-speed characteristics with fuzzy controller. **c** Torque-speed adaptive neuro fuzzy controller. **d** Torque speed characteristics with neuro controller

developed. The proposed neuro controller is proved to be more superior as compared to adaptive by comparing the response of conventional PI, with FLC, Adaptive neuro fuzzy speed controller based IM drive. The results of simulation for induction motor with its characteristics are listed in Appendix 'A'. Figures 7, 8 and 9 shows the torque-speed characteristics, torque and speed responses of conventional, FLC, adaptive neuro fuzzy controller and neuro controller respectively. It appears the rise time drastically decreases when neuro controller is added to simulation model and both the results are taken in same period of time. In neural network based simulation, it is apparent from the simulation results shown in Fig. 7c and d, torque-speed characteristic converges to zero in less duration of time when compared with conventional Controller and FLC, which is shown in Fig. 7a and b. Neuro controller has no overshoot and settles faster in comparison with FLC and adaptive neuro fuzzy controller. It is also noted that there is no steady-state error in the speed response during the operation when neuro controller is activated. In conventional controller, oscillations occur, whereas in neuro controller, adaptive neuro fuzzy and FLC, no oscillations occur in the torque response before it finally settles down as shown in Fig. 8 good torque response is obtained with Neuro controller as compared to conventional, FLC and adaptive neuro fuzzy controller at all time instants and speed response is better than conventional controllers, FLC and adaptive neuro fuzzy controller. There is a negligible ripple in speed response with neuro fuzzy controller in comparison with conventional

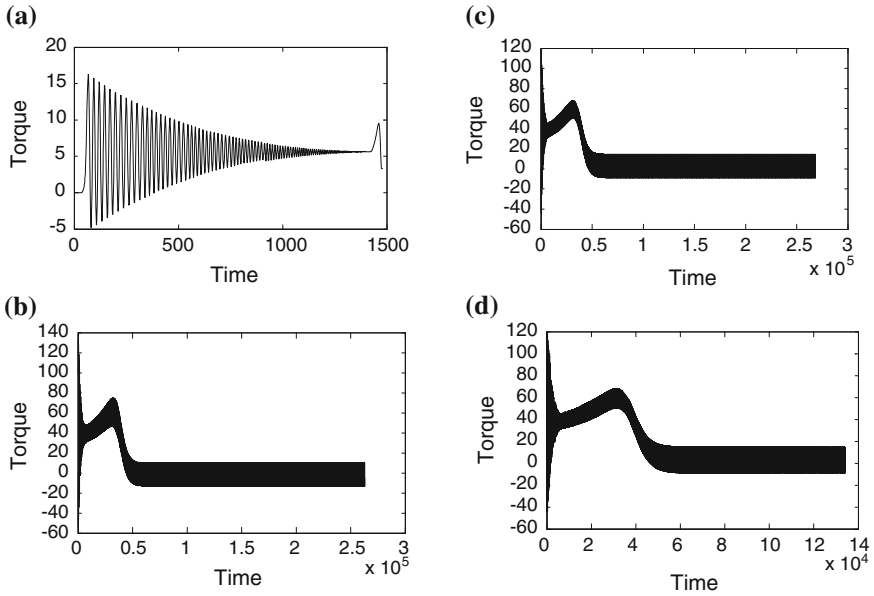


Fig. 8 Torque characteristics. **a** Conventional controller. **b** Fuzzy controller. **c** Adaptive neuro fuzzy controller. **d** Neuro controller

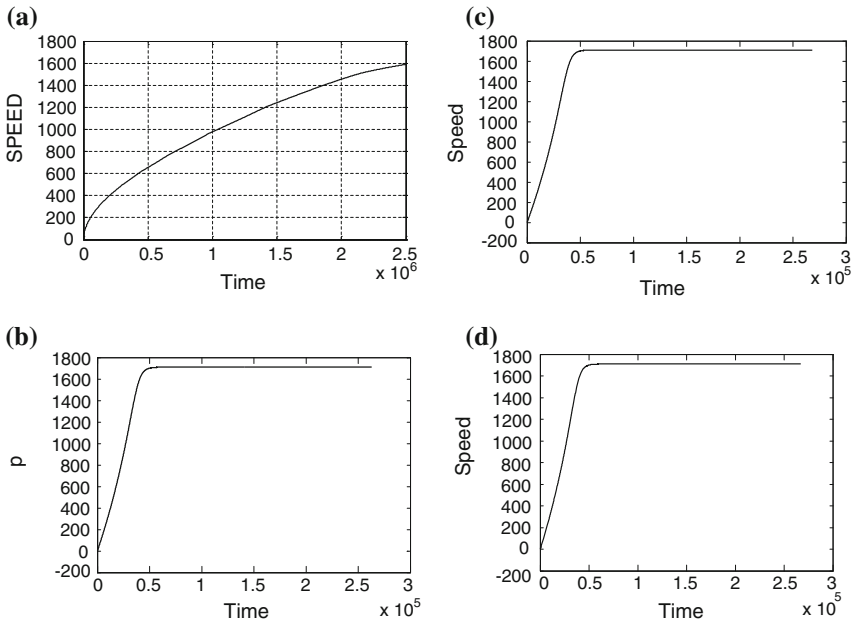


Fig. 9 Speed characteristics. **a** Conventional controller. **b** Fuzzy controller. **c** Adaptive neuro fuzzy controller. **d** Neuro controller

controller, FLC and adaptive neuro controller under dynamic conditions which are shown in Fig. 8. With the neuro controller, speed reaches its steady state value faster as compared to Conventional, FLC and adaptive neuro fuzzy controller.

7 Conclusion

In this paper, comparison of simulation results of the induction motor are presented with different types of controller such as conventional, fuzzy control, Adaptive neuro fuzzy and neuro controller. From the speed waveforms, it is observed that with adaptive fuzzy and neuro controller the rise time decreases drastically, in the manner which the frequency of sine waves are changing according to the percentage of error from favourite speed. According to the direct relation of induction motor speed and frequency of supplied voltage, the speed will also increase. Fuzzy controller has better performance than the conventional controller. By comparing neuro controller, Adaptive neuro fuzzy model with FLC model, it is apparent that by adding learning algorithm to the control system will decrease the rising time more than expectation and it proves neuro controller has better dynamic performance as compared to FLC and Adaptive neuro fuzzy controller.

Appendix A

The following parameters of the induction motor are chosen for the simulation studies:

$V = 220$	$f = 60$	$HP = 3$	$R_s = 0.435$	$R_r = 0.816$	$X_{ls} = 0.754$
$X_{lr} = 0.754$		$X_m = 26.13$	$p = 4$	$J = 0.089$	$rpm = 1,710$

References

1. Chan, T.F., Shi, K.: Applied Intelligent Control of Induction Motor Drives. IEEE Wiley Press (2011)
2. Krause, P.C.: Analysis of Electrical Machinery and Drives System. IEEE Wiley Press (2000)
3. Mohan, N.: Advanced Electric Drives: Analysis, Control Modeling using Simulink. MNPHERE Publication (2001)
4. Menghal, P.M., Laxmi, A.J.: Adaptive neuro fuzzy based dynamic simulation of induction motor drives. IEEE International Conference on Fuzzy Systems, pp. 1–8 (2013)
5. Menghal, P.M., Laxmi, A.J.: Neural network based dynamic simulation of induction motor drives. IEEE International Conference on Power, Energy and Control (2013). doi:[10.1109/ICPEC.2013.6527722](https://doi.org/10.1109/ICPEC.2013.6527722)

6. Menghal, P.M., Laxmi, A.J.: Adaptive neuro fuzzy interference (ANFIS) based simulation of Induction motor drive. *Int. Rev. Model. Simul. (IRMOS)* **5**(5), 2007–2016 (2012)
7. Shi, K.L., Chan, T.F., Wong, Y.K., Ho, S.L.: Modeling and simulation of the three phase induction motor using SIMULINK. *Int. J. Elect. Eng. Educ.* **36**, 163–172 (1999)
8. Dandil, B., Gokbulut, M., Ata, F.: A PI type fuzzy—neural controller for induction motor drives. *J. Appl. Sci.* (2005). doi:[10.3923/jas.2005.1286.1291](https://doi.org/10.3923/jas.2005.1286.1291)
9. Kumar, R., Gupta, R.A., Surjuse, R.S.: Adaptive neuro-fuzzy speed controller for vector controlled induction motor drive. *APEJ* (2009). doi:[14.79e41505757a2a8cab](https://doi.org/10.79e41505757a2a8cab)
10. Denai, M.A., Attia, S.A.: Fuzzy and neural control of an induction motor. *Int. J. Appl. Math. Comput. Sci.* (2002). doi:[10.1.1.135.303](https://doi.org/10.1.1.135.303)
11. Subudhi, B., Anish Kumar, A.K., Jena, D.: dSPACE implementation of fuzzy logic based vector control of induction motor. *IEEE TENCON* (2008). doi:[10.1109/TENCON.2008.4766502](https://doi.org/10.1109/TENCON.2008.4766502)
12. Bose, B.K.: Neural network applications in power electronics and motor drives—an introduction and perspective. *IEEE Trans. Ind. Electron.* (2007). doi:[10.1109/TIE.2006.888683](https://doi.org/10.1109/TIE.2006.888683)
13. Uddin, M.N., Hafeez, M.: FLC-based DTC scheme to improve the dynamic performance of an IM drive. *IEEE Trans. Ind. Appl.* (2012). doi:[10.1109/TIA.2011.2181287](https://doi.org/10.1109/TIA.2011.2181287)
14. Uddin, M.N., Wen, H.: Development of a self-tuned neuro-fuzzy controller for induction motor drives. *IEEE Trans. Ind. Appl.* (2007). doi:[10.1109/TIA.2007.900472](https://doi.org/10.1109/TIA.2007.900472)
15. Uddin, M.N., Radwan, T.S., Rahman, A.: Performance of fuzzy logic based indirect vector control for induction motor drive. *IEEE Trans. Ind. Appl.* (2002). doi:[10.1109/TIA.2002.802990](https://doi.org/10.1109/TIA.2002.802990)
16. Uddin, M.N., Huang, Z.R., Chy, M.M.I.: A simplified self-tuned neuro-fuzzy controller based speed control of an induction motor drive. *IEEE Power Engineering Society General Meeting* (2007). doi:[10.1109/PES.2007.385720](https://doi.org/10.1109/PES.2007.385720)

A Bio-inspired Trusted Clustering for Mobile Pervasive Environment

Madhu Sharma Gaur and Bhaskar Pant

Abstract Pervasive systems are usually highly dynamic, heterogeneous, and resource-restricted where small and powerful dissimilar devices have to establish independent network unknown by the user. There is no fixed infrastructure and centralized access control. The set of connections relies on the convergence of wireless technologies, advanced electronics and the Internet to communicate seamlessly with other devices as tiny sensors. Trusted and Security-critical communication is the key concern in such decentralized and unpredictable environment. Bio-Inspired systems are increasing significant adaptation, reliability and strength in the dynamic and heterogeneous networks where information is ubiquitous. Some specific characteristics of swarms, like their lightweight, transient nature and indirect communication, make this adaptation more demanding. In this paper we explore bio-Inspired systems to look at the trust computation factors and opportunities in autonomic computing environments like mobile pervasive environment and evaluate their trustworthiness. We use standard clustering technique and propose a trust metric in which we observe the node behavior through various trust parameters. In winding up, we put our efforts to represent the cluster formation with honey bee mating to set up general vulnerabilities requirements for compromised node behavior to the system under exploration.

Keywords Clustering · Honey bee mating · Pervasive environment · Trust metric

M. S. Gaur (✉)
GEU, Dehradun, Uttarakhand, India
e-mail: madhu14nov@gmail.com

GL Bajaj Institute of Technology and Management, Greater Noida, India

B. Pant
Department of IT, GEU, Dehradun, Uttarakhand, India
e-mail: pantbhaskar2@gmail.com

1 Introduction

The rapid growth of mobile computing has given rise to the information systems in which user can access the global network regardless of location or time. The words pervasive and ubiquitous mean “existing everywhere.” Pervasive computing devices are completely connected and constantly available. The vision of ubiquitous computing which Mark Weiser described in his 1991 paper [1] is based on the idea that future computers merge with their environment more and more until they become completely invisible for the user. Pervasive systems are usually highly dynamic, heterogeneous, and resource-restricted where small and powerful dissimilar devices have to establish independent network unknown by the user. There is no fixed infrastructure and centralized access control and set of connections relies on the convergence of wireless technologies, advanced electronics and the Internet to communicate seamlessly with other devices as tiny sensors and needs to be self adaptive and self-organizing.

Distributed wireless micro-sensor networks are an important component of the Pervasive computing that relies on the convergence of wireless technologies, advanced electronics and the Internet. A sensor node are location unaware and may not be equipped with GPS, can communicate directly only with other sensors that are within a small distance. However, in reality, sensor nodes are resource-restricted. Due to lack of fixed infrastructure, all the nodes have autonomous to make decisions based on the available information on the relying base station or mobile base station. All nodes are integrated into a wireless mobile pervasive Ad-Hoc network with multi-hop routing ability. Traditional security schemes cannot always be applied to such environments. Therefore, concepts like trust and reputation also applied to gain a certain level of security and confidence among inter-operating nodes. Up to the present, research on the trust management mechanisms of WSNs or MANETs have mainly focused on node’s trust evaluation to enhance the security and robustness where trust evaluation is the key concern to recognize malicious, selfish and compromised nodes which have been authenticated. In this paper we use standard clustering technique and propose a trust metric in which we observe the node behavior through various trust parameters. Clustering is a classic approach for achieving an energy efficient performance in sensor networks. Clustering provides locality of communication through organizing the number of nodes as clusters which saves energy and reduces network contention. In winding up, we put our efforts to represent the cluster formation with honey bee mating scheme, an energy efficient trusted cluster formation and head selection in pervasive mobile environment. Rest of the work is organized as II. Literature review, III proposed Trust Metric, IV A Bio-Inspired Cluster formation and finally conclusion and future scope.

2 Literature Review

We explore bio-Inspired systems to look at the trust computation factors and opportunities in autonomic computing environments like mobile pervasive ad-hoc networks and evaluate their trustworthiness Cho et al. [2], present a complete survey on trust management in MANET and specify that Trust is dynamic, subjective, not necessarily transitive, asymmetric and context-dependent. It can be defined as Direct and Indirect trust. LEACH (Low Energy Adaptive Clustering Hierarchy [3] is a cluster based protocol, which includes distributed cluster formation. For each cluster, a sensor node is selected as a cluster head. The cluster head applies aggregation functions to squeeze the data before transmission to the destination. PTM [4–6] a research sub-item of UBISEC (secure pervasive computing) supported by Europe IST FP6, which presents different models with revised D-S evidence theory and defines the inter-domain dynamic trust management in subjective area. The limitation of the PTM is that it acquires indirect trust value on average without taking the fuzziness, subjectivity and uncertainty into account. Lopez et al. [7] list the best practices that are essential for developing a good trust management system for WSN and make an analysis of the state of the art related to these practices. These references formulate an amazing summary, propose many profound viewpoints and show an additional insight on the trust evaluation field. In addition, other protocols [2, 8, 9] address trust management methods in self-organization networks from different views. A honey bee mating applications on clustering [10, 11] also inspire our proposed approach to present a bio inspired trusted clustering for pervasive environment.

3 Proposed Trust Metric

Our proposed trust metric based on social trust, QoS trust and reliability in terms of packet sent and received parameters to evaluate best possible aspects trustworthiness.

3.1 Trust Metric Parameters

The proposed trust metric key trust parameters are intimacy (for measuring nearness based on interaction experiences) and integrity (for measuring irregularity) to measure social trust derived from social networks. We choose energy (for measuring competence) and selfishness (for measuring uncooperativeness) to measure QoS trust derived from communication networks Table 1.

Here *intimacy evaluates* two node's neighbor nose's interaction experience. It follows the maturity model proposed in [8] where sensor nodes have more positive

Table 1 Trust Metric Parameters

$T_{xy}^{Intimacy}$	Intimacy for measuring nearness or closeness based on past experience
$T_{xy}^{Integrity}$	Integrity for measuring irregularity or the honesty
T_{xy}^{Energy}	Energy for measuring competence or capability
$T_{xy}^{Selfishness}$	Selfishness for measuring uncooperativeness
$T_{X/Y}^{Mobility}$	Mobility to estimate the power consumption and residual energy of any node x
$T_{xy}^{Reliability}$	Reliability As total number of Packets sent by node x and received by node y

experiences. Assuming that a compromised node is malicious and untruthful, integrity component is taken that can efficiently identify whether a node is malicious or not. As a QoS trust metric energy is one of the most important components in the subjective resource-restricted networks. The unselfishness parameter specifies whether a node can cooperatively execute the intended procedure. In pervasive diverse devices or nodes are communicating seamlessly thus proposed approach can be applied in a heterogeneous network with immensely different energy levels and degrees of malicious or selfish behaviors. We apply this trust management approach to a clustered pervasive ad-hoc environment in which a sensor node may adjust its behavior dynamically according to its own operational state and environmental conditions. Here each node is more likely to become selfish in case of low energy level or it has many unselfish neighbor nodes around when it has more compromised neighbors around it.

- $T_{xy}^{Intimacy}(\mathbf{t})$: It ranks the interaction experiences following the maturity model [8]. It is computed by the number of interactions between nodes x and y over the maximum number of interactions between node x and any neighbor node over the time period [0, t].
- $T_{xy}^{Integrity}(\mathbf{t})$: This refers to the confidence of node x that node y is truthful based on node x's direct observations toward node y. Node x calculate approximately T honesty, direct $ij(\mathbf{t})$ by observing a count of suspicious untruthful experiences of node y which node x has observed during [0, t] using a set of anomaly detection rules such as a high inconsistency in the sensor reading or recommendation has been experienced, as well as interval, retransmission, repetition, and delay rules.
- $T_{xy}^{Energy}(\mathbf{t})$: This refers to the belief of node x that node y still has adequate energy (representing competence) to perform its intended function. It may be measured by the percentage of node j's remaining energy. To calculate $T_{xy}^{Energy}(\mathbf{t})$, node x estimates node y's remaining energy by overhearing node y's packet transmission activities over the time period [0, t], utilizing an energy consumption model.
- $T_{xy}^{Selfishness}(\mathbf{t})$: This parameter represents the degree of selflessness of node y as estimated by node x based on direct observations over [0, t].

Furthermore the selfish behavior of node y can be detected using eavesdrop and snooping techniques such as not honestly performing sensing and reporting functions, data forwarding summing that a compromised node must be uncooperative and thus selfish. If node x is not a 1-hop neighbor of node y , node x will use its former experience $T_{c_{ij}}(t - \Delta t)$ and recommendations for Selfishness it is also possible that a node doesn't route a packet from the other nodes or simply drops some packets to save their power or other energy. Thus such selfish nodes cannot have a high trust value because of the data delivery rate. By not providing packet forwarding for low trusted nodes, a network can encourage cooperation and reduce selfishness.

$T_{xy}^{\text{Reliability}}$:

The reliability of nodes can be evaluated in different ways, but, in general, it can be considered as the capability of nodes to respect a service agreement. This is a particular procedure that lies behind the identity certification or the encryption process. In the remaining part of this section, the word trust is used to identify the reliability of nodes. However, the protocol presented here can be easily extended to incorporate identity checking and trusting in the classic sense.

3.2 Algorithm Trust Evaluation (Calc-Trust)

- Step 1: Collect data about a node (X_i to node n , where n is the total no of nodes in a cluster)
- Step 2: Find the Trust threshold values associated to each behavior as described above
- Step 3: Calculate trust value for each parameter [0.0–0.2]
- Step 4: Aggregate all the trust value and find the mean corresponding threshold.
- Step 5: Calculate the corresponding trust value using the formula.

3.3 Trust Calculation

The trust calculation is conducted, particularly between two neighbor nodes in a cluster. When a node X evaluates trust on another node Y at time t . We consider five trust components as described above like intimacy, integrity, energy, selfishness and reliability. The trust value that node X evaluates towards node Y at time t , $T_{xy}(t)$, is represented as a real number in the range of $[0, 1]$ where 0 indicates distrust and 1 complete trust. $T_{xy}(t)$ is computed by

Table 2 Trust parameters and Cumulative Trust levels

Trust Parameters	Trust weight	Cumulative Trust Value	Level of trust
	0.0	0.0	Distrust
Intimacy	0.2	0.2	Very low trust
Integrity	0.2	0.4	Low trust
Energy	0.2	0.6	Partially Trusted
Selfishness	0.2	0.8	Highly trusted
Reliability	0.2	1.0	Fully Trusted

Table 3 Trust Value range of Trust Parameters

Trust Parameters	Trust Value Range of Parameter		
	0.0	0.1	0.2
Intimacy	No familiarity	Partial Intimacy	Fully intimate
Integrity	No Integrity	Partial Integrity	Full Integrity
Energy	No Energy Efficient	Partial Energy Efficient	Fully Energy Efficient
Selfishness	Unselfish	Partial Selfish	Fully selfish
Reliability	No reliable	Partial Reliable	Fully Reliable
	No Trust	Partial Trust	Full Trust

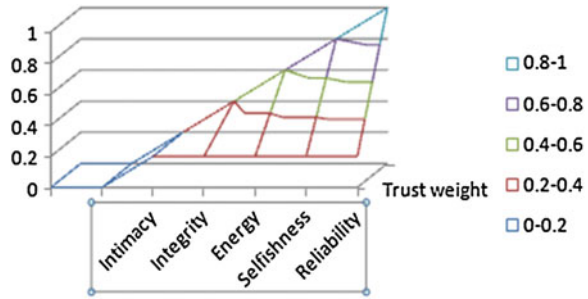
$$T_{xy}(t) = C1 * T_{xy}^{Intimacy} + C2 * T_{xy}^{Integrity} + C3 * T_{xy}^{Energy} + C4T_{xy}^{Selfishness} + C5T_{xy}^{Reliability} \tag{1}$$

Where C1, C2, C3, C4, and C5 are costs associated with these five trust components with C1 + C2 + C3 + C4 + C5 = 1. Deciding the best values of C1, C2, C3, C4, and C5 to maximize system performance is a trust formation. We assume that each trust parameter is equally contributing in the process of Trust formation. Equation 1 can be rewritten

$$T_{xy}(t) = 0.02C1 * T_{xy}^{Intimacy} + 0.2C2 * T_{xy}^{Integrity} + 0.2C3 * T_{xy}^{Energy} + 0.2C4T_{xy}^{Selfishness} + 0.2C5T_{xy}^{Reliability}$$

After collecting the information about nodes X and Y an Algorithm Compute-TRUST will be run to calculate the direct trust of node X about Y. Whenever the cluster head C-Hds inquires X's opinion about Y node, it will send the trust value. Assuming that each trust parameter have equal contribution in the final trust value T_{xy}(t). On the basis of cumulative trust value after each parameter Calculated Suppose For example Tables 2, 3 and Fig. 1.

Fig. 1 Trust metric parameters



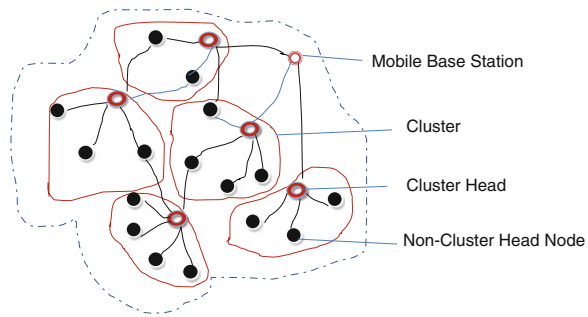
4 A Bio-inspired Trusted Cluster Formation

Clustering represents the different virtual groups of network sensor nodes which are physically neighboring and helps to organize the pervasive ad hoc networks hierarchically. Number of heuristic clustering algorithms has been presented in the literature and discuss about the latest developments in clustering like mobility-aware clustering, energy-efficient clustering, load-balancing clustering and combined-metrics-based clustering. In the mobile pervasive environments, nodes may differ from each other in terms of available resources and degree of mobility. Major resources are the communication, computation power and energy efficiency while the degree of mobility is the relative value to indicate dynamism of a node and average speed of the node. The magnitude of resources and mobility may differ continuously. As a whole, we assume that every node has different willingness value to be a volunteer set by its owner. A node that has abundant resources, a lower degree of mobility and a higher willingness value has a higher chance to be a volunteer. Any node can be a service provider (SP) as well as a service requestor (SR). The volunteers maintain a list of neighbor volunteers and a service directory for its range. We use a Service Discovery based on Volunteers for heterogeneous and uncertain pervasive computing environments. It provides a flexible and adaptable architecture appropriate for dynamic pervasive computing environments. We assume node-to-node connectivity in the network with common network/transport protocols such as TCP/IP. This approach uses a small subset of the nodes called collaborator that performs directory services to other nodes in the system. Here less mobile or nodes with high energy a nodes propose itself as a volunteer node in heterogeneous and uncertain (Fig. 2).

4.1 Trusted Clustering: Proposed Approach

The overall proposed approach is to dynamically organize the pervasive Sensor nodes clusters. Each cluster consists of one cluster head(C-Hd) node and an arbitrary number of clustered nodes(C-Nd). In each cluster, the C-Hd acts as a representative for its C-Nds and as nodes communicates their data over shorter

Fig. 2 Cluster



distances to their respective cluster heads. To save a substantial amount of energy, all the nodes that are not used continuously. The random selection of the cluster head node may obtain a poor clustering set of connections, and cluster head nodes may be redundant for some rounds of operation. Trust is calculated for a sensor node based on past interaction experiences given by neighbor nodes for assessing the reliability. It was also to measure the security of a node by evaluating whether a node is malicious or not. Trustworthy Architecture for such networks provides the trusted communication among the cluster nodes, based on trust and reputation formulations. Mobile pervasive Networks (MP-NETs) consist of a large number of relatively low powered mobile nodes, communicating in a network. Clustering is one of the techniques used to manage data exchange amongst interacting nodes. Each group of nodes has one or more elected Cluster Head(s) C-Hds, where all Cluster heads are interconnected for forming a communication with limited energy sources for longer period of time. Misbehaving nodes and cluster heads can drain energy rapidly and reduce the total life span of the network. To ensure a secure and trusted communication cluster heads with trusted information becomes critical for the overall performance. The Cluster head(s) selection algorithm based on an efficient trust model. This algorithm aims to elect trustworthy stable cluster head(s) that can provide secure communication via cooperative selfish nodes.

4.2 Cluster Formation Algorithm

Clustering algorithm partitions a network into different clusters, creating a network hierarchy in the network. A particular node is elected in a cluster to manage the cluster information is known as the cluster head, and the other nodes are its members.

1. Any node can be volunteer to imitate the cluster formation.
2. Calculate the available Energy expressed by equation

$$En_y = \frac{E_h - E_c}{E_h} \quad (1)$$

where $E_c = E_c + E_{req} + \square \geq 0$

3. Calculate the node stability or degree of mobility by the path planning.
4. Calculate the trust parameters as per pre-define threshold consider as a cluster member node.
5. The node with lowest mobility, high energy availability and highest trust as per above defined factors will be cluster head and will be responsible to provide service to each cluster node instead of each node itself.

Our Proposed cluster head selection algorithm is based on the analyzing the node misbehavior or compromising node detection based on the trust calculation. One of the essential operations in using clustering technique is to select cluster head among the nodes in the network and making a virtual group from the remaining nodes as a cluster around the cluster head node. In our proposed approach this done in a distributed manner, where nodes make autonomous decisions without any centralized control. The algorithm initially assumes that each sensor in the network becomes a cluster head with probability p . Each cluster head then advertises itself as a cluster head to the sensors within its radio range. This advertisement is forwarded to all sensors that are no more than k -hops away from the CH through controlled flooding.

Advantages and Limitations of Trusted Clustering: In our proposed approach, trust is calculated at two levels (a) trust at C-Hd level and (b) trust at clustered nodes CNs (Non-Cluster head node) level. Each C-Hd evaluates the cost of other C-Hds and C-Nds in its cluster while a CNs calculates other CNs in the same cluster in terms of trust value. The peer-to-peer trust costing is regularly updated based on direct or indirect observations. When two nodes are neighbors within a cluster, they evaluate each other based on direct observations. The C-Hd managers accomplish trust evaluation toward all C-Hds in the system. The selection of cluster head based on the most promising highest trust level or reader may refer protocols like HEED, LEACH [C-Leach] for a best possible solution. The description of these is outside the scope. If a C-Hds consumes more energy than a non-cluster head node is compromised, the more energy will be consumed to deal with attacks. Furthermore a selfish node consumes less energy than an unselfish node as its selfish behavior is reflected by stopping sensing functions and randomly dropping messages. Thus, the only secrecy of the system can be quickly sense and expel compromised nodes before a system failure. Considering the proposed approach in subjected area i.e. pervasive environment where the seamless communication relies on baseline technologies and may vary location to location and available infrastructure.

4.3 Honey Bee Mating for Trusted Cluster

Honey-bees mating is a swarm-based intelligence technique used in search optimization, inspired by the process of mating in real honey-bees. The behavior of honey-bees is the communication of their (1) Inherited potentiality. (2)

Environmental and physiological environments and (3) the social conditions of the colony. A typical honey-bee colony consists of a single egg laying long-lived queen (best bee), anywhere from zero to several thousand drones. The colony can be founded in two different ways as “self-governing origin” the colony starts with one or more reproductive females that build the nest, lay the eggs, and feed the larva’s. Later, division of labor takes place and the queen concentrates on egg laying and the workers in brood. A colony of bees is a large family of bees living in one bee-hive. A bee hive is like a big city with many “segments of the settlement”. The queen is the main vital member of the hive because she is the one that keeps the hive going by producing new queen and worker bees. With the help of approximately 18 males (drones), the queen bee will mate with multiple drones one time in her life over several days. The sperm from each drone is planted inside a pouch in her body. She uses the stored sperms to fertilize the eggs. Whether a honeybee will become a queen, a drone, or a worker, depends on whether the queen fertilizes an egg. Since she is the only bee in the colony that has fully developed ovaries, the queen is the only bee that can fertilize A queen bee may live up to 5 or 6 years, whereas worker bees and drones never live more than 6 months. Queens represent the main reproductive individuals which are specialized in eggs laying while Drones are the fathers of the colony. They are haploid and act to amplify their mothers’ genome without altering their genetic composition, except through mutation. Workers are specialized in brood care and sometimes lay eggs. Broods arise either from fertilized or unfertilized eggs. The mating process occurs during mating-flights far from the nest. A mating flight starts with a dance where the drones follow the queen and mate with her in the air. In a typical mating-flight, each queen mates with seven to twenty drones. In each mating, sperm reaches the spermatheca and accumulates there to form the genetic pool of the colony. In the mathematical representation, a drone is represented by a genotype and a genotype marker. Workers which are used to improve the brood’s genotype, represent a set of different heuristics. For example, at one-point of crossover heuristic, the crossover heuristic operator applies to the brood’s genotype with that of a randomly generated genotype where the crossover point is also selected at random. Each queen is characterized with a genotype, speed, energy, and a spermatheca with defined capacity. Spermatheca is defined as a repository of drones. In our proposes honey bee mating Queen is characterized by a fitness function based on trusted calculation based on above define parameters. The mapping of real honey bee and a pervasive network can be viewed in Table 4.

5 Conclusion and Future Work

In this paper, we proposed a bio-inspired trusted clustering for pervasive environment considering trustworthiness based on social and QoS trust parameters. In clustering technique If a C-Hds consumes more energy than a non-cluster head node the C-Hds is compromised, than more energy will be consumed to deal with

Table 4 Mapping of Natural Honey Bee Mating Pervasive Node's Mating

No#	Pervasive Node's (Bees) Mating	Natural Honey Bee Mating
1.	Nodes in pervasive adhoc Network	Bees in Hive
2.	Random node selection to broadcast as Cluster head with minimum mobility and maximum energy based on available base station or mobile base station information	Initial bee population
3.	Trust calculation to elect Cluster head with highest trust value (Queen)	Defining Fitness Function select Best Bee
4.	Drone Bee	Expected Cluster head list as per initial population Queen
5.	Working Bees	Heuristic Search Function
6.	Mating	Cross Over

trust calculation by evaluating peer nodes misbehavior. Furthermore a selfish node consumes less energy than an unselfish node as its selfish behavior is reflected by stopping sensing functions and randomly dropping messages. Thus, the only secrecy of the system can be quickly sense and expel compromised nodes before a system failure For Cluster formation we map subjective network with honey bee mating Honey-bees mating which a swarm-based intelligence technique. This technique is used in search optimization, inspired by the process of mating in real honey-bees to analyze the proposed approach in bee like network structure. As a further work simulation of proposed approach in dynamic scenario is ongoing. Identification of its application areas and implication is our future scope.

References

1. Weiser, M.: The computer for the 21st century. *Sci. Am.* **265**(3), 94–104 (1991)
2. Cho, H., Swami, A., Chen, I.R.: A survey on trust management for mobile ad hoc networks. *IEEE Commun. Surv. Tutor.* **13**(4), 562–583 (2011)
3. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless mi-crosensor networks. *IEEE Trans. Wireless Commun.* **1**(4), 660–670 (2002)
4. Almenare, F., Marin, A., Campo, C., Garcia, R.C.: PTM: a pervasive trust management model for dynamic open environments. In: *Proceedings of the 1st Workshop on Pervasive Security, Privacy and Trust*, Boston, August 2004
5. Almenarez, F., Marin, A., Campo, C., Garcia, R.C.: TrustAC: trust-based access control for pervasive devices. In: *Proceedings of the 2nd International Conference on Security in Pervasive Computing*. pp. 225–238, Boppard, Germany, April 2005
6. Almenarez, F., Marin, A., Diaz, D., Sanchez, J.: Developing a model for trust management in pervasive devices. In: *Proceedings of 4th IEEE Annual International Conference on Pervasive Computing and Communications*. pp. 267–271, Pisa, Italy, March 2006
7. Lopez, J., Roman, R., Agudo, I., Fernandez, C.G.: Trust Management Systems for Wireless Sensor Networks: Best Practices. *Comput. Commun.* **33**, 1086–1093 (2010)
8. Bao, F., Chen, R., Chang, M., 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.* **9**(2), 169–183 (2012)

9. Hsieh, M.Y., Huang, Y.M., Chao, H.C.: adaptive security design with malicious node detection in cluster-based sensor net-works. *Comput. Commun.* **30**, 2385–2400 (2007)
10. Bozorg Haddad, O., Afshar, A., Mariño, M.A., Adams, B.J.: Honey-bee mating optimization (HBMO) algorithm for optimal reservoir operation. *J. Franklin Inst.* **344**, 452–462 (2007)
11. Fathian, M., Amiri, B., Maroosi, A.: Application with honey-bee mating optimization algorithm on clustering. *Appl. Math. Comput.* **190**, 1502–1513 (2007)
12. Li, J.L., Gu, L.Z., Yang, Y.X.: A new trust management model for P2P networks. *J. Beijing Univ. Posts Telecommun.* **32**, 71–74 (2009)
13. Marmol, F.G., Perez, G.M.: Towards pre-standardization of trust and reputation models for distributed and heterogeneous systems. *Comput. Stand. Interfaces* **32**, 185–196 (2010)
14. Heinzelman, WR., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless micro-sensor net-works. In: *Proceedings of 33rd Hawaii International Conference on System Sciences*. pp. 1–10 (2000)
15. *IEEE Commun. Surv. Tutor.* 562–583 (2011)
16. M Wireless Conference (RAWCON'98), pp. 55–58 Aug 1998

Comparison of Adaptive Social Evolution and Genetic Algorithm for Multi-objective DTLZ Toolkit

Nitish Ghune, Vibhu Trivedi and Manojkumar Ramteke

Abstract Test problems play an important role when it comes to evaluate the performance of multi-objective evolutionary algorithms (MOEAs). Among a number of test problems available in literature, DTLZ toolkit has its distinct place. In this paper, this toolkit has been used to measure the potential of a newly developed algorithm based on social evolution, namely, Adaptive Social Evolution (ASE) algorithm. The results are compared with those of binary-coded elitist Non-Dominated Sorting Genetic Algorithm (NSGA-II) and found to be better in terms of convergence and computational time.

Keywords Adaptive social evolution · Genetic algorithm · DTLZ toolkit · MOEAs

1 Introduction

Since last two decades, the multi-objective optimization (MOO) of complex real world problems using evolutionary algorithms (EAs) has emerged as a widely accepted area of research and application. The popularity of EAs lies behind the fact that they are able to find a set of equally good solutions (Pareto front) in a single run. Multi-objective EAs (MOEAs) developed in the past are so successful

N. Ghune (✉) · V. Trivedi · M. Ramteke
Department of Polymer and Process Engineering, Indian Institute of Technology,
Roorkee, India

e-mail: nitishghune@gmail.com

V. Trivedi
e-mail: vibhutrivedi195@gmail.com

M. Ramteke
e-mail: ramtekemanoj@gmail.com

that these inspire researchers to create new algorithms continuously. There are various measures to compare these new algorithms with the existing ones. MOEAs are generally compared on the basis of their convergence speed. However, rapidly evolving technology and increasing computational speeds have established the flexibility of application as a principle criterion for the success of an MOEA. In this context, several hybrid algorithms are developed by amalgamating the useful properties of existing EAs and other search techniques such as fuzzy logic [1], neural network [2], particle swarm optimization [3], ant colony optimization [4], etc. Recently [5], a new algorithm namely, adaptive social evolution (ASE) to meet the above criteria is developed. The algorithm is based on the novel theme of 'social evolution by following heroes' and inherits the virtues of popular EAs like, genetic algorithm (GA), differential evolution (DE) and particle swarm optimization (PSO).

ASE mimics a common trait of social behavior, i.e. evolution of society by following heroes, which is supported by several historical facts [6–8]. Evidently, [7] the impact of heroes on society is found to be more pronounced in times of turmoil and war than in times of peace. This is because the society becomes weak and unstable and thus more vulnerable to revolution or change in these times. This useful observation is aptly incorporated by ASE to increase the convergence speed. The progress of algorithm is altruistic towards the use of better solutions in the starting generations in which the population is far from the optimality and resembles to unstable society. However, natural selection prevails as the algorithm approaches to optimality.

ASE has been successfully applied [5] to a number of MOO problems of varying nature including ZDT4 problem, travelling salesman problem, cone design problem and problems involving various important industrial processes. The aim of this work is to further testify the performance of ASE by applying it to more challenging test problems from existing MOO literature. Further, the results are compared with binary-coded elitist non-dominated sorting genetic algorithm, NSGA-II [9].

2 Description of ASE [5]

ASE comprises of following steps (see Pseudo-code): *initialization*, *fitness calculation*, *ranking*, *following heroes* (FH), *personalized deviation* (PD) and *elitism* [10].

The algorithm initializes with the generation of the population of individuals, N_p . The array of the decision variables resembling the characteristics represents an individual in a simulated social system. The values of these characteristics decide the goodness of a simulated individual. Firstly, the characteristics for N_p individuals are generated sequentially one after another within their prescribed lower and upper limits. The generation of j th characteristic of i th individual can be

random $\left[x_{i,j} = x_j^{Low} + Random\ Number \times \left(x_j^{High} - x_j^{Low} \right) \right]$, deterministic (strictly satisfying some constraints given in the problem) or heuristic (using local search, $x_{i,j} = x_{i,j}^{Local\ Search}$) in nature. The values of characteristics of an individual are then used in the fitness function to measure its goodness (fitness). The process is repeated for N_p individual one after another. Each individual is then ranked [10] according to its goodness.

In the next step, the individuals with the highest amount of goodness (having rank 1) are denoted as *heroes*. Every *hero* then inspires the rest of the population (including other heroes) using following hero (FH) operator on characteristic basis. This operator perturbs the original value of a given characteristic to that of a randomly selected *hero*. In this way all the characteristics of an individual are perturbed to the inspired characteristics using that of different heroes. All N_p members of population undergo this inspiration process one after another. In this, j th characteristic of i th individual is inspired as:

$$x_{i,j}^{New} = x_{i,j}^{Old} + Random\ Number \times \left(x_{h,j}^{Hero} - x_{i,j}^{Old} \right) \quad (1)$$

where, $x_{h,j}^{Hero}$ is j th characteristic of randomly selected *hero*.

The inspired individuals from FH operation are further subjected to personalized deviation (PD) operator. PD operator takes its inspiration from artificial intelligence. In this operator, a characteristic, randomly selected with very small probability is randomly $\left[x_{i,j}^{New} = x_j^{Low} + Random\ Number \times \left(x_j^{High} - x_j^{Low} \right) \right]$, deterministically deterministic (strictly satisfying some constraints given in the problem) or heuristically perturbed between the given upper and lower limits (using local search, $x_{i,j}^{New} = x_{i,j}^{Local\ Search}$). Solutions, thus generated are generally good. Although, some bad solutions are also generated (more prevalently in random generation), they are eliminated using concept of *elitism*. Other deterministic and heuristic search techniques can be added to this operator to increase its efficiency. PD operator increases the flexibility of ASE.

The original as well as perturbed population has both good and bad individuals. For optimization purpose, bad individuals must be removed from the population. ASE performs this ‘*screening*’ of bad individuals by a procedure known as *elitism* which was originally developed for GA. In this step, all N_p individuals from modified population and all N_p individuals from original population are mixed together to form a population of $2N_p$ individuals. These $2N_p$ individuals are again ranked and the best N_p individuals thus selected are called *elites*. These N_p - *elites* are used as the starting population for the next generation. This process continues until the algorithm finds optimal results or the count of user defined maximum generations is completed.

2.1 Pseudo-code of ASE

Initialization

For each Individual (1 to N_p)

```
{
    Initialize Individuals
}
```

Ranking Individuals

For each Generation (1 to Maximum Generations)

```
{
    Rank each Individual
    {
        Ranking and Crowding Distance
    }
    Select Heroes
    {
        If (Rank of Individual = 1) then
            Individual = Hero
    }
}
```

Following Heroes Operation

For each Individual (1 to N_p)

```
{
    If (Random Number  $\leq$  Probability of Following Heroes) then
    For each variable
    {
        Value of Variable = Current Value of Variable + Random Num-
        ber  $\times$  (Value of Variable in Randomly Selected Hero - Current
        Value of Variable)
    }
}
```

Personalized Deviation

For each Individual (1 to N_p)

```
{
    If (Random Number  $\leq$  Probability of Personalized Deviation) then
    For each variable
    {
        Value of Variable = Lower Bound of Variable + Random Num-
        ber  $\times$  (Upper Bound of Variable - Lower Bound of Variable)
    }
}
```

Elitism

Mix the Initial and the Final Population after Personalized Deviation

For each Individual (1 to $2N_p$)

```
{
    Ranking and Crowding Distance
}
Select the Elite Individual (1 to  $2N_p$ )
{
    Individual = Best Individual
}
}
```

Print Individuals

3 DTLZ Problems

Several MOO test problems are present in the literature [11] to evaluate the performance of newly developed algorithms. One of such suite of continuous test problems is DTLZ toolkit given by Deb et al. [12]. These problems are scalable to any number of objectives and thus provide the flexibility to test the algorithm for more rigorous conditions. Total nine problems are given in the test suit. From these only seven (DTLZ1–DTLZ7) are solved in this study (due to side constraints, DTLZ8 and DTLZ9 are omitted). The detailed description of these is given in the Table 1. Among these problems, DTLZ1–DTLZ4 comprise of multiple global optima, DTLZ5 and DTLZ6 gives degenerate Pareto optimal front and DTLZ7 exhibits disconnected Pareto optimal front [11].

The performance of EAs can be evaluated using several characteristics [11] of a given problem such as: *geometry of Pareto optimal front*, *modality*, *mapping of fitness landscape*, *parameter dependencies* and *bias*. The geometry of Pareto optimal front can be *concave*, *convex*, *linear*, *mixed*, *disconnected*, *degenerate*, or combination of these shapes. Finding a Pareto optimal front with different shapes as given above proves the robustness of the algorithm. *Modality* is the measure of optima associated with the given objective function i.e. an objective function with multiple local optima is called *multimodal* (more difficult case) while the one with single optima is called *unimodal*. The *mapping* between the Pareto optimal set and the Pareto optimal front may be *one-to-one* or *many-to-one*. For Pareto many-to-one problems, an EA faces difficulty in finding multiple or alternatively equivalent optima. These optima are important because their availability provides greater degrees of freedom to a decision maker in selecting the solutions. *Parameter dependency* is a very important characteristic to define the difficulty in optimization problem. If for all values of a parameter the set of global optimum is same then the given objective is separable with respect to the given parameter otherwise non-separable. *Bias* is the measure of variation in distribution of set of objective vectors in fitness space mapped for an evenly distributed sample of parameter vectors in the search space. A problem with bias is difficult to solve because it directly affects the convergence speed of EAs towards the Pareto optimal front. For instance, the fitness landscape could be biased in a way so that it becomes difficult to find global optima of a multimodal MOO problem.

Besides having these properties, the main advantage of test problems is that the results for these problems are readily available in open literature which makes the comparison of new EAs with the older ones easy.

Table 1 Details of DTLZ problems [11, 12]

Name of problem	Formulation (all objectives are to be minimized.)	Parameter domains	Characteristics			
			Geometry	Separability	Modality	Bias Many-to-one mapping
DTLZ1	$f_1 = \frac{1}{2}(1+g) \prod_{i=1}^{M-1} y_i$ $f_{m=2:M-1} = \frac{1}{2}(1+g) \left(\prod_{i=1}^{M-m} y_i \right) (1 - y_{M-m+1})$ $f_M = \frac{1}{2}(1+g)(1 - y_1)$ $g = 100 \left[k + \sum_{i=M}^n (y_i - 0.5)^2 - \cos(20\pi(y_i - 0.5)) \right]$	[0, 1]	Separable	Multimodal	No	Yes
DTLZ2	$f_1 = (1+g) \prod_{i=1}^{M-1} \cos(y_i \pi/2)$ $f_{m=2:M-1} = (1+g) \left(\prod_{i=1}^{M-1} \cos(y_i \pi/2) \right) \sin(y_{M-m+1} \pi/2)$ $f_M = (1+g) \sin(y_1 \pi/2); g = \sum_{i=M}^n (y_i - 0.5)^2$	[0, 1]	Concave	Unimodal	No	Yes
DTLZ3	Same as DTLZ2, except the equation for g is replaced from DTLZ1.	[0, 1]	Concave	Multimodal	No	Yes
DTLZ4	Same as DTLZ2, except all $y_i \in y$ are replaced by y_i^α , where $\alpha > 0$.	[0, 1]	Concave	Unimodal	Yes	Yes
DTLZ5	Same as DTLZ2, except all $y_2, \dots, y_{M-1} \in y$ are replaced by $\frac{1+2\epsilon y_i}{2(1+g)}$.	[0, 1]	-	Unimodal	No	Yes
DTLZ6	Same as DTLZ5, except the equation for g is replaced by $g = \sum_{i=M}^n y_i^{0.1}$.	[0, 1]	-	Unimodal	Yes	Yes

(continued)

Table 1 (continued)

Name of problem	Formulation (all objectives are to be <i>minimized</i> .)	Parameter domains	Characteristics			
			Geometry	Separability	Modality	Bias mapping
DTLZ1:	$k = 5$, DTLZ2-DTLZ6: $k = 10$, DTLZ7: $k = 20$; $n = k + M - 1$	$[0, 1]$	Dis-connected	Separable for f_M only	Unimodal except f_M	No
DTLZ7	$f_m = 1 : M - 1 = y_m$ $f_M = (1 + g) \left(M - \sum_{i=1}^{M-1} \left[\frac{f_i}{1 + g} (1 + \sin(3\pi f_i)) \right] \right)$ $g = 1 + 9 \sum_{i=M}^n y_i / k$					

4 Results and Discussion

The seven DTLZ problems are solved using ASE and the performance is compared with that of binary-coded NSGA-II. All the codes are written in FORTRAN 90. The codes are run on Intel Xeon ES-2640 0@2.50 GHz with 16 GB RAM computer operated with Windows 7. The values of ASE parameters namely, probability of FH, PD operators and size of the population are 0.9, 0.1 and 100 respectively. Those for NSGA-II parameters namely, crossover, mutation probabilities and length of binary string and size of the population are 0.85, 0.008, 20 and 100 respectively. All the problems are solved for 1,000 generations.

Generally, *box plots* [10] are used for analyzing the convergence as well as the distribution of the Pareto optimal solutions of MOO problems. In these plots, the *median* of the points should be lying close to the midpoint of the box for a uniform distribution. Also, the distribution of *outliers* in the *box plots* shows the searching effectiveness of an algorithm towards the extreme end of the Pareto optimal front. In general, MOO algorithms found to have difficulties in searching the extreme regions of the Pareto optimal front. Thus, for minimization problem lower the range ($f_{i,High} - f_{i,Low}$) of *box plot* of a given objective better is its convergence. The concepts of *box plots* are used to analyze the DTLZ1–DTLZ7 problems as shown in Fig. 1. The comparative analysis of the results is shown in Table 2. For DTLZ1 and DTLZ3, ASE is found to have considerably better convergence compared to NSGA-II as shown by lower height of the *boxes* plus *whiskers* (see Fig. 1). This is also confirmed by the Pareto optimal front plotted in Fig. 2.

For these problems, the ranges of objectives $\{f_1, f_2, f_3\}$ are given as: $\{2.01, 2.01, 2.01\}^{DTLZ1}$, $\{21.93, 21.93, 171.56\}^{DTLZ3}$ and that of ASE are: $\{0.43, 0.43, 0.50\}^{DTLZ1}$, $\{11.27, 8.83, 69.56\}^{DTLZ3}$, respectively which clearly illustrate better convergence using ASE as compared to NSGA-II. For all other remaining problems, convergence using both ASE as well as NSGA-II is comparable as shown in Fig. 2. Performance of the algorithms are further compared using the concept of *median* (has to lie perfectly at the mid of the boxes for best distribution) in *box plots* and the distribution of end points (better distribution of end points should be present in the *box plots*). Based on the distribution of the points in Fig. 1, ASE performs better in terms of *median* and end points for problems DTLZ2 and DTLZ4. For DTLZ5 and DTLZ6 both ASE and NSGA-II performs better in terms of *median* whereas the distribution of end points is worse in both cases. In contrast to this, results of DTLZ7 are better only in terms of distribution of end points for both ASE and NSGA-II.

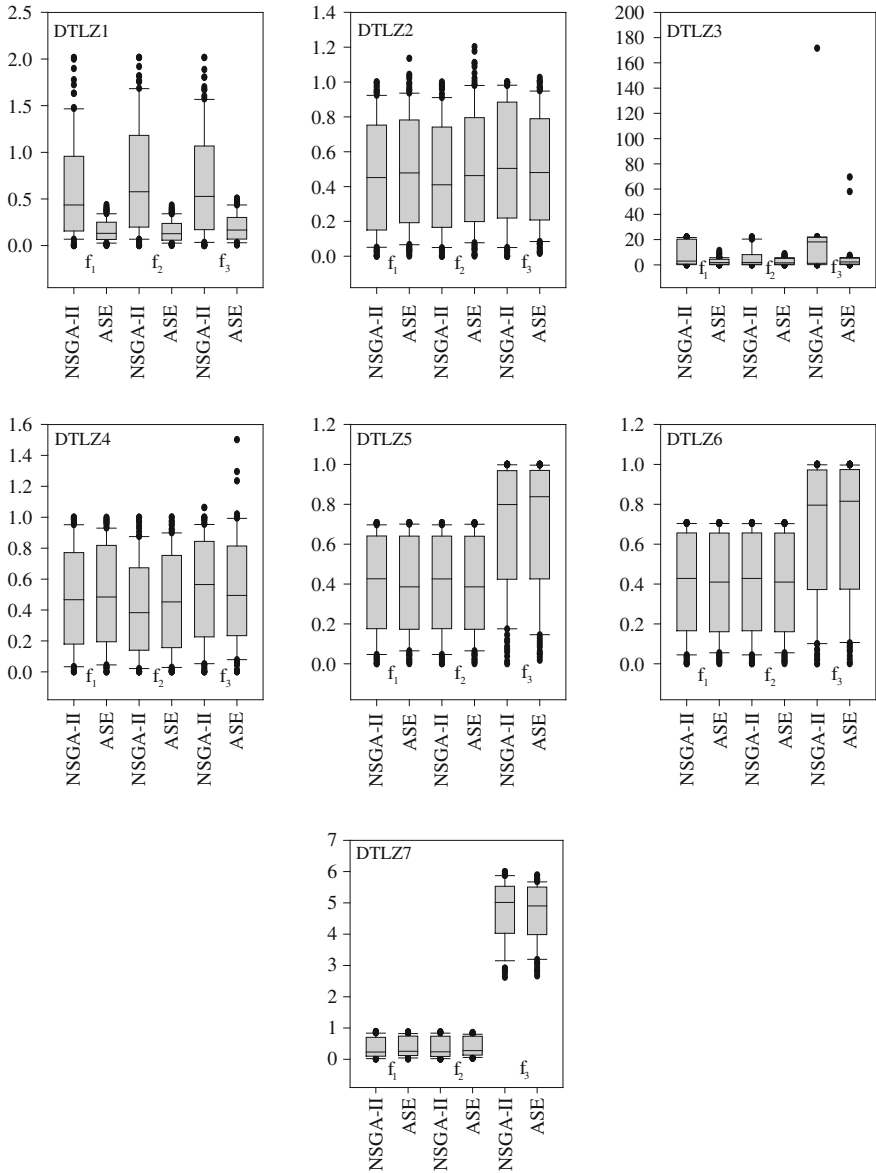


Fig. 1 Box plots for all DTLZ problems

Table 2 Comparison of ASE and NSGA-II results

Problems	Convergence		Distribution of Pareto optimal points			
			Performance based on <i>median</i>		Performance based on <i>distribution of end points</i>	
	ASE	NSGA-II	ASE	NSGA-II	ASE	NSGA-II
DTLZ1	✓	×	—	—	—	—
DTLZ2	✓	✓	✓	×	✓	×
DTLZ3	✓	×	—	—	—	—
DTLZ4	✓	✓	✓	×	✓	×
DTLZ5	✓	✓	✓	✓	×	×
DTLZ6	✓	✓	✓	✓	×	×
DTLZ7	✓	✓	×	×	✓	✓

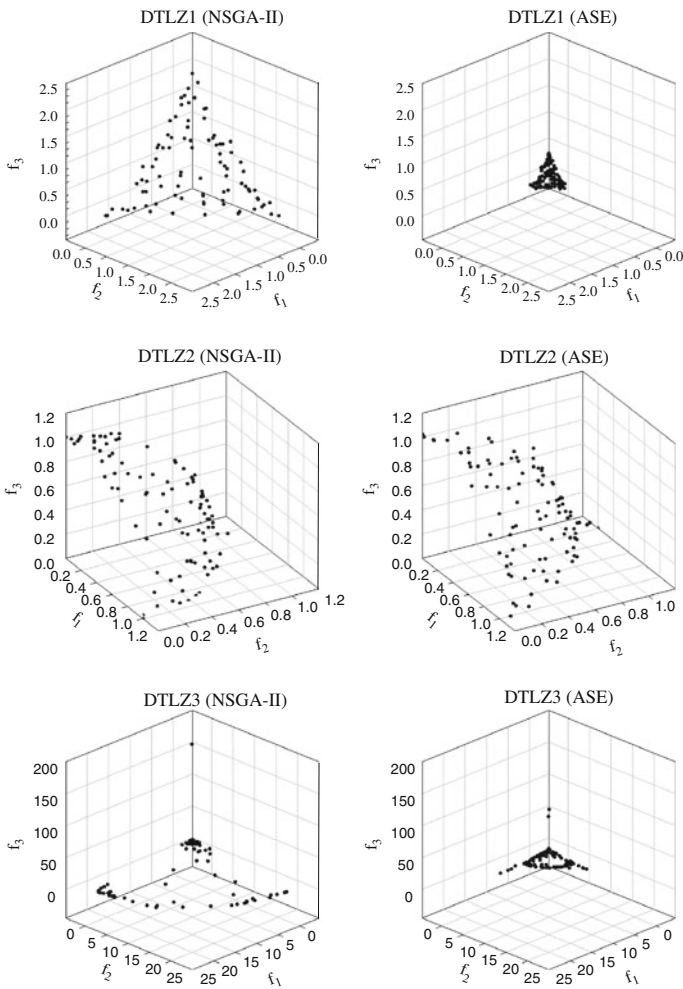


Fig. 2 Pareto plots for problems DTLZ 1–7

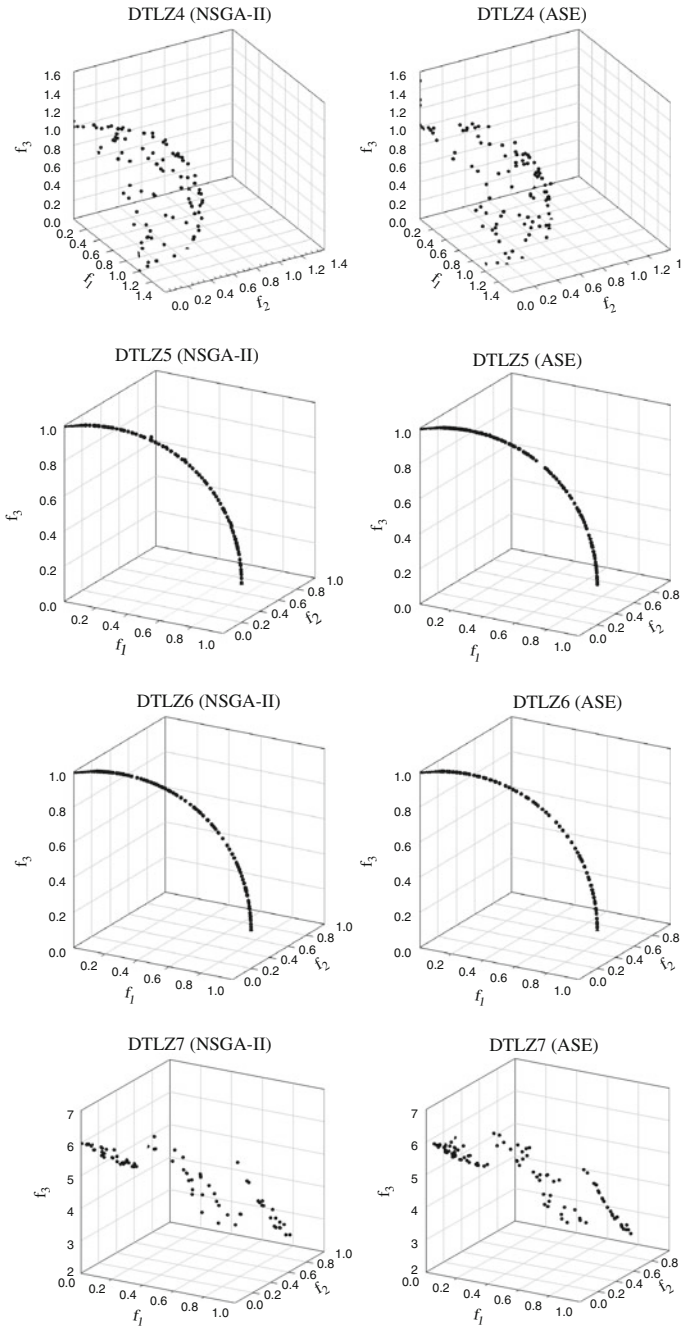


Fig. 2 (continued)

5 Conclusions

Seven benchmark problems from DTLZ test suite are solved using the ASE algorithm. The results are found to be better in comparison to that of NSGA-II. Moreover, ASE takes comparatively lower computational time than NSGA-II for any given problem. This is due to the fact that ASE has less number of operations as compared to genetic algorithm. The successful application of ASE to the DTLZ problems confirms its potential and it paves the way for the future applications of ASE to the real world and industrial problems.

References

1. Herrera, F., Lozano, M.: Adaptation of genetic algorithm parameters based on fuzzy logic controllers. In: Herrera, F., Verdegay, J.L. (eds.) *Genetic Algorithms and Soft Computing*, pp. 95–125. Physica-Verlag, Wurzburg (1996)
2. Wang, L.: A hybrid genetic algorithm-neural network strategy for simulation optimization. *Appl. Math. Comput.* **170**, 1329–1343 (2005)
3. Shi, X.H., Liang, Y.C., Lee, H.P., Lu, C., Wang, L.M.: An improved GA and a novel PSO-GA-based hybrid algorithm. *Inf. Process. Lett.* **93**, 255–261 (2005)
4. Tseng, L.Y., Liang, S.C.: A hybrid meta-heuristic for the quadratic assignment problem. *Comput. Optim. Appl.* **34**, 85–113 (2005)
5. Ramteke, M.: Optimization using adaptive social evolution by following heroes. Submitted (2013)
6. Jeffers, H.P.: *The 100 Greatest Heroes: Inspiring Profiles of One Hundred Men and Women Who Changed the World*. Citadel Press, New York (2003)
7. Grinin, L.E.: The role of an individual in history: a reconsideration. *Soc. Evol. Hist.* **9**, 95–136 (2010)
8. Carlyle, T.: *On Heroes, Hero Worship, and the Heroic in History*. University of California Press, Berkeley (1993)
9. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002)
10. Deb, K.: *Multi-objective Optimization Using Evolutionary Algorithms*, 1st edn. Wiley, Chichester (2001)
11. Huband, S., Hingston, P., Barone, L., While, L.: A review of multiobjective test problems and a scalable test problem toolkit. *IEEE Trans. Evol. Comput.* **10**, 477–506 (2006)
12. Deb, K., Thiele, L., Laumanns, M., Zitzler, E.: Scalable test problems for evolutionary multi-objective optimization. KanGAL Report 2 001 001, Kanpur Genetic Algorithms Lab (KanGAL), Indian Institute of Technology, Kanpur, India (2001)

A Reference Watermarking Scheme for Color Images Using Discrete Wavelet Transform and Singular Value Decomposition

Ravinder Katta, Himanshu Agarwal and Balasubramanian Raman

Abstract This paper presents a reference watermarking scheme in which both original and watermark images are RGB color images. Reference images are obtained from red, green and blue components of the original image by using Discrete Wavelet Transform (DWT) and Directive contrast. Singular Value Decomposition (SVD) is applied on each of the component images (Red channel, Green channel and Blue channel) of watermark and reference images (red, green and blue). Watermark is embedded by modifying the singular values of the reference images with the singular values of the watermark component images. A reliable watermark extraction scheme is developed. Experimental results show that the proposed scheme survives against common signal/image processing attacks.

Keywords Discrete wavelet transform (DWT) · Singular value decomposition (SVD) · Color Image · Reference watermarking

1 Introduction

With the growth of internet and digital technology, the distribution and exchange of digital media have become very easy. At the same time, these have raised many concerns about copyright infringement. Therefore, digital content providers feel insecure to distribute their copyrighted multimedia (text, image, audio and video)

R. Katta (✉) · H. Agarwal · B. Raman
Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee
247667, India
e-mail: kravidma@iitr.ac.in

H. Agarwal
e-mail: him11dma@iitr.ac.in

B. Raman
e-mail: balarfma@iitr.ac.in

content over the internet. For such scenarios, digital watermarking has emerged as an efficient solution for copyright protection.

Digital watermarking is a technique of embedding some data into a multimedia. The embedded data is known as ‘watermark’ and the media in which watermark is embedded is called original/host/cover media. The embedded watermark can later be extracted or detected for various purposes such as copyright protection, data authentication, transaction tracking, broadcast monitoring, proof of ownership, medical imaging and copy control. A watermarking scheme [1] consists of two algorithms namely, watermark embedding algorithm and watermark extraction algorithm. In watermark embedding algorithm, watermark is embedded into a multimedia. In watermark extraction algorithm, watermark is extracted/detected from the watermarked media.

In many applications, a part of original image is available at watermark extraction algorithm instead of complete original image. In such applications, reference watermarking schemes [2–4] are appropriate choice as watermark extraction algorithm of these schemes require reference image (reference image is a small derived part of original image) instead of original image. In reference watermarking, watermark is embedded in reference image. Most of the watermarking schemes have been developed for gray scale images. Watermarking of color images are also equally important.

Bhatnagar et al. [2] proposed a reference watermarking scheme for gray scale images. We have extended the work of Bhatnagar et al. [2]. In this paper, we propose reference watermarking scheme for color images. A color watermark is embedded in a color image. The proposed watermarking scheme is semi-blind as watermark extraction algorithm requires a part of information of original image.

The rest of the paper is organized as follows. In Sect. 2 proposed watermarking technique is discussed. In Sect. 3 experimental results are discussed. Conclusions are made in Sect. 4.

2 Proposed Watermarking Technique

In this section, we present a watermarking scheme for color images. First RGB color image is taken as original image (I). Red (I_R), Green (I_G) and Blue (I_B) components are extracted from I . Discrete Wavelet Transform (DWT) [5] is applied on (I_R), (I_G) and (I_B) separately. At level 1, DWT decomposes an image into four sub bands to be denoted by LL, HL, LH and HH. The low frequency sub band is represented by LL where as the high frequency sub bands are represented by HL, LH and HH. These sub bands are named as A, H, V and D respectively. Daubechies filter coefficients (Db2) are used in DWT. Based on directive contrast [6], three reference images (I_R^{ref}), (I_G^{ref}), (I_B^{ref}) are obtained. A color image (W) is taken as watermark. Red component (W_R), Green component (W_G) and Blue component (W_B) are extracted from W . SVD [7] is applied on (I_R^{ref}), (I_G^{ref}), (I_B^{ref}) as well on (W_R), (W_G), (W_B)

Watermark is embedded by modifying the singular values of (I_R^{ref}) , (I_G^{ref}) with the singular values of (W_R) , (W_G) , (W_B) respectively. Watermark embedding strength is controlled by the parameter α , $\alpha \in (0,1]$. We set $\alpha = 0.035$.

2.1 Watermark Embedding Algorithm

Mathematically original image, watermark image, watermarked image and extracted watermark image can be written as follows:

Original image $(I) = \{I_\beta(i,j)/i = 0, 1, 2, \dots, M-1, j = 0, 1, 2, \dots, N-1, \beta \in \{R, G, B\}\}$

Watermark image $(W) = \{W_\beta(i,j)/i = 0, 1, 2, \dots, M_1-1, j = 0, 1, 2, \dots, N_1-1, \beta \in \{R, G, B\}\}$

Watermarked image $(I') = \{I'_\beta(i,j)/i = 0, 1, 2, \dots, M-1, j = 0, 1, 2, \dots, N-1, \beta \in \{R, G, B\}\}$

Extracted watermark image $(W') = \{W'_\beta(i,j)/i = 0, 1, 2, \dots, M_1-1, j = 0, 1, 2, \dots, N_1-1, \beta \in \{R, G, B\}\}$

1. Apply 3 level DWT on I_β s, $\beta = R, G, B$ separately and the resultant images are denoted by $I_{\beta,3}^\theta$ where $\theta = A, H, V, D$.
2. Define a key $(K) = I_{\beta,3}^{select}$, where $select \in \{A, H, V, D\}$ We set $select = D$.
3. Apply level DWT on $I_{\beta,3}^{select}$, $\beta = R, G, B$ and denote the resultant images by $I_{\beta,4}^{select,\theta'}$, where $\theta' = A, H, V, D$.
4. Find the Directive contrast for all high frequency sub bands (for $\theta' = H, \theta' = V, \theta' = D$) and it is denoted by $C_{\beta,4}^{select,\theta'}$.
5. Initialize high frequency sub bands values to zero whose directive contrast are less than the threshold. The threshold directive contrast is given as follows:

$$T_{\beta,4}^{select,\theta'} = Sort(P \times S_{\beta,4}^{select,\theta'}) \quad (1)$$

where $Sort(O)$ are the sorted directive contrast, $S_{\beta,4}^{select,\theta'}$ is the size of the sub band.

P is the percentage of the wavelet coefficients which we want to retain. We set $P = 1$.

6. Apply 1 level Inverse Discrete Wavelet Transform (IDWT) to form the reference images, to be denoted by $I_{\beta,3}^{ref}$, $\beta = R, G, B$.
7. Apply SVD on each $I_{\beta,3}^{ref}$ and W_β , $\beta = R, G, B$ as follows.

$$I_{\beta,3}^{ref} = U_{\beta,3}^{ref} S_{\beta,3}^{ref} (V_{\beta,3}^{ref})^T \quad (2)$$

$$W_{\beta} = U_{W_{\beta}} S_{W_{\beta}} V_{W_{\beta}} \quad (3)$$

8. Modify the singular values of each $I_{\beta,3}^{ref}$ with the singular values of each W_{β} as

$$\sigma_{I_{\beta,3}^{ref}}^* = \sigma_{I_{\beta,3}^{ref}} + \alpha \sigma_{W_{\beta}} \quad (4)$$

where α controls the watermark embedding strength and $(\sigma_{I_{\beta,3}^{ref}})$, $(\sigma_{W_{\beta}})$ are the singular values of $I_{\beta,3}^{ref}$, W_{β} respectively.

9. Obtain modified reference images as

$$(I_{\beta,3}^{ref})^* = U_{I_{\beta,3}^{ref}} S_{I_{\beta,3}^{ref}}^* (V_{I_{\beta,3}^{ref}})^T, \quad (5)$$

where $S_{I_{\beta,3}^{ref}}^* = \text{diag}(\sigma_{I_{\beta,3}^{ref}}^*)$.

10. Replace $I_{\beta,3}^{select}$ by $(I_{\beta,3}^{ref})^*$ in step 1 and then apply 3 level Inverse Discrete Wavelet Transform on the resultant images.
11. Concatenate the images which are given by step 10 to obtain the Watermarked image (I').

2.2 Watermark Extraction Algorithm

The steps involved in watermark extraction algorithm are as follows.

1. Extract I'_R, I'_G, I'_B component images from the watermarked image (I').
2. Apply 3 level DWT on $I'_{\beta}, \beta = R, G, B$ separately. Based on K , which is saved during watermark embedding process, select the sub bands to obtain the watermarked reference images as follows.

$$I_{\beta,3}^{ref} = I_{\beta,3}^{select}$$

3. Perform Singular Value Decomposition on original reference images ($I_{\beta,3}^{ref}$) and watermarked reference images ($I_{\beta,3}^{select}$) separately.

$$I_{\beta,3}^{ref} = U_{I_{\beta,3}^{ref}} S_{I_{\beta,3}^{ref}} (V_{I_{\beta,3}^{ref}})^T \quad (6)$$

$$I_{\beta,3}^{select} = U_{I_{\beta,3}^{select}} S_{I_{\beta,3}^{select}} (V_{I_{\beta,3}^{select}})^T \quad (7)$$

4. Extract the singular values of watermark image as follows:

$$\sigma_{W, \beta}^{ext} = \frac{\sigma_{I_{\beta,3}^{select}} - \sigma_{I_{\beta,3}^{ref}}}{\alpha} \quad (8)$$

where $\sigma_{I_{\beta,3}^{ref}}$, $\sigma_{I_{\beta,3}^{ext}}$ are singular values of watermarked reference images and original reference images respectively.

5. Obtain the estimate of watermark image as

$$W_{\beta}^{ext} = U_{W_{\beta}}(S_{W_{\beta}}^{ext})(V_{W_{\beta}})^T \quad (9)$$

where $S_{W_{\beta}}^{ext} = \text{diag}(\sigma_{W_{\beta}}^{ext})$, $\beta = R, G, B$.

6. By Concatenating $W_{\beta}^{ext}s$, $\beta = R, G, B$, we obtain extracted watermark.

3 Results and Discussions

3.1 Data Set

In order to explore the performance of the proposed watermarking scheme, MATLAB platform is used and experiments are performed on different images (Fig. 1a) of size 512×512 , namely, Lena, Zelda, Mandrill, Pepper, Living room, Lake, Tree and Airplane. Seven different color images (Fig. 1b) of size 64×64 , namely, IEEE Logo, IITR Logo, RAVI, University Logo, College Logo, IITR and CVGIP are used as watermark images. IITR Logo is embedded in Zelda and Airplane images. IEEE Logo is embedded in Lena image. RAVI, University Logo, College Logo, IITR, CVGIP are embedded in Mandrill, Pepper, Living room, Lake, Tree images respectively. Watermarked image (Fig. 1c) quality is measured using Peak Signal-to-Noise Ratio (PSNR) and the similarity of the extracted watermark (Fig. 1d) is measured using Correlation Coefficient (CC) [2]. Higher the PSNR value result in lower the image degradation. The PSNR of watermarked images and CC values of extracted watermark images are provided in Table 1. The PSNR between the original and watermarked image is computed as follows:

$$PSNR = 10 \log_{10} \left(\frac{I_{\max}^2}{MSE} \right) \quad (10)$$

where I_{\max} is the maximum intensity value of image (I). MSE is the Mean Square Error obtained by the formula

$$MSE = \frac{1}{MN \times 3} \sum_{i=1}^M \sum_{j=1}^N ([I_R(i,j) - I'_R(i,j)]^2 + [I_G(i,j) - I'_G(i,j)]^2 + [I_B(i,j) - I'_B(i,j)]^2) \quad (11)$$

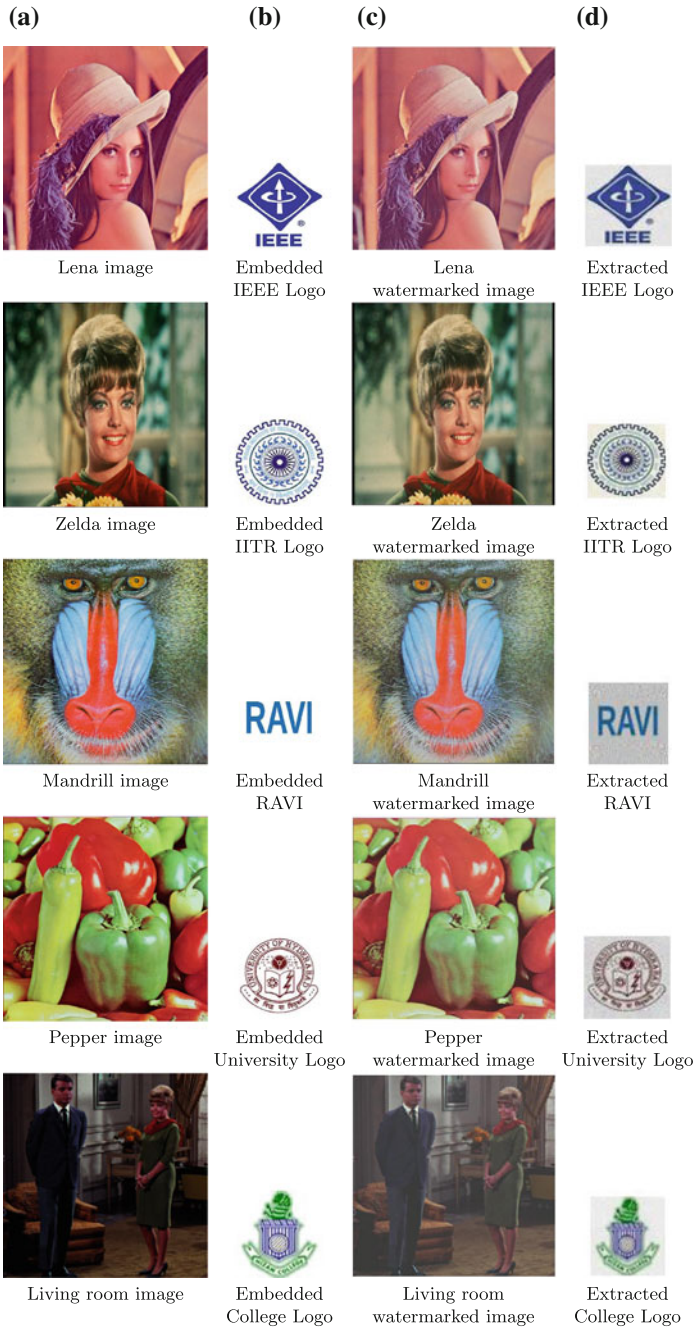


Fig. 1 a Original test images. b Embedded watermarks. c Watermarked images. d Extracted watermarks

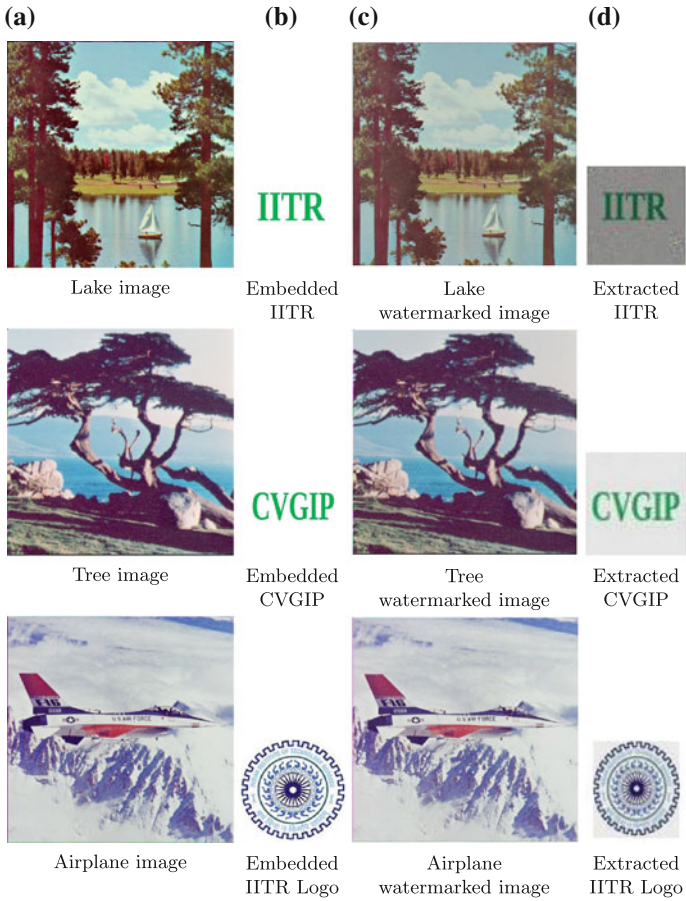


Fig. 1 (continued)

Table 1 PSNR and correlation Coefficient values of test images

Test image	Embedded watermark	PSNR (dB)	Correlation coefficient
Lena	IEEE logo	39.9042	0.9996
Zelda	IITR logo	43.1066	0.9994
Mandrill	RAVI	36.4011	0.9989
Pepper	University logo	39.2070	0.9992
Living room	College logo	41.9876	0.9998
Lake	IITR	36.0837	0.9983
Tree	CVGIP	38.5934	0.9999
Airplane	IITR logo	41.6743	0.9998

3.2 Results

In order to discuss the robustness of the proposed watermarking scheme, the watermarked image is attacked by a variety of common image processing attacks namely, average and median filtering, Gaussian and salt & pepper noise addition, motion blur, rotation, resizing and histogram equalization. After these attacks on watermarked image, the extracted watermark is compared with the original one. For robustness analysis, Zelda watermarked image is used, since this image has the highest PSNR value among all test images.

Filtering. Filtering is one of the most common attack in watermarking. An average filter of size 11×11 is applied on Zelda watermarked image. The resultant image is shown in Fig. 2a. The watermark extracted from Fig. 2a is shown in Fig. 2b.

A median filter of size 11×11 is applied on Zelda watermarked image. The obtained image is shown in Fig. 2c. The watermark extracted from Fig. 2c is shown in Fig. 2d.

Addition of Noise. Adding noise is another method to estimate the robustness of the watermarking scheme. Addition of noise is responsible for the degradation and distortion of the image. Gaussian noise (with mean zero and variance 0.01) is added to Zelda watermarked image. The resultant image is shown in Fig. 3a. The watermark extracted from Fig. 3a is shown in Fig. 3b.

Salt & Pepper noise (with pixel density 0.05) is added to Zelda watermarked image. The obtained image is shown in Fig. 3c. The watermark extracted from Fig. 3c is shown in Fig. 3d.

Motion Blur. Motion blur is performed on Zelda watermarked image. The motion blur length (L) and angle (θ) are 20 pixels and 45° respectively. The watermarked image after motion blur ($L = 20$ pixels, $\theta = 45^\circ$) is shown in Fig. 4a. The watermark extracted from Fig. 4a is shown in Fig. 4b.

Rotation. Zelda watermarked image is rotated (in anti-clock wise direction) by 50° . Zelda watermarked image after rotation is shown in Fig. 4c. The watermark extracted from Fig. 4c is shown in Fig. 4d.

Resizing. First the size of Zelda watermarked image is reduced to 128×128 and again brought to its original size 512×512 . The resultant image is shown in Fig. 5a. The watermark extracted from Fig. 5a is shown in Fig. 5b.

Histogram Equalization. We tested our proposed watermarking method by performing histogram equalization. Zelda watermarked image after applying histogram equalization is shown in Fig. 5c. The watermark extracted from Fig. 5c is shown in Fig. 5d.

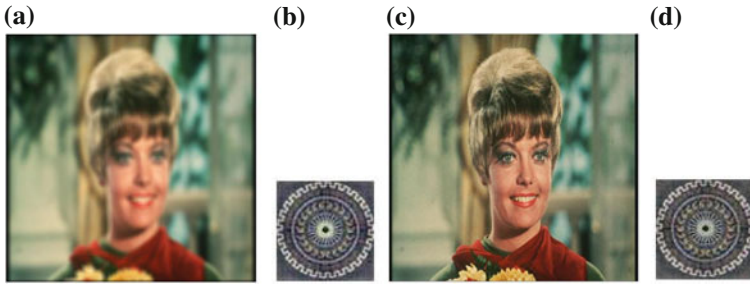


Fig. 2 **a** Zelda watermarked image after applying average filter of size 11×11 (PSNR = 24.6007 dB). **b** Watermark extracted from **a** (CC = -0.6397). **c** Zelda watermarked image after applying median filter of size 11×11 (PSNR = 28.5912 dB). **d** Watermark extracted from **c** (CC = -0.6086)

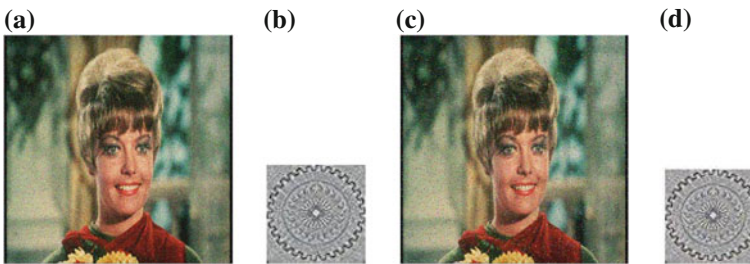


Fig. 3 **a** Zelda watermarked image with Gaussian noise (PSNR = 19.5964 dB). **b** Watermark extracted from **a** (CC = 0.5116). **c** Zelda watermarked image with salt and pepper noise (PSNR = 17.8555 dB). **d** Watermark extracted from **c** (CC = 0.4545)

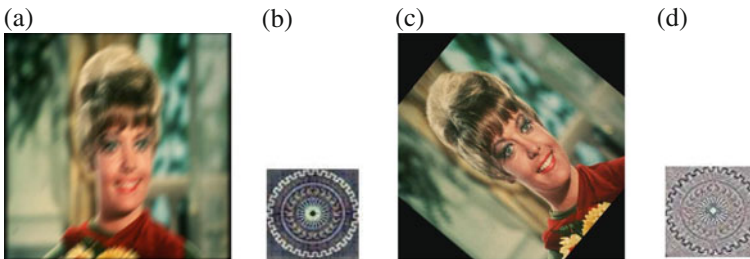


Fig. 4 **a** Zelda watermarked image ($L = 20$ pixels, $\theta = 45^\circ$) (PSNR = 25.2191 dB). **b** Watermark extracted from **a** (CC = -0.5627). **c** Zelda watermarked image after 50° rotation (PSNR = 10.4313 dB). **d** Watermark extracted from **c** (CC = 0.2115)

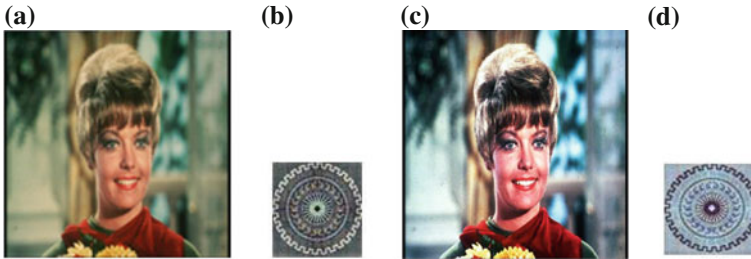


Fig. 5 a Zelda watermarked image after resizing (PSNR = 30.0436 dB). b Watermark extracted from a (CC = -0.1018). c Zelda watermarked image after performing histogram equalization (PSNR = 16.2654 dB). d Watermark extracted from Fig. 5c (CC = 0.8949)

4 Conclusions

A semi-blind reference watermarking scheme is presented for color images in which both the original image and the watermark images are color images. The salient features of the proposed watermarking scheme are as follows.

1. There is no visual degradation between watermarked images and original images.
2. Extracted watermarks are exactly same as that of the embedded watermarks.
3. Several attacks (average filtering, median filtering, Gaussian noise addition, salt & pepper noise addition, motion blur, rotation, resizing and histogram equalization) have been performed on watermarked images. After each attack, the extracted watermark is recognizable, which proves the robustness of proposed watermarking scheme.

Acknowledgments The authors, Ravinder Katta & Himanshu Agarwal gratefully acknowledge the financial support of the University Grants Commission, New Delhi, India through their Junior Research Fellowship (JRF), Senior Research Fellowship (SRF) Schemes respectively for their research work.

References

1. Arnold, M.K., Schmucker, M., Wolthusen, S.D.: Techniques and Applications of Digital Watermarking and Content Protection. Artech House, London (2003)
2. Bhatnagar, G., Raman, B.: A new robust reference watermarking scheme based on DWT-SVD. *Comput. Stand. Interfaces* **31**(5), 1002–1013 (2009)
3. Joo, S., Suh, Y., Shin, J., Kikuchi, H., Cho, S.J.: A new Robust watermark embedding into wavelet DC components. *ETRI J.* **24**(5), 401–404 (2002)
4. Liu, J.L., Lou, D.C., Chang, M.C., Tso, H.K.: A robust watermarking scheme using self-reference image. *Comput. Stand. Interfaces* **28**(3), 356–367 (2006)

5. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989)
6. Toet, A., Van Ruyven, L.J., Valeton, J.M.: Merging thermal and visual images by a contrast pyramid. *Opt. Eng.* **28**(7), 287789 (1989)
7. Liu, R., Tan, T.: An SVD-based Watermarking Scheme for Protecting Rightful Ownership. *IEEE Trans. Multimedia* **4**(1), 121–128 (2002)

Local Search Methods for the Winner Determination Problem in Multi-Unit Combinatorial Auctions

Abdellah Rezoug and Dalila Boughaci

Abstract In this paper, we are interested in the winner determination problem in the multi-unit combinatorial auctions (MU-WDP). In this type of auctions, a set of units of goods has to be auctioned. Bidders request the number of desired units of each good and the total price for the complete bid. Each bid has to be discarded or fully accepted. The objective of the auctioneer is to maximize its revenue. The MU-WDP is known to be NP-hard. In this paper, we propose three local search methods: a local search (LS), a tabu search (TS) and a stochastic local search (SLS) for the MU-WDP. The proposed methods are evaluated on some benchmark problems. The experimental study shows that the SLS algorithm is able to find good solutions for the multi-unit winner allocation compared to both LS and TS methods. Further, experiments against the CPLEX 12.5, show the effectiveness of the proposed SLS method in finding good quality solutions in faster time.

Keywords Multi-unit combinatorial auctions · Winner determination problem · Optimization problems · Local search · Tabu search · Stochastic local search · CPLEX

1 Introduction

The Combinatorial Auction (CA) is an efficient mechanism in which a seller is faced with a set of price offers for various bundles of goods. The aim is to allocate the goods in a way that maximizes the auctioneer revenue. In multi-unit

A. Rezoug (✉) · D. Boughaci
USTHB- FEI- Department of Computer Science, Laboratory of Artificial Intelligence LRIA,
BP 32 EL-Alia, Beb-Ezzouar 16111 Algiers, Algeria
e-mail: abdellah.rezoug@gmail.com

D. Boughaci
e-mail: dboughaci@usthb.dz

combinatorial auctions, each good has a set of occurrences. The buyer instead to bid for a set of goods, he bids for a set of bundle of goods with a price. The multi-unit winner determination problem (MU-WDP) consists in selecting the subset of bids which maximizes the revenue of sellers.

The combinatorial auction is a mechanism that has proven its effectiveness in solving many problems such as resources' allocation in multi-agent systems, the auctions for radio spectrum rights [6], airport slot allocations [21], transportation services [22, 15], course registrations [27], and commercial time slot allocations [29].

Several approaches have been proposed for the single-unit combinatorial auctions version where the number of goods for each bid is exactly equal to one, among these approaches, we cite: Branch on Bids (BoB) [31], Combinatorial Auctions BoB (CABoB) [32], Branch-on-Price [23], Branch-and-Cut [24], Combinatorial Auction Structured Search (CASS) [9], Almost of these solutions are based on the Branch-and-Bound searching method. Other methods are investigated for the single-unit winner determination problem such as: the dynamic programming [28], the linear programming [26], the integer programming [1], the constraint programming to solve the Vickrey combinatorial auction [19].

In order to handle large scale instances of the winner determination problem, the non-exact methods are often preferable. The heuristic methods are then largely used. We can cite: the local search approaches for example: Hybrid Simulated Annealing (SAGII) [12, 13], Casanova [17], the stochastic local search [4] and the mimetic algorithm [5]. We can also find other approaches like: Artificial Fish-Swarm Algorithm (AFSA), the ant colony algorithm [33] and the genetic algorithm hybridized with the ant colony algorithm [7].

Contrary to single-unit WDP, there are few works proposed for solving the multi-unit MU-WDP. Among them, we can find a Branch-and-Bound based algorithm [11], Combinatorial Auctions Multi-Unit Search CAMUS [25] which is a version of CASS adapted to the multi-unit case. An algorithm based on the Particle Swarm Optimization (PSO) approach [8], a local search algorithm [30], a heuristic suggestion based on a Lagrangian relaxation [14].

In this work, we proposed a Stochastic Local Search for solving the multi-unit version of the winner determination problem. We implemented two other meta-heuristics, the simple Local Search (LS) and the Tabu Search (TS) for comparison purposes. The three methods are compared in term of solutions quality and CPU time response. We tested them on ten different benchmarks that we have generated randomly. We then tested the behaviour of the algorithms relatively to the running time and the revenue of the best allocation obtained by the algorithms. To measure the performance of the proposed SLS, we compared it against the CPLEX solver version 12.5.

The rest of this paper is organized as follows. [Section 2](#) gives a background on the problem definition. [Section 3](#) presents our contributions for the multi-unit WDP. The implementation and some experimental results are given in the fourth section. Finally, we conclude and give some future works.

2 Problem Definition

In multi-unit combinatorial auctions, for each good i , we have some number of units q_i of this good for sale, contrary to the single-unit combinatorial auctions where there is only one unit of each good.

For example, suppose there are three goods g_1, g_2 and g_3 . We have 10 units of good g_1 , 4 units of good g_2 and 5 units of good g_3 for sale.

We have the following bids:

- $\{(8g_1, 4g_2, 2g_3); 20\}$
- $\{(2g_1, 2g_2, 4g_3); 10\}$
- $\{(5g_1, 1g_3); 8\}$

This means, for example, that the first bidder wants to buy 8 units of g_1 , 4 units of good g_2 and 2 units of g_3 , and is willing to pay 20 for this.

In general the problem can be stated as follows:

Let us consider G a set of m goods, $G = \{g_1, g_2 \dots g_m\}$ to be auctioned, and μ_i denote the number of available units of good i for sale. Let us consider B a set of n bids, $B = \{B_1, B_2 \dots B_n\}$.

A bid $B_j = \left((S_j^1, S_j^2 \dots S_j^m), P_j \right)$ where $S_j^k \geq 0$ is the requested number of units of good k , and P_j is the price of $B_j (P_j \geq 0)$. Finally the decision variables are defined as follows: $x_j = 1$ if the bid B_j is accepted (a winning bid), and $x_j = 0$ otherwise (a losing bid).

The WDP in multi-unit auctions can be modelled as the following integer program:

$$\text{Maximize } \sum_{j=1}^n P_j \cdot x_j$$

Under the constraints:

$$\sum_{j=1}^n S_j^i \cdot x_j \leq \mu_i$$

$$i \in \{1 \dots m\} \quad x_j \in \{0, 1\}$$

The MU-WDP is the problem of finding an allocation of winning bids that maximizes the auctioneer's revenue under the constraint that for any good i , the sum of units of i over all the winning bids does not exceed μ_i .

3 The Proposed Approaches for MU-WDP

In this section, we propose three local search methods for solving the MU-WDP in combinatorial auctions.

In the following we start with the main shared components of the three proposed approach: the solution representation, the Random Key encoding strategy used to generate initial solution and the objective function used to measure the quality of solutions. Then, we present the LS, TS and SLS proposed methods for the MU-WDP.

3.1 The Solution Representation

A solution for the WDP is an allocation S which can be represented by a vector with a variable length. Each of those components S_i receives the winning bid number.

3.2 The Random Key Encoding

The initial solution is generated at random by using the strategy of the Random Key encoding (RK) introduced by Bean [2] and used mainly for ordering and scheduling problems.

The RK encoding mechanism permits to generate and manipulate only feasible solutions.

The Random Key encoding operates as follows:

- Firstly, we generate n real numbers sequenced by an r order, where n is the number of bids and the r order is a permutation of keys values.
- Secondly, we select the bid having the highest order value to include in the allocation, knowing that we start with an empty allocation. Then, the bid having the second-highest order value is accepted if its acceptance with accepted bid currently in the allocation verifies the constraint that means that for any good i , the sum of units of i over all the winning bids in the current allocation does not exceed mu_i , otherwise it is discarded. This process continues until having examined the n bids. We obtain a subset of bids that can be a feasible solution to the MU-WDP.

3.3 The Objective Function

The quality of a solution S is evaluated by calculating its objective function. The objective function of an allocation is equal to the sum of prices of the winning bids $S = \{B_1, B_2, \dots, B_l\}$.

$$F(S) = \sum_{i=1}^l Price(B_i) \quad (1)$$

where l is the number of element of the allocation S .

3.4 Local Search Method

The local search method is an iterative improvement method that consists of two essential components: finding a neighbour solution and a exploring the neighbourhood solutions to find good ones [16].

More precisely, it consists of the following steps:

- *Step 1*: start with a random initial configuration or an arbitrary solution S_0 .
- *Step 2*: consider this configuration as a solution with the problem $S = S_0$, measure the quality of S by $F(S)$ where F is the objective function to be optimized.
- *Step 3*: choose a neighbour solution S' of S such as $F(S') > F(S)$, replace S by S' .
- *Step 4*: Repeat *Step 3* until for all neighbour S' of S , $F(S') \leq F(S)$.

3.5 Tabu Search

Tabu search is a local search metaheuristic that has been proposed initially by Glover [10]. The proposed TS for the winner determination problem starts from an initial allocation created using the Random Key RK method. Then, it tries to locate for the best allocation by generating neighbour allocation. This is done by applying the three steps:

1. The first step consists of generating a neighbour allocation from the current allocation. A bid O is chosen such as it does not belong to the tabu list and its price is the highest. When O is found, it will be included into the allocation, and added into the tabu list. After that, we update the allocation revenue and a vector some conflict allows controlling the conflicts.
2. The second step involves the conflict elimination. To do that, the bid O_m having the minimal price is found and deleted from the allocation. When a bid is deleted from the solution, its price and goods quantities requested must be omitted. This procedure is repeated while the conflict persists.
3. The third step consists of saving the best allocation. For each neighbour allocation, the revenue is evaluated, so if it is more than the revenue of the best allocation then the neighbour allocation becomes the best allocation S^* .

The search process is repeated for a number of iterations fixed empirically. At the end, we obtain the best solution that gives the optimal revenue to the sellers.

The Tabu search outline for the WDP The TS algorithm is sketched in Algorithm 1.

Algorithm 1 : The Tabu search for the MU-WDP.

Require: an MU-WDP instance, $maxiter$, a collection S ,

Ensure: An improved allocation S

```

1:  $TL$  is the tabu list, initially empty:  $TL = \phi$ ;
2: Generate an arbitrary feasible allocation  $S$  using RK encoding;
3: for  $iter = 0$  to  $maxiter$  do
4:   Choose  $O$  from reparation bids having the maximal price and  $O \cap TL = \phi$ ;
5:    $S = S \cup O$ ;
6:   Update (Revenue and somConflit);
7:    $TL = TL \cup O$ ;
8:   while there is conflict do
9:      $O_m$  in  $S$  having the minimal price;
10:     $S = S - O_m$ ;
11:    Update (Revenue and somConflit);
12:  end while
13:  if  $F(S) > F(S^*)$  then
14:     $S^* = S$ ;
15:  end if
16: end for
17: return the best solution  $S^*$ .
```

3.6 A SLS for MU-WDP

The Stochastic Local Search (SLS) [17, 18] is a local search method that performs a certain number of local steps that combines diversification and intensification strategies to locate a good solution. The intensification phase consists in selecting a best neighbour solution. The diversification phase consists in selecting a random neighbour solution. The diversification phase is applied with a fixed probability $wp > 0$ whereas the intensification phase with a probability $1 - wp$. The SLS has been applied successfully on several optimization problem such as the single-unit winner determination problem in combinatorial auctions [3]. In this section, we propose a stochastic local search variant for the multi-unit winner determination problem (MU-WDP).

The proposed SLS method starts with an initial allocation S generated randomly according to the RK encoding. Then, it performs a certain number of local steps that consists in selecting a O to be added in the current allocation S and in removing all conflicting bids that occur in the current allocation.

At each step, the bid to be accepted is selected according to one of the two following criteria:

1. The first criterion (*step1* of Algorithm 2) consists of choosing the bid in a random way with a fixed probability $wp > 0$.
2. The second criterion (*step2*) consists of choosing the best bid (the one maximizing the auctioneer's revenue when it is selected) to be accepted.

The process is repeated for a certain number of iterations called *maxiter*, that was determined empirically.

The SLS algorithm for Multi-unit WDP The SLS algorithm is sketched in Algorithm 2.

Algorithm 2 : The SLS method.

Require: a WDP instance, *maxiter*, *wp*

Ensure: an allocation *S*

```

1: Generate an initial solution according to RK S
2: for I = 1 to maxiter do
3:   r  $\leftarrow$  random number between 0 and 1;
4:   if r < wp then
5:     O = pick a random bid (*Step 1)
6:   else
7:     O = pick a bid having the maximum price; (*Step 2)
8:   end if
9:   S = S  $\cup$  O;
10:  Update (Revenue and somConflit);
11:  while there is conflict do
12:    Om = the bid having the minimum price;
13:    S = S - Om;
14:    Update (Revenue and somConflit);
15:  end while
16:  if ( $F(S) > F(S^*)$ ) then
17:     $S^* = S$ ;
18:  end if
19: end for
20: return the best allocation S*.

```

4 The Experimental Study

In order to evaluate the performance of the three proposed approaches LS, TS and SLS for solving the MU-WDP, we have implemented them on machine and evaluated them on some benchmarks of the problem.

The C++ programming language is used to implement the proposed approaches for the MU-WDP. Programs are run on Laptop hp Compaq 610, with CPU core 2 duo 2 GHz, 2 GB of memory.

4.1 The Dataset

The dataset used in this study is randomly generated. Ten benchmarks of various sizes have been generated.

Table 1 gives information about the considered benchmarks. The column one gives the number of benchmark and the column two the number of bids. The column three is the number of goods, the column four gives the interval values of the number of goods per bids, the column five is the interval values of the profit of bids, and the column six represents the interval of values of the available quantities of goods. The benchmarks were generated so as to increase the complexity of the problem by creating much competition between the variables of the problem (the bids). We notice that the prices of bids are very brought closer as well as the quantity available of each good.

4.2 The Parameters Tuning

The adjustment of the different parameters of the proposed approaches SLS, TS and LS is fixed by an experimental study.

- The TS parameters are: the maximum number of iterations is fixed to 5,000 iterations. This number of iteration is sufficient and allows the process to find the best solution. The size of the tabu list is set to 40 elements.
- The SLS parameters are: the maximum number of iterations is fixed to 50,000 iterations. The value of probability $wp = 0.87$.
- The LS parameters are: the searching process is stopped according to the increasing of the objective function F, when F is not increasing any more and no bid remains then the process stops.

4.3 The Numerical Results

Table 2 gives the results obtained by the stochastic local search (SLS) algorithm, the local search (LS) and tabu search (TS) for the considered MU-WDP benchmarks. For each method, we give the CPU time (Time in *second*) and the revenue of the best solution found by each approach (Rev is in *Monetary unit*).

In order to show the performance of the SLS approach, we conducted a comparison with the CPLEX solver 12.5 [20]. We note that for each benchmark, the CPLEX is launched for fifteen minutes (15 min) maximum.

Table 2 gives the average results obtained after five runs of the algorithms where *Rev* is the average revenue and *Time* is the average CPU consumed in second.

Table 1 The data set description

Benchmark	Number of bids	Number of goods	Goods/ Bids	Price of bid	Available quantity
B1	1,400	1,500	[0, 110]	[322, 325]	[900, 920]
B2	1,300	900	[0, 200]	[600, 605]	[1,200, 1,230]
B3	800	800	[0, 400]	[400, 420]	[800, 830]
B4	900	900	[0, 250]	[200, 210]	[1,000, 1,100]
B5	700	1,000	[0, 60]	[100, 105]	[200, 220]
B6	1,000	600	[0, 55]	[150, 155]	[100, 115]
B7	1,100	1,500	[0, 70]	[700, 710]	[200, 220]
B8	1,500	1,500	[0, 90]	[800, 820]	[600, 630]
B9	1,500	1,000	[0, 300]	[10,000, 10,500]	[3,000, 3,100]
B10	1,200	1,000	[0, 120]	[500, 510]	[900, 950]

Table 2 Numerical results of SLS, TS, LS and CPLEX

Benchmark	SLS		TS		LS		CPLEX
	Time (s)	Rev (Mu)	Time (s)	Rev (Mu)	Time (s)	Rev (Mu)	Rev (Mu)
B1	0.047	3,573.9	3.222	3,573.9	0.002	3,569.48	3,561.5
B2	1.233	4,839.2	2.265	4,821.14	0.001	4,833.58	4,832
B3	0.078	839.81	4.84	839.81	0.0001	839.12	839.7
B4	3.074	839.61	2.646	839.6	0.001	837.88	838.7
B5	0.936	419.59	2.374	419.4	0.0008	417.24	419.0
B6	0.514	309.8	2.73	309.7	0.0004	308.76	309.8
B7	1.108	2,129.7	2.488	2,129.7	0.0001	2,128.78	2,128.4
B8	1.2	7,378.4	3.877	7,303.85	0.0032	7,035.6	7,351.4
B9	0.062	6,053.9	3.568	5,789.575	0.0004	5,047.2	6,042.7
B10	2.075	5,600.2	3.892	5,098.7	0.0024	5,283.32	5,597.5

s second and *Mu* Monetary unit

According to the obtained results we released the following remarks:

- For all the benchmarks we noticed that the approach SLS is always able to give the same results of revenue, even if we make several runs. The TS always gives the same results of revenue for the benchmarks 1, 3, 4 and 7, whereas LS finds various results for each runs for all benchmarks.
- According to the numerical results, SLS succeeds to find good quality solutions for the all checked benchmarks. For benchmarks 1 and 3 the quality solutions found by SLS is similar to those found by TS. We remark that, the Benchmark 6 is not very complex since CPLEX gave the same revenue as SLS. However, the simple local search LS fails to obtain good results for all the benchmarks compared to both SLS and TS.
- In term of CPU time point of view, LS is faster than both SLS and TS, but the disadvantage is that it falls quickly into the local minima.

5 Conclusion

In this paper, we proposed three local search methods for the multi-unit winner determination problem in combinatorial auctions. The three proposed approaches stochastic local search (SLS), tabu search (TS) and Local Search (LS) have been implemented and tested on some benchmarks generated randomly. Also, we have conducted a comparison with the CPLEX solver 12.5 in order to show the performance of the proposed approach. According to the experimental study, the SLS was shown great performance and effectiveness in solving MU-WDP. The results are very encouraging, and prove that SLS is able to maximize the objective function better and faster than TS and LS. The comparative study with CPLEX confirms our results. We plan to validate the proposed approach on other benchmarks for MU-WDP.

References

1. Anderson, A., Tenhunen, M., Ygge, F.: Integer programming for combinatorial auction winner determination. In: Proceedings of 4th International Conference on Multi-Agent Systems, pp. 39–46. IEEE Computer Society Press (2000)
2. Bean, J.C.: Genetics and random keys for sequencing and optimization. *ORSA J. Comput.* **6**(2), 154–160 (1994)
3. Boughaci, D.: Metaheuristic approaches for the winner determination problem in combinatorial auction. In: Artificial Intelligence, Evolutionary Computing and Metaheuristics, pp. 775–791. Springer, Berlin (2013)
4. Boughaci, D., Benhamou, B., Drias, H.: Local search methods for the optimal winner determination problem in combinatorial auctions. *J. Math. Model. Algorithms* **9**, 165–180 (2010)
5. Boughaci, D., Benhamou, B., Drias, H.: A memetic algorithm for the optimal winner determination problem. *Soft. Comput.* **13**, 905–917 (2009)
6. Caplice, C.G.: An optimization based bidding process: a new framework for shipper carrier relationships. Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA (1996)
7. Chen, L., Hu, S., Chen, X., Lin, Y.: Hybrid algorithm for winner determination problem in combinatorial auctions. In: International Conference on Computational Intelligence and Security, pp. 82–86 (2008)
8. Farzi, S.: Discrete quantum-behaved particle swarm optimization for the multi-unit combinatorial auction determination problem. *J. Appl. Sci.* **10**(4), 291–297 (2010)
9. Fujishima, Y., Leyton-Brown, K., Shoham, Y.: Taming the computational complexity of combinatorial auctions: Optimal and approximate approaches. In: Proceedings of IJCAI–99 (1999)
10. Glover, F.: Future paths for integer programming and links to artificial intelligence. *Comput. Oper. Res.* **13**, 533–549 (1986)
11. Gonen, R., Lehmann, D.: Linear programming helps solving large multi-unit combinatorial auctions (2001)
12. Guo, Y., Lim, A., Rodrigues, B., Zhu, Y.: Heuristics for a brokering set packing problem. In: Proceedings of Eighth International Symposium on Artificial Intelligence and Mathematics, pp. 10–14 (2004)

13. Guo, Y., Lim, A., Rodrigues, B., Zhu, Y.: Heuristics for a bidding problem. *Comput. Oper. Res.* **33**(8), 2179–2188 (2006)
14. Guo, Y., Lim, A., Rodrigues, B., Tang, J.: Using a Lagrangian heuristic for a combinatorial auction problem. *Int. J. Artif. Intell. Tools La-grange-IJAIT*, 21–34 (2006)
15. Graves, R.L., Schrage, L., Sankaran, J.K.: An auction method for course registration. *Interfaces* **23**(5), 81–92 (1993)
16. Hansen, P., Mladenovic, N.: An introduction to variable neighborhood search. In: *Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization*, pp. 433–458. Kluwer Academic Publishers, Boston (1999)
17. Hoos, H.H., Craig, B.: Solving combinatorial auctions using stochastic local search. In: *Proceedings of the 17th National Conference on Artificial Intelligence*, pp. 22–29 (2000)
18. Hoos, H.H., Stutzle, T.: *Stochastic local search: foundations and applications*. Morgan Kaufmann Publishers, Burlington (2004)
19. Holland, A., Osullivan, B.: Towards fast vickrey pricing using constraint programming. *Artif. Intell.* **21**(34), 335–352 (2004)
20. ILOG CPLEX optimization studio 12.5
21. Jackson, C.: *Technology for spectrum markets*. Ph.D. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA (1976)
22. Jones, J.L.: *Incompletely specified combinatorial auction: an alternative allocation mechanism for business-to-business negotiations*. Ph.D. thesis, Warrington College of Business Administration, University of Florida, Gainesville, FL, USA (2000)
23. Kameshwaran, S., Benyoucef, L.: Branch on price: a fast winner determination algorithm for discount auctions. In: Cheng, S.-W., Poon C.K. (eds.) *AAIM 2006, LNCS 4041*, pp. 375–386. Springer, Berlin (2006)
24. Laureano, F.: Escudero, Mercedes Landete, Alfredo Marn: a branch-and-cut algorithm for the winner determination problem. *Decis. Support Syst.* **46**, 649–659 (2009)
25. Leyton Brown, K., Shoham, Y., Tennenholtz, M.: An algorithm for multi-unit combinatorial auctions. In: *Proceedings of the National Conference on Artificial Intelligence (AAAI-00)*, pp. 56–61. Austin, Texas, US (2000)
26. Nisan, N.: Bidding and allocation in combinatorial auctions. In: *Proceedings of the ACM Conference on Electronic Commerce (EC-00)*, pp. 1–12. Minneapolis, US, Oct 2000
27. Rassenti, S.J., Smith, V.L., Bulfin, R.L.: A combinatorial auction mechanism for airport time slot allocation. *Bell J. Econ.* **13**(2), 402–417 (1982)
28. Rothkopf, M., Pekee, A., Harstad, R.: Computationally manageable combinatorial auctions. *Manage. Sci.* **44**(8), 1131–1147 (1998)
29. Sheffi, Y.: Combinatorial auctions in the procurement of transportation services. *Interfaces* **34**(4), 245–252 (2004)
30. Singh, R.J., Sen, A.K., Sarkar, U.K.: A partitioned stochastic search algorithm: application to multi-unit winner determination problem in combinatorial auction. In: *Proceedings of Athens: ATINER'S Conference Paper Series, No: COM2012–0261* (2012)
31. Sandholm, T., Suri, S.: BOB: improved winner determination in combinatorial auctions and generalizations. *Artif. Intell.* **145**, 33–58 (2003)
32. Sandholm, T., Suri, S., Gilpin, A., Levine, D.: CABOB: a fast optimal algorithm for winner determination in combinatorial auctions. *Manage. Sci.* **51**(3), 374–390 (2005)
33. Zheng, G., Lin, Z.C.: A Winner determination algorithm for combinatorial auctions based on hybrid artificial fish swarm algorithm. *Physics Procedia*, 1666–1670, International Conference on Solid State Devices and Materials Science (2012)

Strongly Biased Crossover Operator for Subgraph Selection in Edge-Set Based Graph Representation

Sakshi Arora and M. L. Garg

Abstract The performance of crossover operator is a complex interplay of the various characteristics of the genetic algorithm (GA) and sometimes also of the problem under question. The fundamental design choices in a GA therefore, are its representation of candidate solutions and the operators that will act on that representation. This paper suggests a new strongly biased multiparent crossover operator that offers a strong locality and provides a strong heritability quotient but can still escape the local optima when the sample space is over represented by similar building blocks. The proposed operator uses a dynamic 2-D vector representation for the chromosomes and this data structure may evolve as the execution of the crossover operator proceeds. On a population which consists of solutions near to optimal and mostly lying in the basin of attraction of single optima, the bias towards the optima in the generated offsprings is proportional to the sample size from the search space. Based on this reasoning, the effect of arity of the proposed crossover operator is tested using a population in which all the candidates lie in close neighborhood of each other. To analyze the dynamic search behavior of the proposed crossover operator and the impact of the representation scheme on the locality, heritability and exploratory power of the operator, we suggest a no mutation, zero selection pressure GA model that generates bounded diameter minimum spanning trees from the underlying complete graphs on random and Euclidean instances.

Keywords Genetic algorithm · Arity of operator · Locality · Heritability

S. Arora (✉) · M. L. Garg
School of Computer Science and Engineering, S.M.V.D.U, Katra, India
e-mail: sakshi@smvdu.ac.in

M. L. Garg
e-mail: garg.ml@smvdu.ac.in

1 Introduction

An evolutionary algorithm's structure and parameter settings affect its performance, but the primary determinants of an EA's success or failure are the coding by which its chromosomes represent candidate solutions and the interaction of the coding with the EA's operators, specifically the crossover operator. The performance of crossover operator is a complex interplay of the various characteristics of the genetic algorithm and sometimes also of the problem under question. The fundamental design choices in an evolutionary algorithm therefore, are its representation of candidate solutions and the operators that will act on that representation. These operators may offer varying degrees of locality, heritability, and computational efficiency.

Researchers have described EAs for a number of constraint and simple MST-related problems, particularly the bounded diameter minimum spanning tree (BDMST) problem [1, 2]. Most often, these algorithms have used implicit representations to encode spanning trees and have applied decoding algorithms to identify the spanning trees that chromosomes represent [3]. They have applied positional operators like k-point crossover [4]. An EA's recombination operator should provide strong heritability. Here, this means that the tree produced by recombination of two parent trees should consist mostly of parental edges. We propose representing spanning trees for evolutionary search as Real numbers representing the index of the edge. The reason for choosing this form of edge set encoding is that evaluating edge-sets is fast, as are the operators, particularly recombination operator; under these operators, edge-sets have strong heritability and locality. Applied to this representation, traditional initialization algorithms and positional crossover and mutation operators are unlikely to generate chromosomes that form spanning trees [5], so we describe a new strongly biased multiparent crossover operator k-GCO () that offers a strong locality and provides a strong heritability quotient and can overcome the drawbacks of some of the overly biased crossover operators mentioned in the literature [6]. Further, the traditional crossover operators like the uniform crossover operator with $p = 0.04$ and the 10-point crossover operator, fail to achieve the desired results because the number of bits in the string make it difficult to achieve the desired results with these operators. A better crossover would focus on the positions in which the strings differ, and select exactly k bits to exchange. Crossover is a mechanism of restricting the search space to Hamming closure of two strings [7]. It is reasonable therefore to design crossover operators that focus on this space. The K-Guided Crossover Operator (K-GCO()) operator attempts to increase the scope of the search space to Hamming closure of k strings, where the k strings are all high in fitness, where k is a user determined, problem dependent parameter which denotes the number of parents participating in recombination. The simulation results show that the proposed strategy significantly improves the computational efficiency of GAs. The rest of the paper is organized as follows. In the section that follows, a brief overview of the K-GCO () and the work going on in the field of multiparent

crossover operators using different representation schemes for achieving strong locality and high heritability quotient is provided. Section 3 discusses the different phases of K-GCO () in light of locality and heritability. Experimental results are presented in Sect. 4. Section 5 summarizes the main contributions of the paper.

2 Related Work

2.1 Chromosome Representation Schemes

The significance of crossover operator in controlling GA performance has long been acknowledged in GA research [8]. Evolutionary algorithms have proven effective on several hard spanning tree problems, and much work has been done on representing spanning trees for evolutionary search [9]. Recent studies have indicated the general usefulness of representing spanning trees directly as lists of their edges and applying operators that always yield feasible trees [10]. We adopt the edge-list coding here, but augment each chromosome with the center vertex or vertices of the spanning tree it represents. The fitness of a chromosome is the total weight of its tree, which can be found, by scanning the edge-list, in time that is $O(n)$. The diameter bound renders the initialization and variation operators of the prior EAs inapplicable or ineffective. EAs for spanning tree problems that represent candidate spanning trees directly as sets of their edges have been considered in [9]. Further, in [11] crossover and mutation operators based on edges, paths, and subtree of spanning trees in an EA for the optimum communication spanning tree problem have been suggested. We propose representing chromosomes of spanning trees as real numbers representing the index of edges. This representation can be implemented in an array or a hash table whose entries are the indices of edges from which the pairs of vertices can be extracted using the formula given in Sect. 3.3. The latter data structure allows insertion, lookup, and deletion of individual edges in constant time; both require space that is linear in the number n of nodes. The edge-set based crossover operator-the heuristic crossover operator presented in [6] is a modified version of KruskalRST* crossover. [12] Proposed two other variants of the edge-set based crossover operators. They differ in the strategy of completing the offspring with the edges available in E_0 : **Greedy crossover**: When using this strategy, the edge with the smallest weight is chosen from E_0 . **Inverse-weight-proportional crossover**: This strategy selects each edge from E_0 according to probabilities inversely proportional to the edges' weights. Again, Julstrom and Raidl [3] has examined the performance of the different crossover variants. Due to the construction process all three crossover strategies have a strong bias towards the MST. The bias of the TX operator is already so strong that EAs are only able to find optimal solutions if they are very similar to the MST [6]. According to the construction process the bias of the Greedy crossover is higher than the bias of the TX crossover. Therefore, Greedy crossover also results in low EA performance if

the optimal solution is not the MST. The inverse-weight-proportional crossover introduces a bias to the MST similar to the TX crossover. However, the bias cannot be controlled in a systematic way but depends on the specific weights of the edges.

2.2 Arity of Crossover Operators

Not much work has been done in the field of the arity of the crossover operator. Where, the arity of crossover operator is defined as the number of parents it uses. We have come across studies where the multi-parent aspect has an incidental character, for instance, a certain fixed arity, or even if the definition is general and would allow comparison between different number of parents, this aspect is not given attention. ASPARAGOS [13], applies p-sexual voting recombination applied on the quadratic assignment problem. An interesting attempt has been made to combine genetic algorithms with the Simplex Method resulted in the ternary simplex crossover in [14]. Uniform crossover with an arbitrary number of parents is also used in [15] as schema sampling procedure in a GA, but the multi-parent feature is only a side-effect and is not investigated. r-crossover has been introduced as a generalization of uniform crossover in GAs creating one child from r parents in [16]. The results of (CMX) *operator* (MFX), and *seed crossover operator* (SX) [17] were better on functions having multimodality and/or epistasis. Thus these studies concluded clearly indicate that on multimodal landscapes, an arity of more than two for the recombination operator may benefit the search.

BDMSTs have a fitness landscape known as rugged or multi modal landscape. We therefore, suggest a new heuristic multiparent recombination operator on the edge set encoding scheme of the graphs, applicable to the Bounded Diameter minimum spanning tree generation. The suggested operator has a strong heritability quotient and is less biased than greedy and TX operators and is more biased than KruskalRST*. Along with the encoding scheme, unbiased initialization and tunable arity, the proposed crossover operator is proven to give better performance on Genetic algorithm. The operator is easily extended to generate only feasible spanning trees and to incorporate local, problem-specific heuristics.

In our operator, we make sexuality a graded rather than a Boolean feature by having an arity that can vary. In other words, the operator can be called with a certain number of parents as input and this number can be modified by the user.

3 K-Guided Crossover Operator: (K-GCO ())

Input: 'n': number of nodes in the sample graph, 'k': number of parents, 'd': Diameter constraint

Output: $2k$ number of Offsprings

Phase 1: *K_Parents_Selection*()**Phase 2: *Gene Frequency based Scanning*()**Stage 1: *Tabular_Chromosome_Declaration*()Stage 2: *Initial_Setup*()Stage 3: *Centre_Extraction_Routine*()Stage 4: *Gene_Occurrence_Scanner*()Stage 5: *Centre_Vertex_Incidence_Checker*()**Phase 3: *Diameter_Constrain_handler*()****Phase 4: *Candidate_Tree_Construction*()**Stage 1: *Declaring_Disconnected_Component_handler*()Stage 2: *Call_subroutine_Handling_disconnected_component*()Stage 3: *Call_subroutineVertex_Checker*()Stage 4: *Random_Subgraphs_for_Offsprings*()**3.1 Phase 1: *K_Parents_Selection*()**

This routine selects K parents from the initial population pool whereby each parent is selected via 2-Tournament Selection. 2-Tournament Selection besides roulette wheel selection is one of the most popular parent selection mechanism adopted. The selection pressure exerted by tournament selection method has some randomness and is not heavily biased towards the best individuals as in roulette wheel, and therefore fits better in our model. The increased arity of the operator as explained will introduce the requisite amount of bias in favor of the parent candidates.

3.2 Phase 2: *Gene_Frequency_Based_Scanning*()

This phase generates the tabular chromosome. The proposed chromosome structure differs from the tabular chromosome given by Knowles and Corne [18]. The proposed structure as can be seen in the ensuing phases of *K-GCO*(), is a dynamic 2-D vector which evolves as the as the execution of the crossover operator proceeds.

The sub stages ***Tabular_Chromosome_Declaration*()** and ***Initial_Setup*()** take care of the generation and setting up of the chromosome named as **Parent_vector**, where the first row of the vector represents the edge indices of the edges appearing in the chromosome, the second row of the vector represents the number of occurrences of each edge in the parents selected and is initialized to all ones at the beginning of the GCO run (Fig. 1).

Next, the ***Centre_Extraction_Routine*()** extracts the root nodes of the parent trees from a universal array **U_array** which stores the centers of all the trees in the

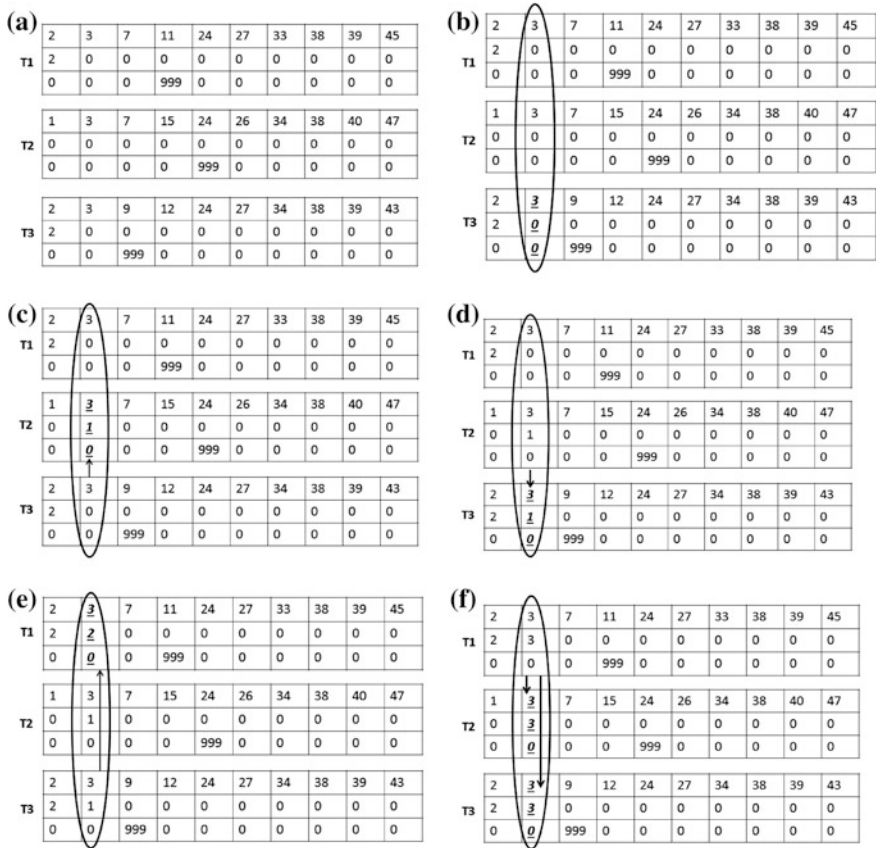


Fig. 1 a–f Gene frequency scanning

initial population. From these parent roots, it randomly selects one of the roots to be the centre node of the offsprings.

The **Gene_Occurrence_Scanner()** subroutine calculates the frequency of occurrence of genes in the parents. As the scanning phase of the GCO proceeds, this row reflects the status of the edges as being present in one, two or k parents respectively. The array *Position_Tracker* keeps a track of the locus in ith parent where a particular gene is present and on each occurrence of that gene in k parents the counter is incremented. Based on the locus reflected in the *Position_Tracker* for a particular parent, the counter is then updated and reflected as the frequency of that gene. The third row **centre_edge** of the vector is a flag initialized to all zeros.

The subroutine **Centre_Vertex_Incidence_Checker()** checks each edge for incidence on centre and updates the third row accordingly. This row will assume significance in the next phase of the GCO i.e. segregation of edges into different sets. Here, the length of the vector is n–1, where n represents the number of vertices in graph.

3.3 Phase 3: Diameter Constraint Handler()

This subroutine ensures the generation of valid offsprings only i.e. offsprings that respect the diameter constraint. Most of the algorithms in literature [19] generate constraint MST's which violate one or the other constraint necessitating the need for applying computationally expensive repair methods or rejecting such MST's altogether. With each of the $2k$ offsprings, a **vertex_depth** vector which is initialized to -1 is associated, which keeps the track of the depth of each vertex from the centre and allows only those edges to be inserted in the offspring which do not violate the diameter constraint. The genes in the chromosomes representing the parent spanning trees are real numbers representing the index of the edge. The vertices are extracted from these edges using the following formula:

$$vertex1 = (index/n) + 1$$

and $vertex2 = (index\%n)$ where n is the number of vertices in the graph.

The depth of the centre in the **vertex_depth** vector is 0. The vertices which are directly connected to the centre will have a value of 1. Vertices not forming direct edge with centre are updated as the execution of this phase proceeds.

3.4 Phase 4: Candidate_Tree_Construction()

The flag **centre_edge** associated with each edge of the parent vector, assumes a value 1 or 0 indicating whether the edge is incident on centre or not resp. The construction of offsprings starts with the extraction of those edges from the k Parent vectors whose flag value is equal to 1.

Thus partially constructed offspring may further witness one of the three cases:

Case 1: When either vertex 1 or vertex 2 but not both is already present in the partially constructed offspring. If one of the vertexes is in constraint handling vector and the other in disconnected components vector then all the vertices having same status would get copied after calculating their depth from the **vertex_depth** vector. The corresponding edge index is stored in the partially constructed offspring vectors $1 \dots k$.

Case 2: When both the vertices are not present in the partially constructed offspring. The sub routine **Handling_Disconnected_Components()** is called **Disconnected_Component_Handler** is a dynamic linked list which stores those pair of extracted vertices which are not yet part of the partially constructed offspring. The variable, **status**, associated with the pair of vertices is assigned some rank. Every time the pair of vertices is not present in the partially constructed offspring and the **Disconnected_Component_Handler** lists, a new incremented rank is assigned. However, if one of the two vertices is present in the list, then same existing rank is assigned. If both the vertices are present in the linked list, having

separate status, then the **Disconnected_Compont_Handler** is updated with the smaller rank value/status.

Case 3: When both the extracted vertices are already present in the offspring. Such vertices are rejected as including the edge between them would violate the constraint of cycle free graph.

The routine **Random_Sungraphs_for_Offsprings()** is then called which tracks from the vertex depth array those vertices whose depth is still -1 and their inclusion does not violate the diameter constraint when paired with a randomly selected vertex from the offspring vector.

4 Experimental Setup

The iterated interplay of selection, variation operators, and replacement strategy in the EA exhibits its own dynamics. In evolutionary search, the behavior of the variation operators is expected to depend on the current population's properties. Therefore, this section studies the complete *dynamic* case and states the results obtained by applying K-GCO() on two different strategies proposed by the authors earlier in [20, 21]. The effect of K-GCO() in searching the multimodal landscape can be studied in two different cases, when the initial population pool consists of strings very close to the optimum and when strings in the initial pool may be present on 2 to 3 different peaks is studied. To study the dynamic search behavior of K-GCO() we propose a zero mutation and no selection pressure GA model so that the characteristics of the offsprings reflect purely the search behavior of K-GCO(). It may be added though that adding ARC_EX_Mut() mutation operator as suggested by authors in [22] may enhance the quality of generated BDMSTs further. In the Intelligent Genetic Algorithm for Tree Construction (I-GAT), the first case is ensured by generating the initial population using CTH algorithm [21]. G-GAT- the Guided Genetic Algorithm for Tree Construction ensures the second case when the initial population strings are generated using DRGH [20]. Both these algorithms ensure a priori knowledge of the optimum, so that the bias of K-GCO() can be constructively utilized to help I-GAT and G-GAT to converge faster.

Tables 1 and 2 illustrates the comparison of results of G-GAT and I-GAT with those of ACO implementation [23], permutation coded EA as well as a Edge-stet coded EA from [5]. Of major interest here ACO as ACO is so far the leading metaheuristic for BDMST. For comparison the same instances, diameter bounds, time limits and termination conditions as in [23] are used. The instances are taken from the Beasley's OR-Library¹ and have been originally proposed for the Euclidean Steiner tree problem. All nodes of a graph are settled within the unit square and the Euclidean distance between any pair of nodes corresponds to the edge costs between these two nodes. Furthermore, as in [23], the tests were grouped into long-term and short-term runs. The long-term runs concentrated on

Table 1 Long term runs on euclidean instances

Instance			Permutation coded EA			ACO		G-GAT		I-GAT	
n	D	nr.	Best	Stddev	P	Best	Stddev	Best	Stddev	Best	Stddev
100	10	1	7.818	0.07	0.01	7.759	0.007	7.751	0.004	7.732	0.03
		2	7.873	0.08		7.849	0.004	7.836	0.005	7.827	0.05
		3	7.99	0.08		7.907	0.014	7.905	0.008	7.902	0.07
		4	8.009	0.07		7.979	0.005	7.976	0.004	7.979	0.06
		5	8.193	0.08		8.164	0.002	8.159	0.001	8.157	0.011
250	15	1	12.44	0.08	0.05	12.245	0.013	12.242	0.009	12.24	0.011
		2	12.237	0.1		12.023	0.015	12.019	0.009	12.02	0.08
		3	12.117	0.08		12.017	0.009	12.015	0.011	12.01	0.007
		4	12.572	0.11		12.474	0.011	12.413	0.08	12.4	0.009
		5	12.358	0.12		12.228	0.016	12.223	0.09	12.21	0.002
500	20	1	17.216	0.1	0.1	16.909	0.018	16.813	0.13	16.78	0.011
		2	17.085	0.11		16.76	0.019	16.67	0.12	16.59	0.11
		3	17.173	0.11		16.929	0.014	16.903	0.07	16.83	0.05
		4	17.215	0.13		16.908	0.026	16.835	0.01	16.8	0.03
		5	16.939	0.11		16.512	0.022	16.503	0.01	16.48	0.03

achieving as good results as possible. In the short-term runs the different algorithms had a relatively short time with respect to the instance size for yielding good solutions. The size of the ant colony was set to ten artificial ants for all test categories. And in case of G-GAT and I-GAT crossover probability of 1 has been used. All test runs have been performed on a PentiumR4 2.8 GHz system.

As in [24] the first five instances of each size $n = 100, 250, 500$ available in the OR-Library were used and the diameter bound was set to 10 for instances with 100 nodes, to 15 for instances with 250 nodes, and to 20 for instances with 500 nodes. G-GAT and I-GAT algorithms terminated either if the last 1,000 iterations yielded no improvement of the best found solution so far or if a time limit was reached. This time limit was set to 2,000 s for instances with 100 nodes, to 3,000 s for instances with 250 nodes, and to 4,000 s for instances with 500 nodes.

Table 1 provides the best value of the BDMST for the four metaheuristics and the standard deviation in these values over 50 runs. The bold highlights in Table 1 show the least weight BDMSTs obtained using these techniques for different values of n and D . It is clear from Table 1 that I-GAT out performs the G-GAT and the other two heuristics for all the values of n and D . However, the cost of initial population generation in case of I-GAT is high and scales proportional to the increase in the value of n . The initial population generation in case of G-GAT is also time consuming, however for larger instances, despite the time utilized in the initial population generation, G-GAT is effective in reaching the optimal solution in comparison to ACO and permutation coded EA. The ability of K-GCO() to work with solutions lying in the basin of attraction of optimum facilitates convergence to the optimum.

Table 2 Short-term runs on euclidean instances

Instance			Time (s)	Edge-set coded EA			ACO		G-GAT		I-GAT	
n	D	nr.		Best	Stddev	ρ	Best	Stddev	Best	Stddev	Best	Stddev
500	20	1	50	19.368	0.17	0.5	17.905	0.035	17.887	0.027	17.993	0.056
		2		19.156	0.13		17.832	0.037	17.802	0.025	17.839	0.047
		3		19.321	0.16		17.99	0.027	17.89	0.028	17.92	0.051
		4		19.464	0.19		18.181	0.023	18.169	0.031	18.21	0.049
		5		19.209	0.17		17.66	0.047	17.58	0.029	17.63	0.045
500	20	1	500	18.47	0.13	0.16	17.347	0.019	17.339	0.032	17.349	0.046
		2		18.442	0.22		17.111	0.027	17.097	0.033	17.009	0.046
		3		18.619	0.18		17.27	0.033	17.22	0.029	17.25	0.051
		4		18.745	0.17		17.294	0.029	17.289	0.028	17.299	0.049
		5		18.197	0.02		17.031	0.029	17.025	0.031	17.027	0.048
1,000	25	1	100	28.721	0.16	-	-	-	25.743	0.032	25.993	0.047
		2		28.607	0.19		-	-	26.347	0.029	26.735	0.043
		3		28.41	0.17		-	-	26.578	0.027	26.987	0.045
		4		28.695	0.21		-	-	26.295	0.025	27.009	0.047
		5		28.396	0.19		-	-	26.831	0.029	26.994	0.049
1,000	25	1	1,000	26.494	0.14	0.4	25.414	0.034	25.398	0.031	25.426	0.043
		2		26.3	0.24		25.281	0.028	25.269	0.033	25.277	0.045
		3		25.762	0.21		24.895	0.051	24.803	0.029	24.835	0.046
		4		26.47	0.15		25.391	0.038	25.332	0.028	25.389	0.043
		5		26.117	0.11		24.975	0.039	24.893	0.029	25.013	0.042

From Table 2 it can be seen that the G-GAT consistently outperforms the other metaheuristics. Considering the 500 node instances, the best values of the G-GAT with tighter time limits, as well as those of the ACO, are even better than the objective values of the best solutions found by any of the EAs. It can be inferred from the results of Table 2 that G-GAT outperforms ACO and permutation encoded as well as edge-set encoded EAs for all test instances under consideration. G-GAT has provided better coverage of the search space and has been found to be successful in providing solutions of better quality in comparison to ACO and edge-set encoded and permutation encoded EAs.

5 Conclusion

The usefulness of Edge-sets in encoding spanning trees and applying operators that always yield feasible trees has been proven in recent studied. Indexed edge encoding has therefore been proposed in this study. Further, a dynamic 2-D structure has been proposed for chromosome representation under the above encoding scheme, which effectively yields only feasible offsprings. The new strongly biased multiparent crossover operator k-GCO() has been proposed that

attempts to increase the search of the recombination to Hamming closure of more than 2 strings. The results indicate that on Euclidean instances I-GAT and G-GAT and consequently K-GCO() provides a strong heritability. The a priori knowledge of the optimum guide the bias of K-GCO towards finding better quality solutions as compared to Permutation encoded and edge-set encoded EAs and ACO on Euclidean instances for generating BDMSTs.

Future research could focus on testing the strength of bias of this operator and its effectiveness on random instances and it can be adapted to work on instances when the underlying graph is incomplete.

References

1. Raidl, G.R., Julstrom, B.A.: Edge sets: an effective evolutionary coding of spanning trees. *IEEE Trans. Evol. Comput.* **7**(3), 225–239 (2003)
2. Fekete, S.P., et al.: A network-flow technique for finding low-weight bounded-degree spanning trees. *J. Algorithms* **24.2**, 310–324 (1997)
3. Julstrom, B.A., Raidl, G.R.: A permutation-coded evolutionary algorithm for the bounded-diameter minimum spanning tree problem. In: *Workshop on Analysis and Design of Representations in 2003 Genetic and Evolutionary Computation Conference's Workshops Proceedings*, pp. 2–7 (2003)
4. Gen, M., Ida, K., Kim, J.: A spanning tree-based genetic algorithm for bicriteria topological network design. In: *Evolutionary Computation Proceedings, 1998. The 1998 IEEE International Conference on IEEE World Congress on Computational Intelligence*, pp. 15–20. IEEE (1998)
5. Rothlauf, F.: On the bias and performance of the edge-set encoding. *IEEE Trans. Evol. Comput.* **13**(3), 486–499 (2009)
6. Rothlauf, F., Tzschoppe, C.: Making the edge-set encoding fly by controlling the bias of its crossover operator, pp. 202–212. Springer, Berlin (2005)
7. Tsutsui, S.: Sampling bias and search space boundary extension in real coded genetic algorithms. In *GECCO*, pp. 211–218 (2000)
8. Srinivas, M., Patnaik, L.M.: Adaptive probabilities of crossover and mutation in genetic algorithms. *IEEE Trans. Syst. Man Cybernet.* **24**(4), 656–667 (1994)
9. Julstrom, B.A.: Encoding rectilinear Steiner trees as lists of edges. In: *Proceedings of the 2001 ACM Symposium on Applied Computing ACM*, pp. 356–360 (2001)
10. Li, Y.: An effective implementation of a direct spanning tree representation in GAs. In: *Applications of Evolutionary Computing*, pp. 11–19. Springer, Berlin (2001)
11. Julstrom, B.A., Raidl, G.R.: Weight-biased edge-crossover in evolutionary algorithms for two graph problems. In: *Proceedings of the 2001 ACM Symposium on Applied Computing*, pp. 321–326 (2001)
12. Kaufman, H.: An experimental investigation of process identification by competitive evolution. *IEEE Trans. Syst. Sci. Cybernet.* **3**(1), 11–16 (1967)
13. Mühlenbein, H.: Parallel genetic algorithms, population genetics and combinatorial optimization. In: *Parallelism, Learning, Evolution*, pp. 398–406. Springer, Berlin (1991)
14. Bersini, H., Seront, G.: In search of a good crossover between evolution and optimization. *Manner Manderick* **1503**, 479–488 (1992)
15. Aizawa, A.N.: Evolving SSE: a stochastic schemata exploiter. In: *Proceedings of the 1st IEEE Conference on Evolutionary Computation*, pp. 525–529. IEEE Press (1994)
16. Eiben, A.E., Raue, P.E., Ruttkay, Z.: Genetic algorithms with multi-parent recombination. In: *Parallel Problem Solving from Nature—PPSN III*, pp. 78–87. Springer, Berlin (1994)

17. Tsutsui, S.: Multi-parent recombination in genetic algorithms with search space boundary extension by mirroring. In: *Parallel Problem Solving from Nature—PPSN V*. Springer, Berlin (1998)
18. Knowles, J., Corne, D.: A new evolutionary approach to the degree-constrained minimum spanning tree problem. *IEEE Trans. Evol. Comput.* **4**(2), 125–134 (2000)
19. Deo, N., Abdalla, A.: Computing a diameter-constrained minimum spanning tree in parallel. In: *Algorithms and Complexity*, pp. 17–31. Springer, Berlin (2000)
20. Arora, S., Garg, M.L.: New hybrid evolutionary algorithm for solving the bounded diameter minimum spanning tree problem (2009)
21. Arora, S., Garg, M.L.: Clustering the data points to obtain optimum backbones for the bounded diameter minimum spanning trees. In: *2011 International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 303–307. IEEE (2011)
22. Arora, S., Garg, M.L.: Neighborhood search for the bounded diameter minimum spanning tree. *Int. J. Emerg. Technol. Adv. Eng.* **3**(2) (2013)
23. Kopinitsch, B.: An ant colony optimisation algorithm for the bounded diameter minimum spanning tree problem. Vienna University of Technology, Institute of Computer Graphics and Algorithms (2006)
24. Gruber, M., Raidl, G.R.: Variable neighborhood search for the bounded diameter minimum spanning tree problem. PhD thesis, Institute of Computer Graphics and Algorithms, Vienna University of Technology, pp. 1187–1194 (2006)

“Color to Gray and Back” Using DST-DCT, Haar-DCT, Walsh-DCT, Hartley-DCT, Slant-DCT, Kekre-DCT Hybrid Wavelet Transforms

H. B. Kekre, Sudeep D. Thepade and Ratnesh N. Chaturvedi

Abstract The paper shows performance comparison of DST-DCT, Haar-DCT, Walsh-DCT, Hartley-DCT, Slant-DCT and Kekre-DCT Hybrid Wavelet Transforms using Normalization for ‘Color to Gray and Back’. The color information of the image is embedded into its gray scale version using hybrid wavelet transform [HWT] and normalization method. Instead of using the original color image for storage and transmission, gray image (Gray scale version with embedded color information) can be used, resulting into better bandwidth or storage utilization. Among the three algorithms considered the second algorithm give better performance as compared to first and third algorithm. In our experimental results second algorithm for Haar-DCT HWT and Walsh-DCT HWT using Normalization gives better performance in ‘Color to gray and Back’ w.r.t all other transforms in method 1, method 2 and method 3. The intent is to achieve compression of 1/3 and to store and send color images as gray image and to be able to recover the color information afterwards.

Keywords Color embedding · Color-to-gray conversion · Transforms · Hybrid wavelet transforms · Normalization · Compression

H. B. Kekre (✉) · R. N. Chaturvedi

Computer Engineering Department, Mukesh Patel School of Technology,
Management and Engineering, NMIMS University, Mumbai, India
e-mail: hbkekre@yahoo.com

R. N. Chaturvedi

e-mail: ratneshnc@gmail.com

S. D. Thepade

R&D, Pimpri Chinchwad College of Engineering, University of Pune, Pune, India
e-mail: sudeepthepade@gmail.com

1 Introduction

Digital images can be classified roughly to 24 bit color images and 8 bit gray images. We have come to tend to treat colorful images by the development of various kinds of devices. However, there is still much demand to treat color images as gray images from the viewpoint of running cost, data quantity, etc. We can convert a color image into a gray image by linear combination of RGB color elements uniquely. Meanwhile, the inverse problem to find an RGB vector from a luminance value is an ill-posed problem. Therefore, it is impossible theoretically to completely restore a color image from a gray image. For this problem, recently, colorization techniques have been proposed [1–4]. Those methods can re-store a color image from a gray image by giving color hints. However, the color of the restored image strongly depends on the color hints given by a user as an initial condition subjectively.

In recent years, there is increase in the size of databases because of color images. There is need to reduce the size of data. To reduce the size of color images, information from all individual color components (color planes) is embedded into a single plane by which gray image is obtained [5–12]. This also reduces the bandwidth required to transmit the image over the network.

Gray image, which is obtained from color image, can be printed using a black-and-white printer or transmitted using a conventional fax machine [6]. This gray image then can be used to retrieve its original color image.

In this paper, we propose three different methods of color-to-gray mapping technique using DST-DCT, Haar-DCT, Walsh-DCT, Hartley-DCT, Slant-DCT and Kekre-DCT Hybrid Wavelet Transforms and normalization [8, 9], that is, our method can recover color images from color embedded gray images with having almost original color images. In method 1 the color information in normalized form is hidden in LH and HL area of first component as in Fig. 1. In method 2 the color information in normalize form is hidden in HL and HH area of first component as in Fig. 1 and in method 3 the color information in normalize form is hidden in LH and HH area of first component as in Fig. 1. Normalization is the process where each pixel value is divided by maximum pixel value to minimize the embedding error [13].

The paper is organized as follows. Section 2 describes hybrid wavelet transform generation. Section 3 presents the proposed system for “Color to Gray and Back”. Section 4 describes experimental results and finally the concluding remarks and future work are given in Sect. 5.

2 Hybrid Wavelet Transform

Kronecker product is also known as tensor product. Kronecker product is represented by a sign \otimes . The Kronecker product of 2 matrices (say A and B) is computed by multiplying each element of the 1st matrix(A) by the entire 2nd matrix(B) as in Eq. 1:

LL	LH
HL	HH

Fig. 1 Sub-band in transform domain

$$\begin{aligned}
 \begin{bmatrix} a1 & a2 \\ a3 & a4 \end{bmatrix} \otimes \begin{bmatrix} b1 & b2 \\ b3 & b4 \end{bmatrix} &= \begin{bmatrix} a1 \begin{bmatrix} b1 & b2 \\ b3 & b4 \end{bmatrix} & a2 \begin{bmatrix} b1 & b2 \\ b3 & b4 \end{bmatrix} \\ a3 \begin{bmatrix} b1 & b2 \\ b3 & b4 \end{bmatrix} & a4 \begin{bmatrix} b1 & b2 \\ b3 & b4 \end{bmatrix} \end{bmatrix} \\
 &= \begin{bmatrix} a1b1 & a1b2 & a2b1 & a2b2 \\ a1b3 & a1b4 & a2b3 & a2b4 \\ a3b1 & a3b2 & a4b1 & a4b2 \\ a3b3 & a3b4 & a4b3 & a4b4 \end{bmatrix}
 \end{aligned} \tag{1}$$

The hybrid wavelet [14] transform matrix of size $N \times N$ (say ‘ T_{CD} ’) can be generated from two orthogonal transform matrices (say C and D respectively with sizes $p \times p$ and $q \times q$, where $N = p * q = pq$) as given by Eq. 2.

$$C = \begin{bmatrix} c11 & c12 & \dots & c1p \\ c21 & c22 & \dots & c2p \\ \vdots & \vdots & \vdots & \vdots \\ cp1 & cp2 & \dots & cpp \end{bmatrix} \begin{matrix} C1 \\ C2 \\ C3 \\ CP \end{matrix} \quad D = \begin{bmatrix} d11 & d12 & \dots & d1q \\ d21 & d22 & \dots & d2q \\ \vdots & \vdots & \vdots & \vdots \\ dq1 & dq2 & \dots & dqq \end{bmatrix} \tag{2}$$

Here first ‘ q ’ number of rows of the hybrid wavelet transform matrix are calculated as Kronecker product of D and $C1$ which is given as:

$$D \otimes C1 = \begin{bmatrix} d11c11 & d11c12 & \dots & d11c1p & d12c11 & d12c12 & \dots & d12c1p & \dots & d1qc11 & d1qc12 & \dots & d1qc1p \\ d21c11 & d21c12 & \dots & d21c1p & d22c11 & d22c12 & \dots & d22c1p & \dots & d2qc11 & d2qc12 & \dots & d2qc1p \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ dq1c11 & dq1c12 & \dots & dq1c1p & dq2c11 & dq2c12 & \dots & dq2c1p & \dots & dqqc11 & dqqc12 & \dots & dqqc1p \end{bmatrix} \tag{3}$$

For next ‘ q ’ number of rows of hybrid wavelet transform matrix Kronecker product of identity matrix I_q and $C2$ is taken which is given by Eq. 4:

$$I_q \otimes C2 = \begin{bmatrix} c21 & c22 & \dots & c2p & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & c21 & c22 & \dots & c2p & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & c21 & c22 & \dots & c2p \end{bmatrix} \tag{4}$$

Similarly the other rows of hybrid wavelet transform matrix are generated as $I_q \otimes C3, I_q \otimes C4, I_q \otimes C3 \dots$ and the last q row are generated as Eq. 5:

$$I_q \otimes CP = \begin{bmatrix} cp1 & cp2 & \dots & cpp & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & cp1 & cp2 & \dots & cpp & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & cp1 & cp2 & \dots & cpp \end{bmatrix} \tag{5}$$

and the final hybrid wavelet transform matrix is given by Eq. 6:

$$T_{cd} = \begin{matrix} \begin{matrix} d11e11 & d11e12 & \dots & d11elp & d12e11 & d12e12 & \dots & d12elp & \dots & d1qe11 & d1qe12 & \dots & d1qelp \\ d21e11 & d21e12 & \dots & d21elp & d22e11 & d22e12 & \dots & d22elp & \dots & d2qe11 & d2qe12 & \dots & d2qelp \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ d41e11 & d41e12 & \dots & d41elp & d42e11 & d42e12 & \dots & d42elp & \dots & d4qe11 & d4qe12 & \dots & d4qelp \\ c21 & c22 & \dots & c2p & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & c21 & c22 & \dots & c2p & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & c21 & c22 & \dots & c2p \\ c31 & c32 & \dots & c3p & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & c31 & c32 & \dots & c3p & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & c31 & c32 & \dots & c3p \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ cp1 & cp2 & \dots & cpp & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & cp1 & cp2 & \dots & cpp & \dots & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \dots & cp1 & cp2 & \dots & cpp \end{matrix} \end{matrix} \tag{6}$$

3 Proposed System

In this section, we propose a two new color-to-gray mapping algorithm and color recovery method. The ‘Color to Gray and Back’ has two steps as Conversion of Color to Gray Image with color embedding into gray image and Recovery of Color image back.

3.1 Color-to-Gray Step

1. First color component (R-plane) of size $N \times N$ is kept as it is and second (G-plane) and third (B-plane) color component are resized to $N/2 \times N/2$.
2. Second and Third color component are normalized by multiplying each pixel by 255 to minimize the embedding error.
3. Hybrid wavelet transform applied to first color components of image.
4. First component to be divided into four subbands as shown in Fig. 1 corresponding to the low pass [LL], vertical [LH], horizontal [HL], and diagonal [HH] subbands, respectively.
5. **Method 1:** LH to be replaced by normalized second color component, HL to be replace by normalized third color component.
Method 2: HL to be replaced by normalized second color component, HH to replace by normalized third color component.
Method 3: LH to be replaced by normalized second color component, HH to replace by normalized third color component.
6. Inverse hybrid wavelet transform to be applied to obtain Gray image of size $N \times N$.

3.2 Recovery Step

1. Hybrid wavelet transform to be applied on Gray image of size $N \times N$ to obtain four sub-bands as LL, LH, HL and HH.
2. **Method 1:** Retrieve LH as second color component and HL as third color component of size $N/2 \times N/2$ and the the remaining as first color component of size $N \times N$ by replacing LH and HL by zeros.
Method 2: Retrieve HL as second color component and HH as third color component of size $N/2 \times N/2$ and the the remaining as first color component of size $N \times N$ by replacing HL and HH by zeros.
Method 3: Retrieve LH as second color component and HH as third color component of size $N/2 \times N/2$ and the the remaining as first color component of size $N \times N$ by replacing LH and HH by zeros.
3. De-normalize Second and Third color component by multiplying it by 255.
4. Resize Second and Third color component to $N \times N$.
5. Inverse Hybrid wavelet transform to be applied on first color component.
6. All three color component are merged to obtain Recovered Color Image.



Fig. 2 Test bed of image used for experimentation

4 Results and Discussion

These are the experimental results of the images shown in Fig. 2 which were carried out on DELL N5110 with below Hardware and Software configuration.

Hardware Configuration:

1. Processor: Intel(R) Core(TM) i3-2310 M CPU@ 2.10 GHz.
2. RAM: 4 GB DDR3.
3. System Type: 64 bit Operating System.

Software Configuration:

1. Operating System: Windows 7 Ultimate [64 bit].
2. Software: Matlab 7.0.0.783 (R2012b) [64 bit].

The quality of 'Color to Gray and Back' is measured using Mean Squared Error (MSE) of original color image with that of recovered color image. This is the experimental result taken on 10 different images of different category as shown in Fig. 2. Figure 3 shows the sample original color image, original gray image and its gray equivalent having colors information embedded into it, and recovered color image using method 2 for Haar-DCT Hybrid Wavelet Transforms. As it can be observed that the gray images obtained from these methods appears almost like the original gray image, which is due to the normalizing as it reduces the embedding error.

The quality of the matted gray is not an issue, just the quality of the recovered color image matters. So, it is observed from Table 1 that images with higher

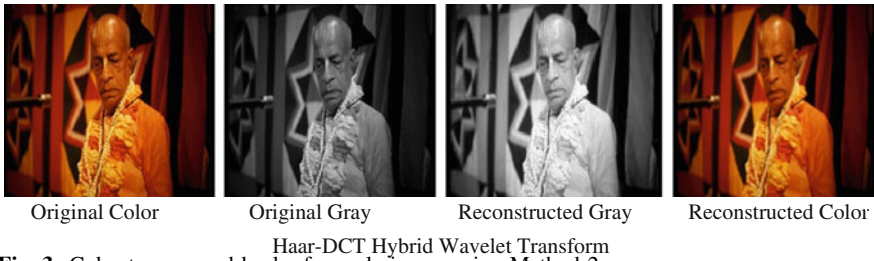


Fig. 3 Color to gray and back of sample image using Method 2

Table 1 MSE of original color w.r.t recovered color image (Method 1)

	Hybrid wavelet transform					
	DST-DCT	Haar-DCT	Walsh-DCT	Hartley-DCT	Slant-DCT	Kekre-DCT
Img 1	480.9374	493.5136	493.5136	892.1496	890.4227	839.4608
Img 2	259.7902	121.3339	121.3339	249.2582	245.2057	227.5410
Img 3	297.0636	280.6049	280.6049	514.9844	522.0832	473.7449
Img 4	200.5412	116.5854	116.5854	277.0018	259.2094	248.1132
Img 5	119.6799	41.9297	41.9297	229.9677	177.7682	180.5100
Img 6	89.5618	77.6847	77.6847	144.9556	143.0372	136.3717
Img 7	491.2481	278.6486	278.6486	336.6507	321.9532	328.8107
Img 8	239.9905	84.2198	84.2198	132.5979	128.1068	130.4926
Img 9	156.4034	99.8884	99.8884	179.1137	180.4896	165.7150
Img 10	432.5759	409.9195	409.9195	510.0717	512.0223	499.3100
Average	276.7792	200.4329	200.4329	346.6751	338.0298	323.0070

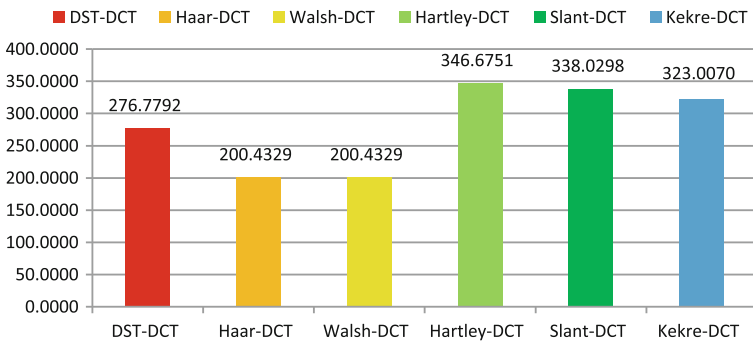


Fig. 4 Average MSE of original color w.r.t recovered color (Method 1)

granularity yields higher MSE and images with lower granularity yields lower MSE and Fig. 4 shows that among all the hybrid wavelet transform tested for method 1 Haar-DCT HWT and Walsh-DCT HWT gives least MSE between Original Color Image and the Recovered Color Image.

Table 2 MSE of original color w.r.t recovered color image (Method 2)

	Hybrid wavelet transform					
	DST-DCT	Haar-DCT	Walsh-DCT	Hartley-DCT	Slant-DCT	Kekre-DCT
Img 1	385.1494	386.8131	386.8100	603.8175	599.5178	564.9952
Img 2	168.6186	94.2015	94.1704	159.7410	153.6871	149.0953
Img 3	227.8618	219.2991	219.3052	334.4131	339.0268	323.3665
Img 4	132.7009	96.5039	96.5038	210.8963	203.4048	192.9994
Img 5	69.6195	25.5616	25.5645	92.2081	64.8220	72.7382
Img 6	68.1849	62.8353	62.8430	102.1031	105.4698	100.8149
Img 7	361.9534	248.9195	248.9873	271.0513	265.8341	267.1678
Img 8	146.4059	62.8723	62.8704	87.5342	84.3602	85.9032
Img 9	110.3315	83.1181	83.1172	131.0886	135.0848	123.5375
Img 10	359.3651	347.9029	347.9204	399.2862	398.4582	393.7425
Average	203.0191	162.8027	162.8092	239.2139	234.9666	227.4361

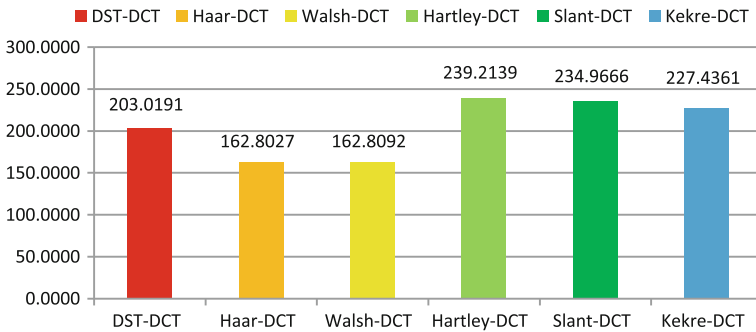


Fig. 5 Average MSE of original color w.r.t recovered color (Method 2)

Table 3 MSE of original color w.r.t recovered color image (Method 3)

	Hybrid wavelet transform					
	DST-DCT	Haar-DCT	Walsh-DCT	Hartley-DCT	Slant-DCT	Kekre-DCT
Img 1	406.6703	423.3483	423.3483	680.3915	701.8834	667.4160
Img 2	167.2253	99.7293	99.7293	185.4235	191.4894	170.1190
Img 3	242.0641	236.8208	236.8208	393.6213	404.1614	360.9510
Img 4	140.2271	90.4560	90.4560	153.0065	144.1542	141.4685
Img 5	71.3757	35.3552	35.3552	168.7440	143.1994	136.7483
Img 6	69.9702	64.5713	64.5713	105.1292	102.7064	97.8116
Img 7	349.3383	240.0452	240.0452	282.4399	275.0722	278.7444
Img 8	156.0923	78.2887	78.2887	108.5049	108.4739	108.6994
Img 9	110.9705	82.5429	82.5429	124.4476	125.2376	116.4830
Img 10	377.7542	368.5165	368.5165	431.1616	438.1089	427.2105
Average	209.1688	171.9674	171.9674	263.2870	263.4487	250.5652

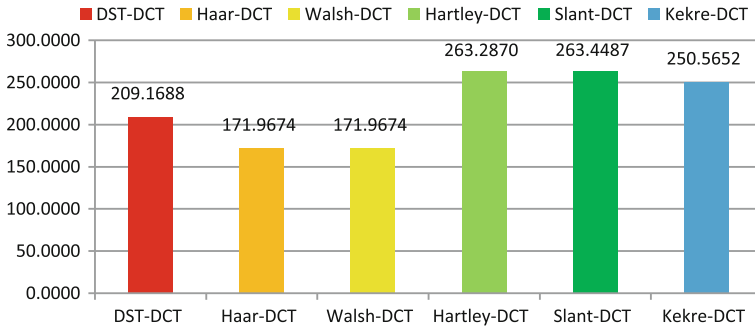


Fig. 6 Average MSE of original color w.r.t recovered color (Method 3)

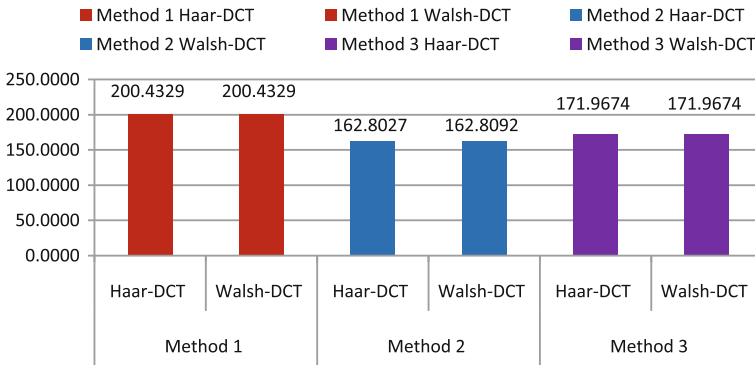


Fig. 7 Average MSE comparison for original color w.r.t recovered color image for the best results of all the 3 methods

In Table 2 it is observed that images with higher granularity yields higher MSE and images with lower granularity yields lower MSE and Fig. 5 shows that among all the hybrid wavelet transform tested for method 2 Haar-DCT HWT and Walsh-DCT HWT gives least MSE between Original Color Image and the Recovered Color Image.

And similarly from Table 3 it is observed that images with higher granularity yields higher MSE and images with lower granularity yields lower MSE and Fig. 6 shows that among all the hybrid wavelet transform tested for method 3 Haar-DCT HWT and Walsh-DCT HWT gives least MSE between Original Color Image and the Recovered Color Image.

From Figs. 4, 5 and 6 for Method 1, Method 2 and Method 3 we can observe that for Haar-DCT HWT and Walsh-DCT HWT we get the best results. To evaluate the best performance among all the three methods, the best results of all the three methods are compared with each other as in Fig. 7.

From Fig. 7 it can be observed that by comparing best results of Method 1, Method 2 and Method 3. Haar-DCT HWT and Walsh-DCT HWT using Method 2 gives best results by obtaining better quality of recovered color image.

5 Conclusion and Future Work

This paper have presented three method to convert color image to gray image with color information embedding into it in two different regions and method of retrieving color information from gray image. These methods allows one to achieve 1/3 compression and to store and send color image as gray image by embedding the color information in a gray image. These methods are based on DST-DCT, Haar-DCT, Walsh-DCT, Hartley-DCT, Slant-DCT and Kekre-DCT Hybrid Wavelet Transforms using Normalization technique. Haar-DCT HWT and Walsh-DCT HWT using method 1, method 2 and method 3 are proved to be the best approach with respect to other hybrid wavelet transforms used in method 1, method 2 and method 3. But among all the methods, method 2 using Haar-DCT HWT and Walsh-DCT HWT gives the best results for ‘Color-to-Gray and Back’. Our next research step could be to test other hybrid wavelet transforms for ‘Color-to-Gray and Back’.

References

1. Welsh, T., Ashikhmin, M.: Mueller K (2002) Transferring color to grayscale image. Proc. ACM SIGGRAPH **20**(3), 277–280 (2002)
2. Levin, A., Lischinski, D., Weiss, Y.: Colorization using optimization. ACM Trans. Graph. **23**, 689–694 (2004)
3. Horiuchi, T.: Colorization algorithm using probabilistic relaxation. Image Vis. Comput. **22**(3), 197–202 (2004)
4. Yatziv, L., Sapiro, G.: Fast image and video colorization using chrominance blending. IEEE Trans. Image Process. **15**(5), 1120–1129 (2006)
5. Kekre, H.B., Thepade, S.D.: Improving ‘Color to Gray and Back’ using Kekre’s LUV Color Space. In: IEEE International Advance Computing Conference 2009, (IACC 2009), pp. 1218–1223, Thapar University, Patiala
6. de Queiroz, R.L., Braun, K.M.: Color to gray and back: color embedding into textured gray images. IEEE Trans. Image Process. **15**(6), 1464–1470 (2006)
7. Kekre, H.B., Thepade, S.D., Parkar, A.: An extended performance comparison of colour to grey and back using the Haar, Walsh, and Kekre wavelet transforms. Int. J. Adv. Comput. Sci. Appl. Special issue on artificial intelligence (IJACSA), 92–99
8. Kekre, H.B., Thepade, S.D., Chaturvedi, R., Gupta, S.: Walsh, Sine, Haar & Cosine transform with various color spaces for ‘color to gray and back’. Int. J. Image Process. (IJIP) **6**(5), 349–356 (2012)
9. Kekre, H.B., Thepade, S.D., Chaturvedi, R.: Improved performance for “color to gray and back” for orthogonal transforms using normalization. Int. J. Comput. Eng. Res. **3**(5), 54–59 (2013)

10. Kekre, H.B., Thepade, S.D., Chaturvedi, R.: New faster ‘color to gray and back’ using normalization of color components with orthogonal transforms. *Int. J. Eng. Res. Technol. (IJERT)* **2**(4), 1880–1888 (2013). ISSN: 2278-0181
11. Kekre, H.B., Thepade, S.D., Chaturvedi, R.: ‘Color to gray and back’ using normalization of color components with Cosine, Haar and Walsh Wavelet. *IOSR J. Comput. Eng. (IOSR-JCE)* **10**(5), 95–104 (2013). e-ISSN: 2278-0661, p-ISSN: 2278-8727
12. Kekre, H.B., Thepade, S.D., Chaturvedi, R.: Information hiding for “color to gray and back” with Hartley, Slant and Kekre’s wavelet using normalization. *IOSR J. Comput. Eng. (IOSR-JCE)* **10**(6), 50–58 (2013). e-ISSN: 2278-0661, p-ISSN: 2278-8727
13. Kekre, H.B., Thepade, S.D., Chaturvedi, R.N.: Novel transformed block based information hiding using cosine, sine, Hartley, Walsh and Haar transforms. *Int. J. Adv. Eng. Technol.* **6**(1), 274–281 (2013). ©IJAET ISSN: 2231-1963
14. Kekre, H.B., Sarode, T.K., Thepade, S.D.: Inception of hybrid wavelet transform using two orthogonal transforms and it’s use for image compression. *Int. J. Comput. Sci. Inf. Secur.* **9**(6), 80–87 (2011)

Music Genre Classification Using Music Information Retrieval and Self Organizing Maps

Abdul Nafey Ahmad, Chandra Sekhar and Abhinav Yashkar

Abstract Recent times have seen a wide proliferation of use of Genre as basis of music classification and documentation. This paper proposes a novel approach of music genre classification using unsupervised Neural Network method—Kohonen Self Organizing Maps (SOM). Various music features like timbral features (attack slope), spectral distribution of energy (brightness) and tonality (major or minor) were extracted from the audio files using MATLAB toolbox. An inventory of a number of musical pieces was developed for building clusters of different genre. A database containing the extracted musical features was then clustered using SOM. Each cluster encapsulated a different genre of music. The centroid of each cluster was taken as the representative point of that genre in the considered n dimensional musical feature space. Then a number of songs were then mapped onto the said space and classified into different genre depending on their Euclidean distance from the center of each cluster.

Keywords Self organizing maps · Music genre classification · Music information retrieval · SOM · Neural classification · Soft computing

1 Introduction

The current era has seen a development of social networks where music and musical content is heavily shared. For example, the popular video sharing website YouTube has got thousands of musical videos with cumulative views in billions. A radical increase in the amount and diversity of music available has caused an increasing emphasis on finding means to classify music in meaningful ways. Music is generally classified into different Genres like Rock, Pop, Country, Soul etc. on

A. N. Ahmad (✉) · C. Sekhar · A. Yashkar
Department of Civil Engineering, Indian Institute of Technology, Kharagpur, India

the basis of manual perception. These genres differ not only in sound, mood and history but also target audience.

Even though the genre-wise classification of music is ubiquitous there is seemingly a dearth of technology that can automatically classify the music. One of the reasons for this is that genre is considered to be a somewhat subjective classification [1]. There are songs that might be classified into multiple genres, or some which fall into none of the different genre's identified. Nevertheless, automatic genre classification can find use in radio stations, audio-visual archive management, entertainment and others.

Although it is hard to precisely define the specific characteristics of a music genre, it is generally agreed, that audio signals of music belonging to the same genre contain certain common characteristics since they are composed of similar types of instruments and having similar rhythmic patterns [2]. Unsurprisingly, humans have an uncanny sense of correctly identifying the genre better than most digital solutions. Broadly speaking, humans are apt at genre classification; studies have shown that humans can accurately predict a musical genre based on a 250 ms long audio clip [3].

Thus, it seems possible to develop automated systems which can pick up higher level features of digital sound clip and use it to classify the sound. One very important tool in this regard is the MIR toolbox for MATLAB from researchers at University of Jyvaskyla [4, 5]. This exposes a number of functions which facilitates analysis of sound files. Selection of specific sound features among those provided by the toolbox enables the classification of genre.

2 Feature Selection

The process of selection of music features was largely empirical. A huge number of features are available for selection. The functions exposed by the toolbox can be arranged under three main categories: rhythm, pitch and timbre. Rhythm refers to temporal (horizontal) arrangement of music. It is related with concepts such as time signatures, beats and tempo of the songs. Next category is of functions which are related to pitch. This has to do with tonality, harmony and modes. This is sometimes called the vertical arrangement of music. Finally, we use some functions which expose the timbral content of the music. Timbre has often been called the color of music. It has close relation with the unique feel of sound of every instrument.

Among these three groups there are a huge number of functions, not all of which can be or should be used. Selection of the features from among all the available choices was done empirically. All the probable features were tested on small sets and their variation among different songs was observed. Those functions that hardly showed any variance or behaved completely randomly were discarded. Some functions were, like key strength, tempo, selected by default, because on basis of musical theory they have to play a crucial role in judgment of genre. These

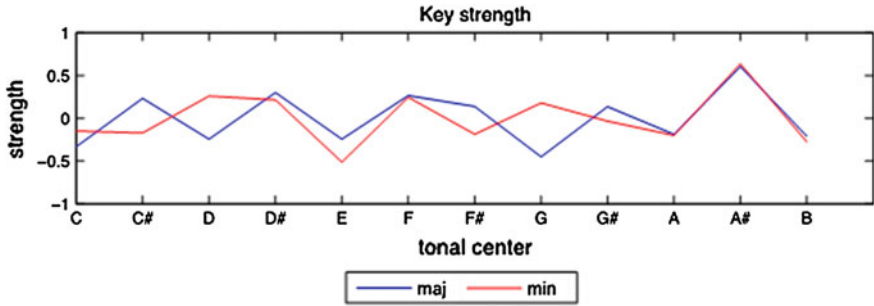


Fig. 1 Graph of key strength across different keys (source MIRtoolbox 1.5 user guide)

experiments finally led to a finite group of reliable indicators, each of which is explained subsequently.

2.1 Pitch

Key Strength [6]: The key strength is a score between +1 and -1 associated with each possible key candidate from C to B including sharp notes. The song is evaluated for key strength in each of the twelve keys and both major and minor key strength are evaluated. Generally, a happy sounding song will have a higher major key strength whereas high minor key strength will lend a darker hue to the song (Fig. 1).

Inharmonicity: This evaluates the amount of frequencies that are not multiples of fundamental frequency as a value between 0 and 1. More precisely, it is the amount of energy contained in vibration outside the fundamental frequency.

2.2 Rhythm

Pulse Clarity [7]: It is a measure of clarity of pulsation indicating the strength of the beats in the track. Generally highly rhythm based genres like hip-hop or dance music will have greater pulse clarity.

Tempo [8]: This is the measure of the “speed” of track. This measures the beat of the track. Usually higher beats per minute (bpm) lend to the track an aggressive persona, whereas a low bpm is used with more emotional songs.

2.3 Timbre

Brightness: This is a value between 0 and 1 which first finds a cutoff frequency and then evaluates how much energy is contained in all the frequencies higher than cutoff.

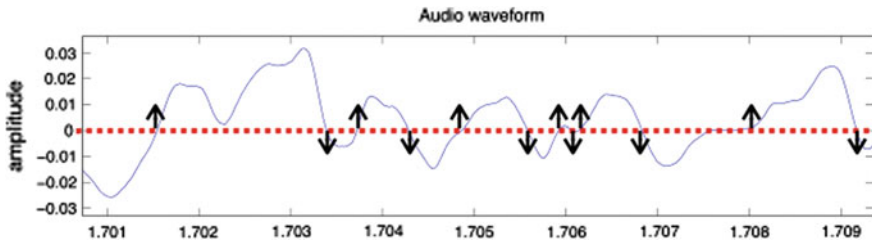


Fig. 2 Zero cross reflects the number of time the signal crosses zero line (*source* MIRtoolbox 1.5 user guide)

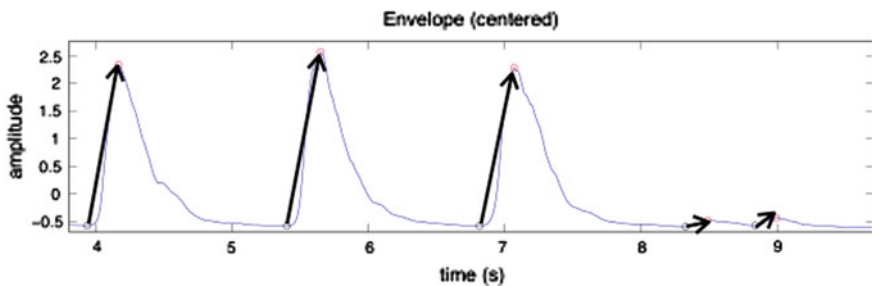


Fig. 3 Attack slope is the slope of the rising edge of audio signal (*source* MIRtoolbox 1.5 user guide)

Zero Cross: A simple indicator of noisiness consists in counting the number of times the signal crosses the X-axis (or, in other words, changes sign) (Fig. 2).

Roughness: This measures the sensory dissonance, or roughness, related to the beating phenomenon caused by two sinusoids which are close in frequency.

Attack Slope: The rising slope of the signal is referred to as attack slope. It is related to harshness of the sound (Fig. 3).

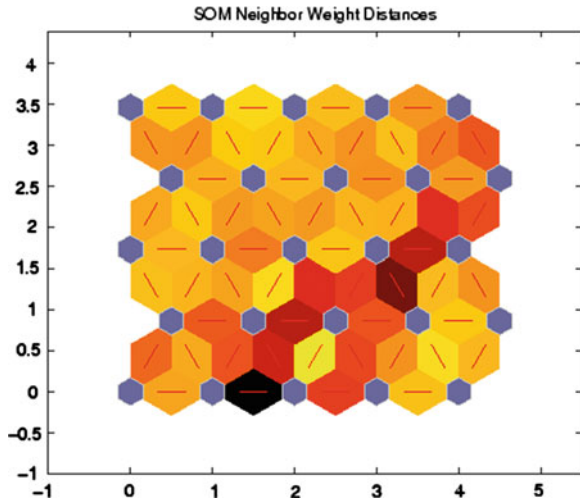
Irregularity: The irregularity of a spectrum is the degree of variation of the successive peaks of the spectrum.

Rolloff: Estimate the frequency under which 0.85 of the total energy lies.

3 Kohonen Self Organizing Map

Neural networks are artificial constructs that are capable of machine learning and pattern recognition. They are represented as a system of interconnected neurons mimicking the wiring of natural neural networks in our brain. Each of the connections between neurons has got a value associated with them called the weight of neurons. As the neural network is exposed to new data these values adjust

Fig. 4 Self-organizing map



themselves in an attempt to learn the patterns of data. Neural networks have been used in widely varying fields and they have much proved to be very reliable both in terms of speed and quality of results.

Self-organizing maps are type of artificial neural network that are trained using unsupervised learning. The output is a 2-dimensionl discretized representation of the input space of the training samples. SOM consist of components called nodes or neurons. The nodes are associated with a weight vector which has the input data and the position on the map space. The nodes are placed in a 2-dimensional regular spacing, generally in a hexagonal or a rectangular grid. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the data space vector (Fig. 4).

SOM can be used to locate and create clusters in data. Clusters in data are regions of homogeneity which can be treated to be comprised of data points having similar numeric values. We are employing SOM in this project to make sense of a huge data and to logically associate data points into different categories. Clusters provide an easy way to do it. Neural network tool ClusterizerXL is proprietary software that can be installed to Excel as a plug in and provides the functions for easy clustering.

4 Methodology

Here we describe the methods that we employed in our process in a step-by-step manner.

4.1 Selection of Genres

There are a wide number of genres available in the music scene today. Online sources such as Wikipedia list hundreds of genres. Overtime genres accumulate different changes, leading to creation of new genres. Thus, these can be arranged into hierarchical structures depending on history of each genre. However, there are some broad classifications in genres that have remained static throughout the years and remain unchanged with time. Some of these are Rock, Hip-Hop, Pop. These are broad categories and each of them has distinctive sound. In our analysis the basic categories that we have used are: Rock, Pop, Hip-Hop (Rap).

4.2 Selection of Sample

The sample we selected had to be representative of the kind of clustering we wanted. As there had to be three clusters (one each for Rock, Pop, Hip-Hop), a number of songs widely associated with each of these classifications were selected so that all the categories were represented in the initial set. There are some songs which cannot be easily categorized in either of the classes; such songs were carefully avoided. The idea while building the set was to have each of the three distinct sounds present.

4.3 Selection of Sound Clip

Music information retrieval is a heavy on processor and analyzing the whole waveform of the song can take a good amount of time. So instead of full song, only 10 s long chunks were taken. It was ensured that each chunk was representative of the tone and the texture of the whole song. Thus, in rap songs the chorus section was avoided as they usually sound more like a pop track. Also, instrumental sections in rock songs were avoided.

4.4 Feature Extraction

The sound clips were then processed by the MATLAB code to extract features. The code digitized the sound clip and executed various functions on it to extract their features. The MATLAB code had to be streamlined and optimized for proper operation. Finally, the code took around one minute to process each song and compute the value of each of the feature.

4.5 Data Entry and Clustering

Data was copied from MATLAB to an Excel spreadsheet for further evaluation and clustering. Finally after collection of all data Neuro Tools ClusterizerXL was used to clusterize the whole of data into three distinct clusters. The results obtained have been recorded in a following section.

4.6 Normalization

The values selected were to bring uniformity to data. To normalize the value in a column, the values were placed on a scale where 0 corresponded to minimum value in that column and 1 corresponded to the maximum. This was essential because some columns might differ in values by orders of magnitude.

4.7 Testing

After points for each of the cluster were located we moved onto finding a center for each cluster. For this we assumed that each of the features formed an axis of an n-dimensional space. Here each point can uniquely represent a song. Next, the centroid of all the points in a cluster was located. This point represents the cluster. Now, each cluster can be identified by this centroid. The distance from centroid will represent the relative tendency of a song to be classified into a particular group. So songs were tested by mapping onto this space for accuracy.

5 Sample of Dataset

Hereby is presented a sample of the dataset in Table 1. A point of interest is that there are always some that end up in cluster contrary to the one where they are expected to go. This is inevitable due to inaccuracy of the tools employed as well as other subjectivity of genre classifications. But on the whole these discrepancies have minor effect and comprise of less than ten percent of all data points.

6 Result

Features were extracted from a set of eighty eight songs comprising mainly of popular songs that have dominated the music boards recently. There was equitable distribution of songs from the three main genres: Pop, Rock and Hip-Hop. Feature

Table 1 A sample of dataset

Cluster	Expected Cluster	Song Name	Artist	Genre	Major Key Strength	Minor Key Strength	Pulse Clarity	Brightness	Roughness	Zero Cross	Attack Slope	Tempo	Rolloff	Irregularity	Inharmonicity	Entropy
1	2	BLUE TRAIN	JOHN COLTRANE	POP (JAZZ)	0.618	0.409	0.443	0.629	826	2206	2.4E+06	150	10014	0.145	0.500	0.914
1	1	BOULEVARD OF BROKEN DREAMS	GREEN DAY	ALT. ROCK	0.765	0.580	0.221	0.577	4037	1524	8.4E+06	165	7344	0.192	0.476	0.905
1	1	IN MY PLACE	COLDPLAY	ROCK	0.798	0.539	0.514	0.557	3722	1356	2.9E+06	141	6253	0.207	0.494	0.894
1	1	PAYPHONE	MAROON 5	ROCK	0.722	0.722	0.601	0.598	2306	1588	1.3E+07	110	9125	0.268	0.480	0.912
1	1	FADE	STAINED	ALT. ROCK	0.616	0.615	0.660	0.587	3934	1380	1.0E+07	117	7683	0.181	0.464	0.906
1	2	LOVE STORY	TAYLOR SWIFT	POP (COUNTRY)	0.833	0.651	0.585	0.633	3449	1819	7.1E+06	119	9354	0.166	0.479	0.923
1	1	LEAVE OUT ALL THE REST	LINKIN PARK	ALT. ROCK	0.761	0.879	0.597	0.612	4133	1601	6.1E+06	107	8627	0.244	0.478	0.911
1	1	GUILTY	RASMUS	ROCK	0.641	0.780	0.477	0.553	5571	1235	7.6E+06	135	6414	0.424	0.468	0.897
1	1	UNWANTED	AVIRL LAVIGNE	ALT. ROCK	0.520	0.661	0.402	0.653	2609	1984	3.9E+06	108	7821	0.111	0.457	0.913
1	1	GIVE ME YOUR NAME	DEAD BY SUNRISE	ALT. ROCK	0.783	0.671	0.462	0.541	1775	1215	6.5E+06	129	8283	0.323	0.485	0.897
1	1	FAINT	LINKIN PARK	ALT. ROCK	0.595	0.632	0.517	0.590	5566	1775	1.1E+07	135	7263	0.179	0.483	0.912
1	1	YOU GIVE LOVE A BAD NAME	BON JOVI	ROCK	0.724	0.614	0.283	0.707	5572	2491	4.5E+06	123	7631	0.126	0.494	0.925
1	1	NO CURTAIN CALL	MAROON 5	ROCK	0.589	0.681	0.701	0.532	4115	1142	1.2E+07	168	7097	0.224	0.463	0.896
1	1	SUMMER OF 69	BRYAN ADAMS	ROCK	0.803	0.827	0.436	0.589	2080	1481	4.9E+06	139	9390	0.248	0.478	0.913
1	1	BURN IT DOWN	LINKIN PARK	NU METAL	0.718	0.629	0.547	0.509	1632	1231	7.3E+06	110	8749	0.215	0.459	0.894
1	1	CARNIVAL OF RUST	POETS OF THE FALL	ALT. ROCK	0.566	0.748	0.300	0.635	2968	1311	6.7E+06	175	9885	0.382	0.492	0.923
1	1	CHOP SUEY	SYSTEM OF A DOWN	ROCK	0.730	0.619	0.430	0.585	2602	1493	5.5E+06	125	6906	0.555	0.482	0.895
1	3	THE REAL SLIM SHADY	EMINEM	HIP HOP	0.584	0.631	0.518	0.559	2591	1586	2.1E+07	105	8229	0.204	0.473	0.908
1	2	RHYTHM DIVINE	ENRIQUE	POP	0.560	0.728	0.803	0.618	1671	1615	5.2E+06	102	9973	0.332	0.469	0.925
1	1	FADE TO BLACK	METALLICA	ROCK	0.578	0.791	0.545	0.627	269	1752	1.4E+06	111	8950	0.238	0.484	0.902
2	1	LIKE SUICIDE	SEETHER	ROCK	0.567	0.786	0.560	0.500	1743	1478	1.3E+06	189	4691	0.116	0.437	0.875
2	1	DEATH OF DEATH	IRON MAIDEN	ROCK	0.479	0.639	0.108	0.403	1304	675	2.8E+06	152	5742	0.113	0.480	0.869
2	2	BLUE SKIES	DUKE ELLINGTON	POP (JAZZ)	0.509	0.502	0.376	0.457	545	1087	1.7E+06	107	3676	0.115	0.430	0.853
2	2	FIX YOU	COLDPLAY	POP (JAZZ)	0.812	0.695	0.152	0.363	233	870	3.8E+05	74	3541	0.261	0.483	0.827
2	2	MY FUNNY VALENTINE	FRANK SINATRA	POP (JAZZ)	0.311	0.375	0.237	0.314	431	482	1.0E+06	104	3703	0.209	0.497	0.836
2	2	I'VE GOT YOU UNDER MY SKIN	FRANK SINATRA	POP (JAZZ)	0.577	0.565	0.215	0.404	48	851	3.2E+05	126	4309	0.190	0.478	0.866
2	2	NAIMA	JOHN COLTRANE	POP (JAZZ)	0.429	0.573	0.088	0.563	40	661	3.3E+05	73	4365	0.392	0.441	0.853
2	2	LATE GOODBYE	POETS OF THE FALL	ROCK	0.804	0.689	0.260	0.521	647	1346	5.4E+05	162	7542	0.213	0.480	0.890
2	2	DO I MOVE YOU	NINA SIMONE	POP (JAZZ)	0.816	0.885	0.243	0.541	904	1198	2.6E+06	107	5322	0.160	0.462	0.881
2	2	QUIT PLAYING GAMES WITH ME	BACKSTREET BOYS	POP	0.586	0.471	0.796	0.632	1995	1843	1.4E+07	107	11186	0.236	0.479	0.902
2	2	SOMEBODY'S ME	ENRIQUE	POP	0.780	0.749	0.414	0.574	1292	1356	1.9E+06	171	7416	0.193	0.478	0.906
2	2	SWEET NOTHING	FLORENCE + THE MACHINE	POP (DANCE)	0.568	0.713	0.810	0.638	3347	1927	1.1E+07	128	9043	0.128	0.453	0.919
2	2	A REMARK YOU MADE	WEATHER REPORT	POP	0.603	0.713	0.235	0.247	74	426	3.1E+05	184	3423	0.845	0.445	0.798
2	2	EMPTINESS	GAJENDRA VERMA	POP	0.758	0.598	0.395	0.371	425	968	2.7E+06	166	7434	0.279	0.475	0.866
2	2	DON'T KNOW WHY	NORAH JONES	POP	0.533	0.398	0.460	0.465	454	873	2.5E+06	175	7091	0.223	0.446	0.876
2	2	SKIPALONG	LENKA	POP	0.560	0.687	0.618	0.382	1049	998	8.8E+06	65	6657	0.177	0.464	0.870
2	1	WAKE ME UP WHEN SEPTEMBER RAINS	GREEN DAY	ROCK	0.723	0.549	0.426	0.495	634	1264	1.4E+06	70	7544	0.239	0.484	0.883
2	2	MOANIN'	ART BLAKEY & THE JAZZ MESSENGERS	POP (JAZZ)	0.438	0.550	0.530	0.652	224	1485	9.0E+05	128	4990	0.067	0.477	0.885
2	2	BE YOURSELF	AUDIOSLAVE	POP	0.575	0.489	0.574	0.409	329	833	6.6E+05	117	6410	0.163	0.479	0.870
2	1	CRAWLING	LINKIN PARK	ALT. ROCK	0.488	0.472	0.575	0.491	564	1052	6.4E+06	105	6440	0.177	0.457	0.885
2	2	GIANT STEPS	JOHN COLTRANE	POP (JAZZ)	0.218	0.330	0.398	0.488	319	1103	1.7E+06	145	6223	0.171	0.469	0.882
3	1	AMERICAN IDIOT	GREEN DAY	ROCK	0.393	0.492	0.665	0.630	3087	1556	1.9E+07	186	7205	0.420	0.489	0.911
3	1	ONE MORE NIGHT	MAROON 5	ROCK	0.536	0.552	0.432	0.614	1355	1603	8.1E+06	186	9083	0.342	0.469	0.917
3	1	EAT YOU ALIVE	LIMP BIZKIT	ALT. ROCK	0.411	0.541	0.528	0.505	3537	1160	1.1E+07	138	8658	0.110	0.470	0.901
3	3	DEATH OF AUTO-TUNE	JAY-Z	HIP HOP	0.208	0.225	0.137	0.577	3413	1470	1.2E+07	86	8500	0.205	0.474	0.913
3	3	EMPIRE STATE OF MIND	JAY-Z	HIP HOP	0.452	0.462	0.643	0.598	4087	1722	2.3E+07	173	8293	0.335	0.464	0.913
3	1	FALLING	STAINED	ALT. ROCK	0.334	0.503	0.323	0.693	3315	2345	8.1E+06	112	7845	0.275	0.491	0.919
3	1	ON TO THE NEXT ONE	JAY-Z	HIP HOP	0.344	0.423	0.416	0.632	4262	1592	1.4E+07	178	8395	0.230	0.474	0.916
3	1	DECODE	PARAMORE	ALT. ROCK	0.543	0.652	0.318	0.582	4765	1486	5.7E+06	161	6346	0.142	0.469	0.904
3	3	INTERNATIONAL LOVE	PITBULL FEAT CHRIS BROWN	HIP HOP	0.366	0.300	0.446	0.601	2186	1557	7.0E+06	120	8699	0.300	0.471	0.909
3	3	CLIQUE	KANYE WEST FEAT LIL B	HIP HOP	0.336	0.498	0.346	0.561	1759	1312	1.0E+07	109	9356	0.485	0.477	0.907
3	3	AIRPLANE	B.O.B	HIP HOP	0.307	0.258	0.266	0.517	1873	1421	7.6E+06	186	5705	0.124	0.467	0.889
3	1	CALIFORNIACTION	RED HOT CHILI PEPPERS	ROCK	0.565	0.610	0.608	0.590	5817	1195	9.3E+06	191	6898	0.150	0.466	0.891
3	3	BUSINESS	EMINEM	HIP HOP	0.134	0.229	0.678	0.527	991	1492	1.8E+07	193	6325	0.112	0.468	0.889
3	3	LOSE YOURSELF	EMINEM	HIP HOP	0.284	0.400	0.148	0.586	1680	1610	8.0E+06	171	7994	0.190	0.457	0.900
3	3	STAN	EMINEM	HIP HOP	0.551	0.507	0.277	0.559	3263	1066	1.5E+07	158	8495	0.256	0.494	0.909
3	3	SUPERMAN	EMINEM	HIP HOP	0.252	0.305	0.592	0.624	2763	1589	1.3E+07	130	9925	0.362	0.477	0.924

values for each of the song was extracted and stored in a spreadsheet. These values were then normalized as explained above. Next they were clustered using standard clustering tools. The values of clusters were averaged to obtain the location of centroid of each of the clusters in n-dimensional feature space. The coordinate of these centroids are listed in the Table 2. Note that not all of the columns are represented due to space considerations.

Table 2 Averaged and normalized cluster values

Sl. No.	Cluster genre	Key strength		Pulse clarity	Brightness	Attack slope	Tempo	Rolloff
		Major	Minor					
1.	Rock	0.742	0.571	0.525	0.649	0.239	0.485	0.487
2.	Pop	0.623	0.592	0.403	0.432	0.099	0.481	0.263
3.	Hip-hop	0.360	0.492	0.437	0.651	0.432	0.668	0.465

Table 3 Verification of the results

Sl. No.	Name of song	Artist	Dist. from cluster			Resultant cluster	Expected cluster
			1	2	3		
1.	Nobody’s home	Avril Lavigne	0.233	0.458	0.623	1	1
2.	Paradise	Coldplay	0.461	0.733	0.693	1	1
3.	Praan	G. Schyman	0.756	0.366	0.967	2	2
4.	Country roads	John Denver	0.757	0.588	1.13	2	2
5.	Spies	Coldplay	0.741	0.861	1.095	1	1
6.	When the wind blows	Iron Maiden	1.009	0.695	1.281	2	1
7.	My heart will go on	Celine Dion	1.034	0.704	1.208	2	2
8.	Mockingbird	Eminem	0.486	0.597	0.464	3	3
9.	When I’m gone	Eminem	0.973	0.970	0.798	3	3
10.	Without me	Eminem	0.925	0.92	0.722	3	3
11.	Airplanes	B.o.B	0.787	0.576	0.637	2	3

6.1 Verification of Results

These above values represent the centroid of each of the three clusters that have been formed. The verification of results is of utmost importance if this model has to have any validity. For verification a test set of songs is developed which is then mapped on the feature space. The distance of each song from the cluster centroids are calculated. The listed genre of the song should correspond to the genre that is suggested by distance comparison. The results are presented in Table 3.

In Table 3, results of the test on a total sample set of eleven songs is shown. The distance shown in the table is, as explained earlier, is the Euclidean distance from centroids of the clusters. The three clusters 1, 2 and 3 represent the genres of Rock, Pop and Hip-hop respectively. If a Rock song is tested, it is expected to be located closest to cluster 1. Here, the last column (expected cluster) shows the most likely cluster of the song on basis of its listed genre. The second last column (resultant column) is the actual cluster in which the song falls on basis of its song feature values. In an ideal scenario the expected and resultant values will always be equal, but in this case they are equal for 9 out of 11 samples. This means an accuracy of 81.8 %; which is quite a satisfactory value.

7 Conclusion

The subjectivity of genre descriptions imposes certain limits on objectively writing subroutines for genre identification. However, the versatile nature of soft computing tools allows us to circumvent these shortcomings. In particular we have seen that certain audio features map are very good predictors of genre distinction. Moreover, these are robust systems which are self-correcting and very efficient. Still, there remains much research to be done, especially in discovering better audio features and correlating them to genre identity. A successful module of genre identification will find use all across the spectrum of music industry. It can be used for building automated playlists or for suggesting music to user, illustrate just two of its uses. Among other things, this research paper hints at the possibility of building larger datasets and incorporating more clusters to build more versatile software. However, even on its own, this model can be relied upon for genre identification and classification.

References

1. Li, T., Tzanetakis, G.: Factors in automatic musical genre classification of audio signals. In: IEEE WASPAA (2003)
2. Bagci, U., Erzin, E.: Automatic classification of musical genres using inter-genre similarity. *Signal. Process. Lett., IEEE* (2007)
3. Tzanetakis, George, Cook, Perry: Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **10**(5), 293–302 (2002)
4. Lartillot, O., Toivainen P.: A matlab toolbox for musical feature extraction from audio. In: International conference on digital audio effects, Bordeaux (2007)
5. Lartillot, O., Toivainen, P., Eerola, T.: A Matlab Toolbox for Music Information Retrieval. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R. (eds.) *Data analysis, machine learning and applications. Studies in classification, data analysis, and knowledge organization*, pp. 261–268. Springer, Heidelberg (2008)
6. Gómez, E.: Tonal description of music audio signal. In: Ph.D. thesis, Universitat Pompeu Fabra, Barcelona (2006)
7. Lartillot, O., Eerola, T., Toivainen, P., Fornari, J.: Multi-feature modeling of pulse clarity: design, validation, and optimization. In: International conference on music information retrieval, Philadelphia (2008)
8. Scheirer, E.D.: Tempo and beat analysis of acoustic musical signals. *J. Acoust. Soc. Am.* **103–1**, 588–601 (1998)
9. McKay, C., Fujinaga I.: Automatic genre classification using large high-level musical feature sets. *ISMIR* (2004)

Implementation of a Private Cloud: A Case Study

Prachi Deshpande, S. C. Sharma and S. K. Peddoju

Abstract Availability of Cloud environment setup is sometimes difficult at personal level for a researcher, especially a beginner. In this case, open source tools, softwares can help a great deal to build and deploy a private cloud. Deploying cloud for research purpose is quiet time consuming due to scattered information and variety of options. In this paper, an attempt has been done for providing collective information of various measures to deploy a private cloud. In-depth analysis of different real-time errors during installation of a cloud setup along with solutions is also reported in this paper. This attempt will certainly helpful in minimizing the time and efforts to setup the cloud and makes deployment easier. Open source operating system (OS) Ubuntu-12.04 long time support (LTS) and OpenNebula cloud computing framework are used for deployment of this private cloud.

Keywords Cloud · Deployment · Hypervisor · KVM · OpenNebula · Open source · VM

1 Introduction

Rapid information processing and access is becoming need of the day. Evolution of cloud computing has its roots to fulfill this ever increasing demand of processing information in real-time manner. Cloud consists of a cluster of resources,

P. Deshpande (✉) · S. C. Sharma · S. K. Peddoju
Indian Institute of Technology Roorkee, Roorkee 247667, Uttarakhand, India
e-mail: deprachi3@gmail.com

S. C. Sharma
e-mail: scs60ft@gmail.com

S. K. Peddoju
e-mail: prof.sateesh@gmail.com

services, and technologies distributed over network. Principle aim of cloud computing is to reduce the cost of service usage, infrastructure development and maintenance, availability and very importantly security in data processing [1]. According to [2], cloud is a parallel and distributed computing system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements (SLA) established through negotiation between the service provider and consumers. According to National institute of standards and technology (NIST) [3], cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Cloud computing is a assimilation of features of different technologies and standards like virtualization of resources, web services and architectures, distributed computing, and data center automation and autonomic computing [4]. It inherits some of its characteristics from mainframe computer, client-server model, grid computing, utility computing, peer-to-peer [5] computing and autonomic computing [2, 6]. In the infrastructure as a service (IaaS) layer of cloud, basic services like computers (physical or virtual), servers, virtual local area networks, firewalls, file based and raw storage are provided to user.

The users need to pay as per its utilization. ‘Amazon elastic compute (EC) 2’, ‘HP cloud’, ‘Rackpace’ are some of the IaaS providers. Computing platform like OS and programming language platform is provided to users via platform as a service (PaaS). ‘Cloud Fondry’, ‘Google App Engine’ and ‘Jelastic’ are some of the PaaS providers. Different application development is made possible by virtue of web servers. In software as a service (SaaS) layer, cloud providers install and operate application software in the cloud. Users can access the software from cloud clients [2]. Based on coverage area and distribution, the cloud deployment models are classified as a private cloud, public cloud, hybrid cloud, community cloud. Many organizations when share the capabilities of cloud, it is referred as community cloud [4]. For research purpose, a private cloud is sufficient to carry out the experimentations. The basic information and various tools for a cloud setup are discussed in Sect. 2. Installation of cloud setup is conversed in Sect. 3 along with its trouble shooting aspect. The performance of the cloud setup is presented in Sect. 4.

2 Backbone for a Private Cloud Installation

This section provides off-the-shelf information for the installation of a private cloud. For research purpose, private cloud has to be built by using available open source platforms. These platforms are acting as a tool for managing the cloud resources and services. The key virtualization technology in a cloud setup is a

hypervisor. Hypervisor is a software which is used in between the hardware and OS for virtualization purpose. It allows hardware virtualization so that multiple virtual machines (VM) can run on a single host. Memory management, CPU scheduling are also performed by hypervisor. This virtualization technology falls into different categories as full virtualization, paravirtualization, and hardware assisted virtualization [4]. Few of the hypervisors available in the market are discussed below. From the feature comparison, one can decide about the suitability of a specific hypervisor according to working environment and desired results.

- **XEN**—It allows separation of the domains which lead to strong isolation and security. The OS acting as domain0 (Dom0) has direct communication with hardware of host. Dom0 manage and launch domainUser (DomU) which is a guest OS. Network management is based on ‘First-in-first-Out (FIFO)’ scheduling. XEN can provide paravirtualization and full virtualization. ‘Dom0’ and ‘DomU’ structure makes XEN different from other hypervisors due to its isolation property. XEN can be compared with kernel virtual machine (KVM) when hardware VM with XEN is provided, as it is similar to KVM, which provides full virtualization [7].
- **KVM**—It is an open source product by ‘RedHat’. It is a module that can be used for converting Linux kernel into a hypervisor. It has features like live migration of VM, support for different types of guest OS, isolation policy decision for VM by administrator to provide security, and hardware devices having Linux support.
- **VMware VSphere**—VSphere is the product by VMware. The main features of VSphere are distributed locking, supported with sharing cell, load balancing based on CPU topology, disk management using latency aware priority based scheduler for network management and priority based network input output control and distributed switches [8].

Different parameters can be used to compare hypervisors to choose among available options. KVM is the solution for virtualizing Linux kernel and it provides full virtualization by direct communication with hardware. KVM is opted in this work as there is no performance issues of host are observed.

Major issues in managing the cloud resources are ensuring the availability of resources on demand and load balancing while providing services to the users [9]. There are different types of virtual infrastructure (VI) managers available for this purpose such as:

- **Eucalyptus**: This is the first open-source framework provided for building ‘IaaS’ clouds. It has compatibility with ‘Amazon EC2’ and has XEN, KVM, VMware back ends, virtual networking, and EC2 compatible command line interface (CLI) and Web portal interfaces [10].
- **OpenNebula**: It is the most famous cloud setup open source platform. Four application programming interface (API) available with it, which makes it different from other VI service providers. These APIs are ‘XML-RPC’ and ‘Libvirt’ for local interaction, ‘EC2 (Query)’ APIs and the OpenNebula cloud

API (OCA) for remote operation purpose. It has compatibility for XEN, KVM, and VMware hypervisors. Also it has powerful security management, virtual networking compatibility with other open source products like open Vswitch, dynamic resource allocation and VM migration [11].

- **oVirt:** It is a virtualization management application and similar in concept with VSphere of VMware. ‘oVirt’ management interface can manage hardware nodes, storage and network resources, and deployment and monitoring of virtual machines running in data center. It can be built with KVM hypervisor and on the ‘RHEV-M’ management server released by RedHat, live migration of VM, user and group based security solutions are supported [12].
- **Apache cloud stack:** It is a project of Apache software foundation (ASF) open source software for deploying public and private infrastructure-as-a-service (IaaS) clouds. Its features include compatibility with hosts running XEN Server/XCP, KVM, and/or VMware ‘ESXi’ with VSphere. It has ‘Amazon S3/EC2’ compatible API. Network services like firewall, NAT, VPN are provided. Web based user interface for managing cloud is also provided by this platform [13].
- **Nimbus:** It is an open-source toolkit focused on providing IaaS capabilities. Nimbus provides Linux-based controller, EC2-compatible, modeling and accessing stateful resources with web services resource framework (WSRF) interfaces; XEN and KVM backend; interface to the Amazon EC2 public cloud and virtual networks. Table 1 provides a brief summary of capacities of different open source platforms [14].

2.1 Private Cloud Setup

In the present work, to build a private cloud infrastructure, the open source software ‘OpenNebula’ is used. The particular choice is due to its features like:

- (a) Deployment can be done with almost all Linux distributions.
- (b) Physical infrastructure management with monitoring and control facilities like creation and deletion of physical hosts and VM, setting and monitoring intervals, VM instance management etc.
- (c) Availability can be achieved with the support of hooks. They can notify error state of host so that the VMs on failed host can be redeployed on another host. Persistent backend keeps OpenNebula daemon safe even after crash for recovery.
- (d) User security management through ‘auth’ subsystem for managing users, additional support of *SSH*, and X.509.
- (e) It has a great deal of adaptability to manage any hardware and software combination along with integrating capacity with any other cloud product.

Figure 1 shows a private cloud setup using OpenNebula. One has to use the OpenNebula facilities to create and manage their own VM and virtual networks. It has following components:

Table 1 Open source platform for cloud setup

Cloud software (IaaS)	Virtual network	Interface to public cloud	Hypervisors compatibility	Security options	Dynamic resource allocation
Eucalyptus	Yes	EC2	XEN, KVM	Firewall filters (security groups), SSH and WS-security	Yes
OpenNebula	Yes	Amazon EC2	XEN, KVM, VMware	Authentication mechanisms and access control lists, firewall	Yes
oVirt	Yes		KVM	Supports Red Hat directory server, Free IPA or microsoft active directory for user and administrator authentication	Yes
Apache cloud stack	Yes	AWS EC2, S3	VMware, KVM, Citrix XEN server, and Citrix XEN cloud platform	Secure single sign-on, isolation of user for managing resources, firewall	Yes
Nimbus	Yes	EC2	XEN, KVM	Public key based authentication and authorization. Use of globus certificate credentials	Require integration with OpenNebula

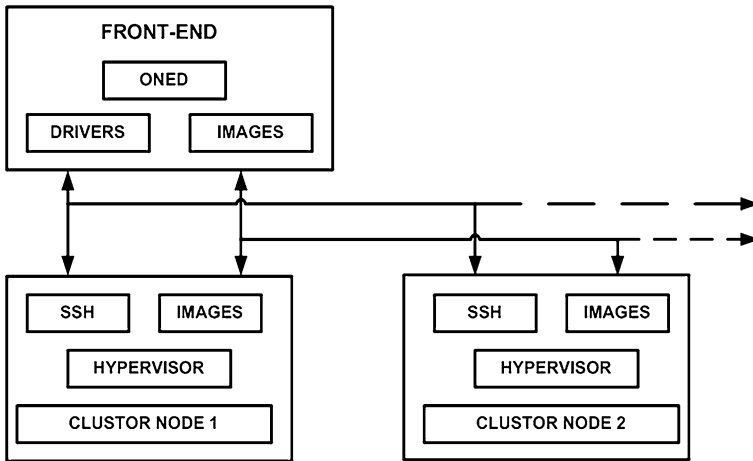


Fig. 1 OpenNebula cloud architecture [11]

- **Front-end:** OpenNebula installation is performed on this machine and it must have access to image repository and network connection with cluster nodes.
- **Cluster nodes:** It is the host on which hypervisor is installed. It provides the resources to the VM.
- **Image repository:** It is a storage medium that holds base images of the VM.
- **OpenNebula daemon:** It is the core service of the system. It managing service called 'oned'.
- **Drivers:** These are programs used by the core to interface with a specific cluster subsystem, i.e. a given hypervisor or storage file system.
- **Oneadmin:** It is the administrator of the private cloud that performs any operation on the VM, virtual networks, nodes or users.

For effective use and deployment of a private cloud, one must be familiar with the internal architecture of OpenNebula. The details of the OpenNebula internal architecture can be found in [11]. The OpenNebula internal architecture is made-up of three layers: the uppermost layer is of tools provided with Open Nebula. The middle layer is known as the core. It consist of VM, host, and virtual network management components. The last layer is of drivers. Drivers are required to plug-in different virtualization technologies into core. The following section describes the important parameters from internal architecture of OpenNebula which is essential to be known for installation of a cloud setup.

2.1.1 Tools

This layer represents various tools available with OpenNebula. CLI is provided with OpenNebula to manually control the VI under 'ONE daemon'. For this

Table 2 Commands available in OpenNebula

Command	Interpretation
onevm	To submit, control and monitor VM
onehost	To add delete and monitor hosts
Onevnet	To add delete and monitor virtual networks

purpose various commands are available with OpenNebula. Few of them are listed in Table 2.

The second tool is scheduler. The scheduler is an independent entity in the OpenNebula architecture. It uses the XML-RPC interface provided by ‘ONE daemon’ to invoke actions on VM. The scheduler distributed in the technology preview provides the following functionality:

- *User-driven consolidation:* The scheduler allocates VM to those cluster nodes which congregate the capacity constraints of the VM.
- *Matchmaking policy:* The scheduler sort out resources which do not meet the constraints of the VM. Then a grading process is made to select the best Host. This is accomplished by pertaining the ranking expression defined in the VM template. Third party tools can also be created using XML-RPC API or OpenNebula-client API.

2.1.2 OpenNebula Core

The core consists of a set of components to control and monitor VMs and hosts. The core performs its actions by invoking suitable drivers. The functional components of OpenNebula core are summarized in Table 3.

2.1.3 Drivers

OpenNebula has a set of pluggable modules to interact with specific middleware (e.g. virtualization hypervisor, file transfer mechanisms or information services). These modules are known Drivers. Few important drivers of OpenNebula are listed out in Table 4.

3 Cloud Setup Using OpenNebula

OpenNebula can be used for cloud setup in two modes as:

- (a) *System-wide mode:* In this mode OpenNebula-services like binaries and configuration files etc. are kept at standard locations under root file system.

Table 3 Functional units of OpenNebula core

Unit	Application
Request manager	To handle client request
VM manager	To manage and monitor VMs
Host manager	To manage and monitor physical resources
Database manager	Persistent storage for ONE data structure 'Sqlite3' is the default database used
Transfer manager	To manage transfer operations i.e. file, VM
Virtual network manager	To manage virtual networks

Table 4 Drivers and functionality

Drivers	Functionality
Information driver	To support information management
Transfer driver	To support transfer operation
VM driver	To manage VM functionality

(b) *Self-contained mode*: In this mode one can place OpenNebula distribution in self contained form at a particular directory.

Once acquainted with the capacities of OpenNebula platform, one can move for the setup of an own private cloud. In this work, the cloud setup is carried out in self-contained mode. For this purpose, the required frontend and cluster machine preferences are as:

- Frontend- Ubuntu-12.04 Desktop LTS with opennebula-3.8.3
- Cluster node- Ubuntu-12.04 Desktop LTS with KVM hypervisor installed.

As various Linux distributions are available, the selection of OS for setup depends on ease of working with the OS and the desired features. Ubuntu has a user friendly GUI and supports most of the virtualization technologies like '*Libvirt*' and KVM. The OpenNebula private cloud installation can be carried out with following steps:

1. Connection of frontend and cluster in LAN.
2. Creation of user account '*oneadmin*' and folder '*/srv/cloud/one*', at front end and cluster nodes.
3. Network file sharing in between frontend and cluster node: Ubuntu provides network file sharing facility by mounting the directory '*/srv/cloud/one*' that is created at frontend and at cluster node. At frontend side, add IP address of cluster and the options required for mounting a directory in '*/etc/exports*' file.
4. At cluster node side, add IP address of frontend and directory to be mounted in '*/etc/fstab*' file.

Table 5 Cloud installation and troubleshooting

Installation Setup	Errors	Solution
Network file sharing		
Front end side	Errors are not prominent for NFS	Set the parameters correctly to avoid errors
Cluster side	Bad syntax' error in <i>fstab</i> or <i>mtab</i> file <i>mount.nfs</i> is unable to mount the directory	It occurs due to improper syntax specially spaces This error occurs due to arguments in ' <i>fstab</i> ' or ' <i>mtab</i> ' file
OpenSSH setup		
Frontend side	SSH cannot connect to host port 22, connection refused	Make sure ' <i>sshd</i> ' configuration file must have port 22 line uncommented Still error persists use ' <i>netstat</i> ' command to check if TCP port 22 is blocked
	After copying public key Command prompt asks to enter a password of cluster machine	The hostname and the IP address of the cluster machine can be added in ' <i>/etc/hosts file</i> ' of frontend Check permission of ' <i>ssh</i> ' directory
Open nebula Installation on frontend machine	Cannot create database or user	To check particular MYSQL version installed on machine and issue above commands according to MYSQL version compatible syntax on frontend
Host creation hurdles	Onehost list shows error status for host	Password less connection between frontend and host must be lost try to SSH host from frontend Try to ping host from frontend
VM deployment error	Unable to open disk path: permission denied Unable to create domain	Add user ' <i>oneadmin</i> ' to disk group at cluster side Add user ' <i>oneadmin</i> ' to KVM group at cluster side

(continued)

Table 5 (continued)

Installation Setup	Errors	Solution
VM deployment error	Unknown OS type ' <i>hvm</i> '	Check if KVM module is loaded using command ' <i>lsmod grep kvm</i> '
GUI errors	<i>Novnc</i> icon active but does not show VM screen error [Failed to connect to server(1006)]	<p>Ensure correct loading of the module. If it is loaded then try '<i>modprobe kvm</i>'</p> <p>If same error persists then check if '<i>lsmod grep kvm</i>' returns '<i>kvm_intel</i>'. Still errors persist, then try '<i>modprobe kvm kvm-intel</i>' at cluster side</p> <p>kill existing proxy</p> <p>You can get proxy details using command '<i>ps aux grep websockify</i>'</p> <p>Relaunch proxy manually to get reason of error</p> <p>Tip: till you resolve the problem you can use virtual machine manager on host to see VM</p>

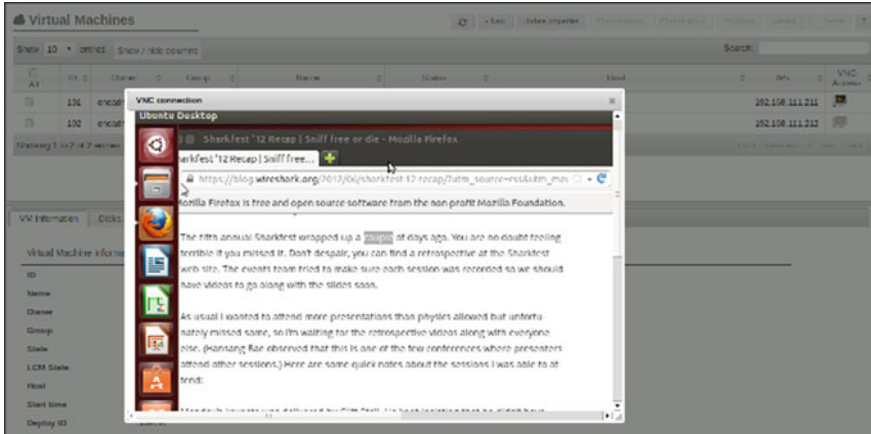


Fig. 2 VM creation

5. After steps 1–4, mount `/srv/cloud/one` using command line at cluster node side.
6. Establishment of communication between front end and clusters using `'openSSH'` facility secure shell access through key generation is performed. Generation of the key at frontend and copying it to cluster node side allows passwordless entry into the host machine.
7. Installation of dependency softwares and packages in frontend as per requirement of OpenNebula version.
8. Installation of OpenNebula in the frontend.
9. Configuration of hypervisor in cluster nodes—Installing bridge utilities and changing `'libvirtd.conf'` and `'qemu.conf'` files.
10. Installation of `'Sunstone'` graphical user interface [15].

Starting with the network file sharing, Table 5 provides possible errors with solutions to mitigate them for speedy cloud installation.

4 Results and Discussion

This paper provides the details of the available open source solutions for building private cloud. OpenNebula private cloud can be setup with the help of this paper by minimizing the possible errors with less time. In the present work, a P-IV system with 2-GB RAM at front end and 8-GB RAM at backend is used for cloud setup. Figure 2 depicts the functioning of cloud in terms of VM creation. It indicates that the OpenNebula cloud setup is successfully functioning. Internet can be accessed via the running VM for OpenNebula just like normal OS. The access to the outside world is provided to VM by `'Libvirt'` bridge.

Fig. 3 Network transmission and reception rate for VM

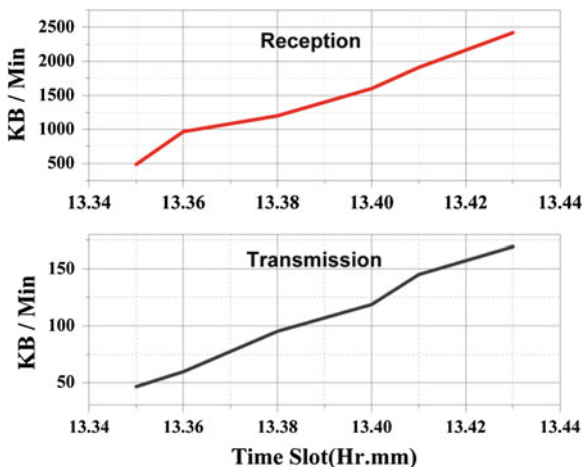
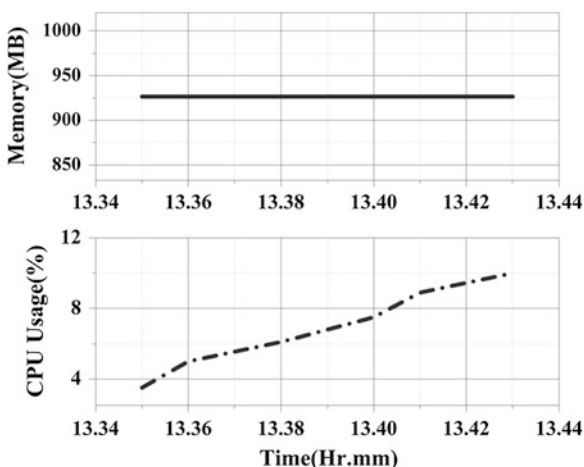


Fig. 4 CPU and memory usage of VM



The performance of the cloud setup has been verified by internet access through the VM. Initially the cloud setup is operated for a time slot of 10 min. Figure 3 shows the network transmission and reception for VM. The transmission and reception rate is relative with the corresponding increase or decrease of the network traffic. It indicates the successful deployment of cloud setup as the communication with the external world (internet) is carried out through VM.

Figure 4 shows the CPU usage by VM. It shows that CPU usage is varying with time slot. Also the maximum CPU usage is 12 % at its peak whereas memory of VM is constant as of allocated at the time of its creation. This indicates that host is dynamically adjusting the workload by using only 12 % CPU capacity and thus the remaining CPU and memory capacity can be made available to the other users (VM) on demand. This is the main difference between a cloud based system and a traditional system.

5 Conclusions

Parting the traditional information processing methodologies, cloud technology is growing rapidly. Hence it has got attention of researchers and academicians for its improvement. In this context, a private cloud set up is carried out using various open source tools. The present work highlights the requirements, procedure and tools for deployment of a private cloud. The practical solutions, based on the personal experience, to the various errors that arise during cloud set up are also listed out here. A KVM hypervisor is used here as it doesn't show any compatibility issues with the host. The performance of the cloud is verified by the cost functions such as transmission and reception rate, CPU usage and memory usage.

It shows that only 12 % of the CPU capacity is used during the peak traffic load and keeping rest of the capacity available for other users. The present attempt will certainly helpful to setup a cloud environment for research purpose with minimal efforts.

References

1. What is Cloud Technology (IEEE 802.16). www.freecloudinfo.com
2. Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., Brandic, I.: Cloud computing and emerging IT platforms: vision, hype, and reality for delivering computing as the 5th utility. *Future Gener. Comp. Syst.* **25**, 599–616 (2009)
3. Brown, E.: NIST issues cloud computing guidelines for managing security and privacy. *NIST Spec. Publ.* 800–144 (2012)
4. Voorsluys, W., Broberg, J., Buyya, R.: *Introduction to cloud computing: principles and paradigms*. Wiley Press, New York (2011)
5. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. *J. Internet Serv. Appl.* **1**(1), 7–18 (2010)
6. Buyya, R., Calheiros, R.N., Li, X.: Autonomic cloud computing: open challenges and architectural elements. In: *Proceedings of 3rd International Conference of Emerging Applications of Information Technology (EAIT 2012)*, Kolkata, India. IEEE Press, USA (2012)
7. XEN project Governance followed by individuals for developing and using Xen products. <http://www.xenproject.org/>
8. Hwang, J., Zeng, S., Wu, K., Wood, T.: A component-based performance comparison of four hypervisors. In: *13th IFIP/IEEE International Symposium on Integrated Network Management (IM) Technical Session* (2013)
9. Sotomayor, B., Montero, R.S., Lorente, I.M., Foster, I.: Virtual infrastructure management in private and hybrid clouds. *IEEE Internet Comput.* **13**(5), 14–22 (2009)
10. Eucalyptus-products and services. <http://www.eucalyptus.com>
11. Official documentation on open nebula cloud framework. www.opennebula.org/documentation
12. oVirt 3.2 release notes. http://www.ovirt.org/OVirt_3.2_release_notes
13. Apache cloud stack. <http://cloudstack.apache.org/>
14. Nimbus infrastructure implementation. <http://www.nimbusproject.org/doc/nimbus/>
15. Cloud-b-lab: Tutorials on cloud computing, OPENNEBULA 3.4.1 in Ubuntu 12.04 Precises—single machine installation for learning and testing purpose

Modified Activity of Scout Bee in ABC for Global Optimization

Kedar Nath Das and Biplab Chaudhur

Abstract Artificial Bee Colony (ABC) algorithm is a bio-inspired technique motivated by the intelligent foraging behavior of honey bee swarm. ABC mainly depends on the activity of employed bees, onlooker bees and scout bees. It is a practice that during simulation, if no further improvement in the population is found within an allowable number of cycles, the employed bee becomes scout and reinitializes the population by its standard equation. But there is a chance of losing the best individuals achieved so far. In this paper, a modification in scout bee activity is proposed, with an insertion of a modified Quadratic Approximation namely qABC. The effectiveness of the proposed qABC over most recent variants of ABC is analyzed through a set of Benchmark problems. The experimental confirms that qABC outperforms its other variants.

Keywords Artificial bee colony algorithm · Swarm intelligence · Bio-inspired algorithm

1 Introduction

Optimization plays an important role in many real-world situations. Unfortunately, when the problem in hand is highly complex, traditional methods become handicapped. As an alternate paradigm, bio-inspired optimization techniques provide at least the near optimal solution. In recent years, Artificial Bee Colony Optimization

K. N. Das (✉) · B. Chaudhur

Department of Mathematics, National Institute of Technology, Silchar 788010, India
e-mail: kedar.iitr@gmail.com

B. Chaudhur

e-mail: biplabs2008@gmail.com

(ABC) has become popular to solve optimization problems, as it requires very less parameters to fine tune.

By simulating the foraging and intelligent behavior of real honey bees swarms, D. Karaboga, in 2005 [1, 2] pioneered the ABC. It is much competent to Genetic Algorithm (GA), Differential Evolution (DE), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO) etc.

Since 2005, it has become a challenge to the researchers to increase the efficiency of ABC in a faster and better way. As a result, many variants of ABC algorithm came out during last few years. Gao et al. [3] proposed a modified artificial bee colony (MABC) algorithm and ABC/best, which improve the exploitation around the best solution of previous iteration. Gao et al. [4, 5] propose two improve version of ABC. A new search technique called ABC/best/1 and ABC/rand/1 is used to speed up the convergence speed. Luo et al. [6] proposed a converge-onlookers ABC (COABC) algorithm in order to improve the exploitation, based on the update times of employed bee (UTEb) food source and claimed that this algorithm performs remarkably better in terms of solution quality and convergence rate. Akay and Karaboga [7] proposed a modified ABC algorithm. They implemented two new search techniques to ABC on frequency and magnitude of the perturbation to improve the convergence rate. Li et al. [8] proposed an improved ABC (I-ABC) algorithm in which the best-so-far solution, inertia weight and acceleration coefficients are introduced to modify the search mechanism. Chaotic ABC (CABC) algorithm is proposed by Atlas [9], to improve the convergence speed of ABC. In their paper, many chaotic maps for parameters adaption are introduced to original ABC.

It is seen that, ABC and its variants achieve the near optimal solution. But most of the cases it struck in some local optima, due to the lack of proper balance between the exploration and the exploitation in the search space. In order to improve further the efficiency of ABC, a robust and simple hybridized ABC algorithm is proposed in this paper.

Rest of this paper is organized as follows. The next Sect. 2 presents the outline of ABC. Section 3 describes the component used in the proposed hybrid system. The proposed hybrid technique is explained in Sect. 4. In Sect. 5, experimental setup and result analysis are presented. The conclusion is drawn in Sect. 6.

2 Artificial Bee Colony Algorithm

Artificial Bee Colony (ABC) algorithm starts with a set of possible solutions called swarm. Swarm consists of employed bees, onlooker bees and scouts. Once all employed bees complete the search process, they share the nectar amount information of the food sources and their location information with the onlooker bees in the hive through waggle dance. An onlooker bee accesses the nectar information and location taken from all employed bees and chooses a food source with a

probability proportionate to their nectar amount. The major steps of the ABC algorithm are as follows.

In the first step, randomly an initial population of SN individuals are generated. Each individual $x_i (i = 1, 2, \dots, SN)$ is a D dimensional vector. Here D refers to the problem dimension. Then the each solution is generated by the following equation.

$$x_{i,j} = xL_j + \mu(xU_j - xL_j) \tag{1}$$

where $i = 1, 2, \dots, SN, j = 1, 2, \dots, D$ and μ is a random number within $[0,1]$. xL_j and xU_j are the lower and upper bound of the variable j . The fitness of each food source is calculated by Eq. (2).

$$fitness_i = \begin{cases} \frac{1}{1+f_i} & \text{if } f_i \geq 0 \\ 1 + abs(f_i) & \text{if } f_i < 0 \end{cases} \tag{2}$$

where f_i is the objective function value and $fitness_i$ is the fitness value of the i th solution.

At the second step, each employed bee of the food source x_i generates a new food source v_i in the neighbourhood of its present position by Eq. (3).

$$v_{i,j} = x_{i,j} + \phi_{i,j}(x_{i,j} - x_{k,j}) \tag{3}$$

where $k \in 1, 2, \dots, SN$ and $j \in 1, 2, \dots, D$ are randomly chosen indexes, provided $k \neq i$, the random number $\phi_{i,j} \in [-1,1]$.

After generating new food source (solution) v_i , a greedy selection technique is used. If the fitness value of v_i is better than the corresponding x_i then replace x_i by v_i , otherwise keep the food source x_i unaltered.

After getting all information from employed bee, each onlooker bee selects a food source with a probability based on its fitness by Eq. (4).

$$p_i = \frac{fitness_i}{\sum_{n=1}^{SN} fitness_n} \tag{4}$$

where $fitness_i$ is the fitness value of the i th position, proportional to the nectar amount of the food source and SN is the number of food sources equals to the number of employed bees and onlooker bees. Clearly, better the fitness of an individual, higher the probability of getting it selected for the next population.

In the onlooker bee phase, an artificial onlooker bee chooses its food source depending on Eq. (4). A new food source is generated by Eq. (3) in the neighbourhood of x_i and its fitness is evaluated. A same greedy selection takes place.

At the last stage, if the position of the food source is not improved further over a predefined number of trials (*limit*), then the food source is treated as an abandoned one by its employed bee and the corresponding employed bee becomes a scout bee. Now the scout bee starts searching for new food source and replaces the abandoned food source by Eq. (1).

In the ABC algorithm, while onlookers and employed bees carry out the exploitation process in the search space, the scouts control the exploration process. For robustness, exploration and exploitation processes must be carried out together.

3 Component in the Proposed Hybrid System

For better understanding the proposed method presented in Sect. 4, it is essential to know the working principle of few of modified quadratic approximation as discussed below.

3.1 Modified Quadratic Approximation

A Random Search Technique (RST) for Global optimization was first introduced by Mohan and Shankar [10]. Later it is modified as Quadratic Approximation (QA) [11]. In QA, one 'child' is generated (as shown in Fig. 1) from the selection of three parents C_1 , C_2 and C_3 as follows in Eq. (5), where C_1 is considered as best fit individuals and C_2 , C_3 are random. The *child* generated is as follows:

$$child = 0.5 \left(\frac{\sum (C_2^2 - C_3^2) f(C_1)}{\sum (C_2 - C_3) f(C_1)} \right) \quad (5)$$

where $f(C_1)$, $f(C_2)$ and $f(C_3)$ are the fitness function values at C_1 , C_2 and C_3 respectively.

In this paper, a Modified Quadratic Approximation (abbreviated as mQA) is proposed. Unlike QA, mQA picks C_1 , C_2 and C_3 randomly with a condition that, $C_1 \neq C_2 \neq C_3$. Rest mechanism of mQA remains the same as QA.

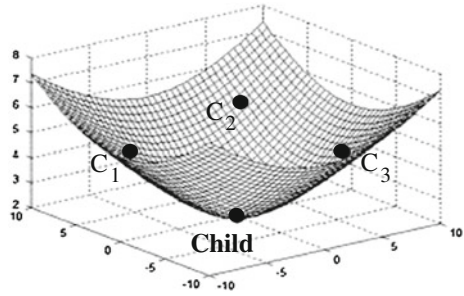
4 Proposed qABC Algorithm

Storn et al. [12] proposed a Differential Evolution (DE) for solving optimization problem. It consists of three major operators namely mutation, crossover and selection. By considering two variants of mutation operator, authors suggested two types of DEs as follows.

$$DE/best/1: \quad V_i = X_{best} + F(X_{r_1} - X_{r_2}) \quad (6)$$

$$DE/best/2: \quad V_i = X_{best} + F(X_{r_1} - X_{r_2}) + F(X_{r_3} - X_{r_4}) \quad (7)$$

Fig. 1 Formation of child by QA



where $i = 1, 2, \dots, SN$ and r_1, r_2, r_3 and r_4 are mutually different random integer number in the range of $1, 2, \dots, SN$. $F < 1.0$ is a positive real number, known as scaling factor or amplification factor. It controls the rate at which the population evolves. Based on this two variant of DE, Gao et al. [4] proposed two different strategies of ABC algorithm given in (8) and (9).

$$\text{ABC/best/1: } v_{i,j} = x_{best,j} + \phi_{i,j}(x_{r_1,j} - x_{r_2,j}), r_1 \neq r_2 \tag{8}$$

$$\text{ABC/best/2: } v_{i,j} = x_{best,j} + \phi_{i,j}(x_{r_1,j} - x_{r_2,j}) + \phi_{i,j}(x_{r_3,j} - x_{r_4,j}); r_1 \neq r_2 \ \& \ r_3 \neq r_4 \tag{9}$$

where $i = 1, 2, \dots, SN$ and $r_1, r_2, r_3, r_4 \in 1, 2, \dots, SN$ different from the base i . $x_{best,j}$ is the best individual in the current population and $j \in 1, 2, \dots, D$ is a randomly chosen index, $\phi_{i,j}$ is a random number in the range $[-1,1]$. $x_{k,j}$ is a random individual in the population and hence the solution search by the Eq. (3) is random enough for exploration. At the same time it is also poor at exploitation.

In [6], the solution search equation is

$$v_{i,j} = x_{i,j} + rand[-1, 1] * fit(x_b) * (x_{i,j} - x_{b,j}) \tag{10}$$

where x_b is the global best food source $x_{b,j}$ stands for the j th element of the global best food source. In the onlooker stage the update process is done by

$$v_{b,j} = x_{b,j} + \psi * (x_{b,j} - x_{k,j}) \tag{11}$$

Where ψ is a uniform random number in $[0,C]$, C is non-negative constant. Each onlooker updates the global best food source by (11).

The ABC/best/1 and COABC algorithm performs well on finding better solution of the objective function. However, after certain cycle of simulation an abandoned case arises. In such instance scout replaces the food source arbitrarily by using Eq. (1). Most of the authors in the literature including [4, 6] use this concept. Therefore there is a chance of losing the best values obtained so far. As a result ABC might not be strong enough to balance the exploration and exploitation behaviour.

Keeping in view the above, in this paper, the searching behaviour of scout bee is slightly modified by new exploration technique. Particularly if a food source is not improved further within the predefined limit, the source is assumed as abandoned by its employed bees and each associated bee becomes a scout to search for a new food source randomly. The scout replace the abandoned food source by the Eq. (5), or if the Eq. (5) fails to create new food source then the abandoned food source is replaced by the Eq. (1), which helps to avoid of getting trapped in local optima many a time. By this process the exploration and exploitation together hopefully improve the performance.

4.1 qABC Algorithm

1. Set Maximum Cycle Number (MCN), Population Size, Limit, crossover probability p_c and Scaling parameter.
2. Initialize the food sources, and evaluate the population, $trail_i = 0$, ($i = 1, \dots, SN$), $cycle = 1$. **REPEAT**
3. Search the new food source for employed bees by (1) and evaluate it.
4. Apply a greedy selection process and select the better solution between the new food source and the old one.
5. If solution does not improve, $trail_i = trail_i + 1$; otherwise $trail_i = 0$.
6. Apply Tournament selection to choose a food source for onlooker bees.
7. Search the new food source for onlooker bees by (3) and evaluate it.
8. Apply a greedy selection and select the better solution between the new food source and the old one.
9. If solution does not improve, $trail_i = trail_i + 1$; otherwise $trail_i = 0$.
10. If $\max(trail_i) > limit$, replace this food source with a new child produced by modified quadratic approximation scheme or by (1) and set $trail_i = 0$. Memorize the best solution achieved so far.
11. Apply elitism.
12. Apply Laplace Crossover (not for $cycle = 1$).
13. $cycle = cycle + 1$. **UNTIL**
14. $Cycle = MCN$.

5 Experimental Setup and Result Analysis

To compare the efficiency and the accuracy of the proposed qABC, 18 well-known benchmark functions with different dimensions are considered from [4] and [6]. The authors solved them with five different scenarios by considering five different experiments, which have been repeated in this present study too. For a fair comparison all the parameter settings have been kept the same as [6] expect

Table 1 Results obtained by ABC, COABC and qABC algorithms

Function	Min	Gen	Algorithm	Mean	SD	Time(s)
Matyas	0	1000	ABC	6.03E-07	3.64E-07	174.61
			COABC	4.40E-07	4.63E-07	47.97
			qABC	1.38E-12	1.37E-12	0.033
Booth	0	1000	ABC	1.68E-17	1.38E-17	142.05
			COABC	6.19E-23	2.06E-22	59.45
			qABC	9.40E-38	9.40E-38	0.0326667
Michalewicz	-1.80	1000	ABC	-1.80	6.83E-16	192.17
			COABC	-1.80	6.83E-16	98.41
			qABC	-1.80	4.16E-18	0.0093333
6 Hump Camel Back	-1.03	1000	ABC	-1.03	7.20E-17	164.78
			COABC	-1.03	1.76E-16	81.97
			qABC	-1.03	5.91E-14	0.0083333
Goldstein-Price	3	1000	ABC	3.00	1.47E-03	178.67
			COABC	3.00	3.21E-06	68.66
			qABC	3.00	2.95E-18	0.0133333

Table 2 Results obtained by ABC, COABC with UTEB=1 and qABC algorithms

Dimension	Algorithm	Mean	SD
10	ABC	8.11E-17	1.14E-17
	COABC	2.66E-49	1.18E-48
	qABC	4.71E-53	9.07E-53
30	ABC	5.78E-16	1.01E-16
	COABC	8.98E-18	1.35E-17
	qABC	5.26E-55	1.14E-54
50	ABC	1.28E-11	1.46E-11
	COABC	4.44E-11	5.14E-11
	qABC	5.03E-34	8.02E-34

population size, which is kept fixed at 10*D. For f9 the crossover probability p_c is 0.0017 where as for the rest of the problems the p_c is 0.978. The location parameter a is fixed as 0 and scaling parameter b ranges from 0.15 to 0.65. The algorithm is coded in C++ and simulated in Linux platform. The mean and standard deviations (SDs) are recorded for 30 and 20 independent runs as explained below.

In order to analyse the performance of placing all onlookers on the solution, tournament selection is used. The same set of five 2-dimensional benchmark functions is collected from [6]. The experiment shows that the onlookers carry out neighbourhood search and update continuously, the optimum solution can be obtained by all three algorithms. Comparative results are reported in Table 1, it contains the mean, SD and time of best objective function value for 30 independent runs. For each function the best result is reported in bold face figure. From Table 1, it is observed that expect 6 Hump Camel Back function, qABC gives better result for the rest.

Table 3 Results obtained by COABC and qABC algorithms with UTEB = 1, 2, 3 respectively

D	Algorithm	UTEB = 1		UTEB = 2		UTEB = 3	
		Mean	SD	Mean	SD	Mean	SD
10	COABC	2.66E-49	1.18E-48	1.88E-102	6.35E-102	8.62E-155	3.67E-154
	qABC	3.54E-97	3.54E-97	5.06E-109	5.06E-109	3.94E-136	3.94E-136
30	COABC	8.98E-18	1.35E-17	4.08E-34	5.66E-34	4.31E-50	1.45E-49
	qABC	1.77E-53	3.53E-53	2.18E-55	9.15E-55	3.78E-58	4.28E-58
50	COABC	4.44E-11	5.14E-11	1.02E-20	1.23E-20	6.09E-30	1.27E-29
	qABC	1.67E-27	4.99E-27	7.61E-36	2.29E-36	3.07E-36	3.21E-36

Table 4 Results obtained by ABC, qABC and COABC algorithms with UTEB = 1, 2, 3 respectively

Fun	MCN		ABC	COABC			qABC
				UTEB = 1	UTEB = 2	UTEB = 3	
f1	1000	Mean	3.41E-17	5.13E-53	2.45E-107	1.42E-152	1.00E-123
		SD	8.03E-18	1.62E-52	3.37E-107	4.44E-152	0
		Time(s)	186.54	55.04	72.36	92.18	0.154
f2	2000	Mean	5.43E-16	3.23E-19	1.53E-35	1.46E-50	1.39E-54
		SD	1.12E-16	4.69E-19	2.09E-35	6.19E-50	2.27E-54
		Time(s)	323.30	107.77	155.42	201.92	6.0915
f3	1000	Mean	6.03E-07	4.45E-07	1.33E-07	1.87E-09	1.55E-43
		SD	3.64E-07	4.63E-07	2.14E-07	2.72E-09	2.19E-43
		Time(s)	174.61	47.97	77.58	87.67	0.069
f4	2000	Mean	2.40E-01	1.68	2.20E-01	8.00E-02	1.44E-01
		SD	4.60E-01	2.08	5.70E-01	1.00E-01	9.61E-02
		Time(s)	392.00	159.81	238.38	241.00	11.00
f5	1000	Mean	-4.69	-4.69	-4.69	-4.69	-4.65
		SD	5.61E-11	1.04E-15	8.40E-16	7.62E-16	5.73E-02
		Time(s)	216.61	111.33	127.45	133.47	0.089
f6	2000	Mean	3.13E-14	1.03E-11	3.41E-14	0	2.03E-15
		SD	3.90E-14	1.32E-11	3.87E-14	0	8.31E-16
		Time(s)	366.09	129.89	160.70	225.16	3.08
f7	1000	Mean	-1.03	-1.03	-1.03	-1.03	-1.03
		SD	7.20E-17	1.76E-16	2.16E-16	2.10E-16	2.17E-19
		Time(s)	164.78	81.97	91.83	107.64	0.0335
f8	2000	Mean	3.93E-13	1.17E-08	5.24E-14	3.57E-14	6.01E-16
		SD	1.55E-13	8.07E-09	1.35E-14	3.57E-15	1.99E-16
		Time(s)	426.83	222.74	254.38	353.30	4.56

In the second and third experiment, the sphere functions of three different dimensions are used. For the 2nd experiment UTEB is taken as 1 and for the 3rd experiments UTEB is taken as 1, 2, 3. Comparative results of these experiments are presented in Tables 2 and 3 respectively. UTEB in these tables represent the update times of employed bee. The best mean and SD obtained for this function is

Table 5 qABC performance comparison of ABC,COABC and the state-of-art algorithms in [4] with UTEB = 3

Fun	D	MCN		ABC	ABC/best/ 2	ABC/best/ 1	COABC	qABC
f1	30	1000	Mean	6.99E-10	4.37E-22	1.57E-27	4.47E-27	3.75E-31
			SD	5.91E-10	2.14E-22	1.14E-27	5.99E-27	1.13E-31
	60	2000	Mean	1.94E-09	1.57E-20	2.42E-25	1.28E-26	1.71E-27
			SD	8.33E-10	4.90E-21	1.09E-25	6.07E-27	2.12E-27
f2	30	1000	Mean	2.36E-06	2.73E-12	3.45E-15	1.82E-14	6.10E-23
			SD	8.32E-07	4.61E-13	8.79E-16	1.10E-14	1.39E-23
	60	2000	Mean	8.30E-06	2.14E-11	5.94E-14	3.30E-14	5.86E-22
			SD	8.93E-07	4.39E-12	1.28E-14	5.47E-15	2.64E-22
f4	30	1000	Mean	0	0	0	0	3.69E-31
			SD	0	0	0	0	4.07E-32
	60	2000	Mean	0	0	0	0	7.43E-28
			SD	0	0	0	0	7.49E-28
f6	30	1000	Mean	6.63E-03	0	0	9.09E-14	0
			SD	1.71E-02	0	0	5.49E-14	0
	60	2000	Mean	3.03E-01	0	0	4.43E-13	0
			SD	4.53E-01	0	0	2.48E-13	0
f8	30	1000	Mean	8.73E-09	4.47E-08	4.23E-11	0	3.41E-18
			SD	1.47E-08	1.05E-07	2.16E-11	0	1.35E-18
	60	2000	Mean	4.46E-09	2.18E-10	0	0	2.95E-17
			SD	6.68E-09	4.33E-10	0	0	3.45E-17
f9	30	1000	Mean	2.05E + 2	0	0	0	1.41E + 02
			SD	1.63E + 2	0	0	0	0
	60	2000	Mean	6.93E + 2	3.67E-11	3.74E-11	0	1.05E + 03
			SD	1.39E + 2	3.17E-12	2.59E-12	0	0
f10	30	1000	Mean	1.02E-05	1.89E-11	1.26E-13	9.40E-13	5.04E-16
			SD	4.15E-06	4.75E-12	3.48E-14	7.58E-13	1.92E-16
	60	2000	Mean	2.05E-05	1.08E-10	3.40E-13	1.19E-12	2.98E-15
			SD	5.54E-06	4.97E-12	6.35E-14	6.73E-13	1.01E-15
f13	30	1000	Mean	1.38E-04	7.34E-10	1.32E-14	2.22E-08	2.22E-18
			SD	6.38E-05	1.01E-09	5.04E-15	2.30E-08	1.57E-18
	60	2000	Mean	1.32E-03	2.68E-08	4.46E-13	2.05E-06	1.64E-17
			SD	1.21E-03	1.84E-08	4.13E-13	5.81E-06	9.60E-18
f19	4	2000	Mean	1.71E-01	2.41E-02	2.14E-03	9.35E-03	9.78E-07
			SD	6.94E-02	2.18E-02	2.21E-03	4.64E-03	6.91E-07
f22	3	1000	Mean	2.07E-03	3.55E-05	2.57E-12	0	0
			SD	3.91E-03	5.63E-05	1.12E-11	0	0
	4	2000	Mean	4.06E-02	1.35E-03	2.51E-06	0	2.71E-20
			SD	4.84E-02	3.27E-03	7.54E-06	0	0

reported in bold face. It is observed from Tables 2 and 3 that after updating employed bee, the solution gets closer to the global optimum. Thus it helps to enhance the capability of escaping from local optima. From Table 3, qABC

algorithms outperforms all variants of COABC, in terms of best objective function values except $UTEB = 3$ for dimension 10.

In the fourth experiment, the performance of qABC is widely analysed. For this a set of eight different kinds of benchmark functions [6] including sphere function, are used. The set consists of unimodal, multimodal, regular, irregular, separable, non-separable and multi-dimensional functions. The value of common parameters used in this paper for all algorithms are same as [6]. The UTEB are set to 1, 2 and 3. The comparative result of qABC is reported in Table 4. It contains the mean, SD and time to achieve the best objective function value for 20 independent runs. For each function the best result is reported in bold face figure. From Table 4 it is found that qABC outperforms ABC and all the variants of COABC, except f1 and f5.

In the last experiment, to further verify the performance 10 benchmark functions of two different dimensions are selected from [4]. The parameter of qABC is the same as that in [4]. For a fair comparison with [6], the limit is set to $0.6 * SN * D$ as in [6], $UTEB = 3$ and the experiment is repeated 30 times independently. The results obtained by qABC, ABC, state-of-art algorithms in [4] and COABC are listed in Table 5. The comparison results show that qABC is superior to ABC, ABC/best/2, ABC/best/1 and COABC except f9.

6 Conclusion

In this paper, a robust hybridization of a modified Quadratic Approximation with Artificial Bee Colony Optimization (namely qABC) is proposed. A slight modification in the activity of the scout bee is incorporated to enhance the better exploring the search space. The numerical results of qABC have been compared with the recently reported ABC in the literature. The experimental results demonstrate that qABC algorithm can well balance the exploration and exploitation. It outperforms its competitors like COABC and its recent variants. Hence it is more reliable. In the other hand, it also impacts less standard deviations and consuming less computational time that confirms the stability of its mechanism.

References

1. Karaboga, D.: An idea based on honeybee swarm for numerical optimization, Technical Report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, (2005)
2. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *J. Global Optim.* **39**, 459–471 (2007)
3. Gao, W., Liu, S.: A modified artificial bee colony algorithm. *Comput. Oper. Res.* **39**, 687–697 (2012)
4. Gao, W., Liu, S., Huang, L.: A global best artificial bee colony algorithm for global optimization. *J. Comput. Appl. Math.* **236**, 2741–2753 (2012)

5. Gao, W., Liu, S.: Improved artificial bee colony algorithm for global optimization. *Inf. Process. Lett.* **111**, 871–882 (2011)
6. Luo, J., Wang, Q., Xiao, X.: A modified artificial bee colony based on converge-onlookers approach for global optimization. *Appl. Math. Comput.* <http://dx.doi.org/10.1016/j.amc.2013.04.001>(2013)
7. Akay, B., Karaboga, D.: A modified artificial bee colony algorithm for real-parameter optimization. *Inf. Sci.* **192**, 120–142 (2012)
8. Li, G., Niu, P., Xiao, X.: Development and investigation of efficient artificial bee colony algorithm for numerical function optimization. *Appl. Soft Comput.* **12**(1), 320–332 (2012)
9. Alatas, B.: Chaotic bee colony algorithm for global numerical optimization. *Experts Syst. Appl.* **37**, 5682–5687 (2010)
10. Mohan, C., Shanker, Kusum: A Random Search Technique for Global Optimization Based on Quadratic Approximation. *Asia Pac. J. Oper. Res.* **11**, 93–101 (1994)
11. Deep, K., Das, K.N.: Quadratic approximation based hybrid genetic algorithm for function optimization. *AMC, Elsevier* **203**, 86–98 (2008)
12. Storn, R., Price, K.: Differential evolution- a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **23**, 689–694 (2010)

A Multistage Model for Defect Prediction of Software Development Life Cycle Using Fuzzy Logic

Harikesh Bahadur Yadav and Dilip Kumar Yadav

Abstract In this paper, a multistage model for software defect density indicator is proposed using the top most reliability relevant metrics and Fuzzy Inference System (FIS). Prediction of defect in each phase of software development life cycle (SDLC) is desirable for effective decision-support and trade-off analysis during early development phases. The predictive accuracy of proposed model is validated using nine real software projects data. Validation results are satisfactory.

Keywords Fuzzy logic · Software defect · Software defect indicator · Software reliability · Software metrics · Fuzzy inference system

1 Introduction

People are working under direct or indirect influence of software now a day. Safety-critical software should be highly reliable as failure of this software may cause injury or death to human beings. Apart from safety-critical systems, software has become the integral part of most of complex applications. Thus it is very important to ensure that the underlying software will perform its intended functions correctly. Hence, software reliability analysis has become a major research area. The reliability of a software system depends upon number of remaining defects. Therefore, it is necessary to predict the software defect for each phase of SDLC.

H. B. Yadav (✉) · D. K. Yadav
Department of Computer Applications, National Institute of Technology,
Jamshedpur 831014, India
e-mail: yadavaharikesh@gmail.com

D. K. Yadav
e-mail: dkyadav1@gmail.com

Software reliability is an important factor of software quality. Software reliability, which is defined by IEEE “The probability that software will not cause the failure of a system for a specified time under specified conditions” [1]. Reliability is requested to be assured in almost all safety-critical system. Software reliability model were designed to quantify the likelihood of software failure [2, 3]. The termination of the ability of a functional unit to perform its required function called failure [1]. And defect can defined as a product anomaly, examples include such things as omissions and imperfections found during early life cycle phases.

Software reliability has played an important role in each early software development phases [4]. Various software reliability models have been developed in the last three decade [3]. The model can alienated into two categories first one Software reliability prediction model and second one Software Reliability estimation model. Majority of model based on failure data. Since failure data are not available in the early phases of SDLC. Therefore, qualitative values of reliability relevant software metrics are used in this model. These qualitative values are associated with imprecision and uncertainty hence fuzzy logic is used in the proposed model.

The rest of the paper is organized as follows: Sect. 2 describes the related work, proposed model is explained in Sect. 3 and result is discussed in Sect. 4.

2 Related Work

The existing software reliability models are neither universally predicting software reliability nor generally tractable to users. Lots of study in the past has been made for software reliability estimation and prediction by various models [3, 5]. Gaffney and Davis [6], Gaffney and Pietrolewicz [7] proposed a phase based model for predicting reliability by using the fault statistics. Air force’s Rome Laboratory developed a model for early software reliability prediction [8, 9]. The model is mainly based on the software requirement specification and data collected by the organization. Agresti and Evanco proposed a model to predict defect density on the basis of process and product characteristics [10]. Smidts suggested reliability prediction model based on the requirements change request during SDLC. There are many factors which are affecting the software reliability in SDLC. Thirty two factors are identified which have impact on the software reliability in initial phase [11]. In another study Li and Smidt [12] identified thirty software metrics which influence the software reliability.

A casual model for defect prediction with Bayesian Net is developed by Fenton et al. [13]. The main feature is that it does not require detailed domain knowledge and it combines both qualitative and quantitative data. Pandey and Goyal [14] have proposed an early fault prediction model using process maturity and software metrics. They have considered the fuzzy profiles of various metrics in different scale and have not explained the criteria used for developing these fuzzy profiles. The method level metrics are used in most of the fault prediction models [15].

Recently Yadav et al. [16] proposed a software defect prediction model in which they had considered only the uncertainty associated over the assessment of software size metric and three metrics of requirement analysis phase. Radjenovic [17] reported that the object oriented and process metrics are more successful in finding the faults compared to traditional size and complexity metrics. Can et al. [18] suggested a model for software defect prediction in which they used benefit of non-linear computing capability of Support Vector Machine and parameters optimization capability of Particle Swarm Optimization. However, the software reliability is a function of number of residual defect in the software and reliability relevant metrics play an important role in software defect prediction.

Therefore, a casual defects prediction model is developed using the reliability relevant metrics which will be able to predict the number of defect present at the end of each phase of SDLC.

3 Proposed Model

The model, which is a FIS, is presented in the Fig 1. In the Requirement Phase (RP) top three reliability metrics such as Requirement Specification Change Request (RCR), Requirement Fault Density (RFD), and Requirement Inspection and Walk through (RIW) are used as input to the fuzzy inference system. The output of RP (Requirement phase Defect density indicator (RPDDI)) taken as an input along with Cyclomatic Complexity (CC) and Design Review Effectiveness (DRE). Design phase Defect density indicator (DPDDI) along with programmer capability (PC) and process maturity (PM) are taken as input in design phase.

Similarly output of Coding phase (coding phase defect density indicator (CPDDI)) and two other testing phase metrics such as Staff Experience (SE), Quality of Documented Test Cases (QDT) are taken as input to the testing phase.

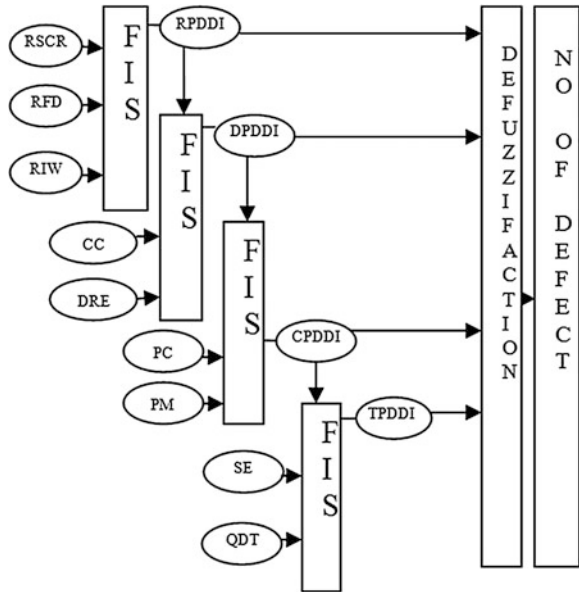
The following steps are involved in this proposed model

- Step 1: Define membership function of each input and output variable
- Step 2: Defining Fuzzy Rule Base
- Step 3: Fuzzy Inference and Defuzzification

3.1 Define Membership Function of Each Input and Output Variable

In the proposed model, defect of each phase of SDLC is predicted based on the measures present in first four phases (i.e. requirements analysis, design, coding and testing) of SDLC. Therefore, proposed model leverages the most relevant top measures [12] from first four phases of SDLC.

Fig. 1 Proposed model architecture



There are many method of membership value assignment such as rank ordering, intuition, inference etc. [19–21]. In the proposed model membership function of all the input and output metrics are defined with the help of domain expert. In this model triangular and trapezoidal membership are considered for representing the linguistic state (L, M, H) for input metrics and (VL, L, M, H, VH) for output metric of fuzzy variable [19–21]. The range of all input and output metrics are taken 1.

Fuzzy profiles for individual software metrics are illustrated in Figs. 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14.

3.2 Defining Fuzzy Rule Base

Fuzzy rule is defined in the form of IF_THEN conditional statement Where IF part of the rule is known as antecedent and THEN part is consequent. In the proposed model Consequent value is assessed by the domain expert. Fuzzy Rules for first four phase of SDLC are illustrated in Figs. 15, 16, 17, 18.

Fig. 2 Requirements specification change request

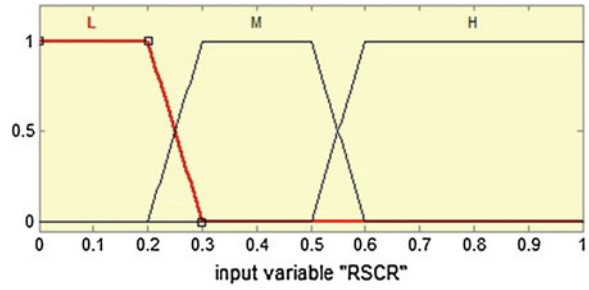


Fig. 3 Requirement fault density

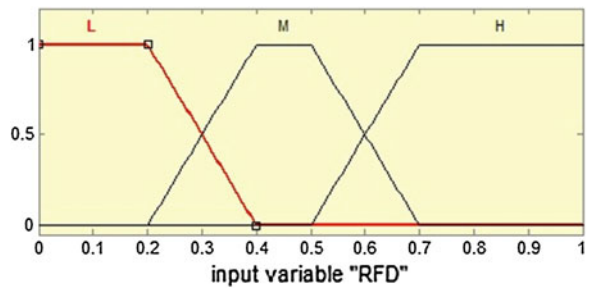


Fig. 4 Requirement inspection and walk through

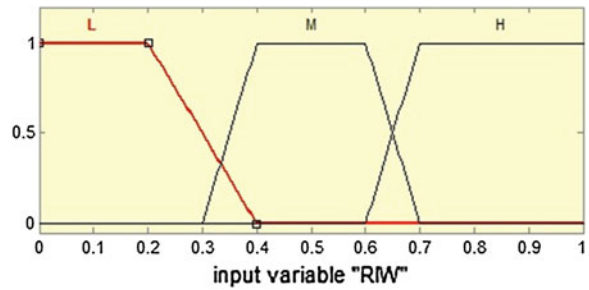


Fig. 5 Requirement phase defect density indicator

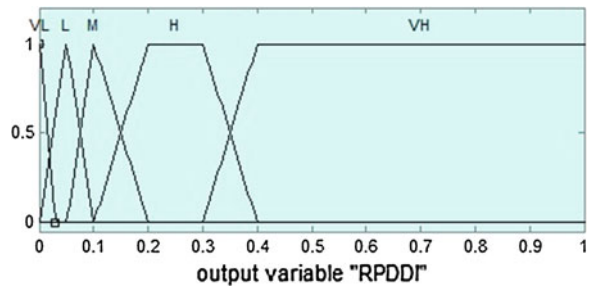


Fig. 6 Cyclomatic complexity

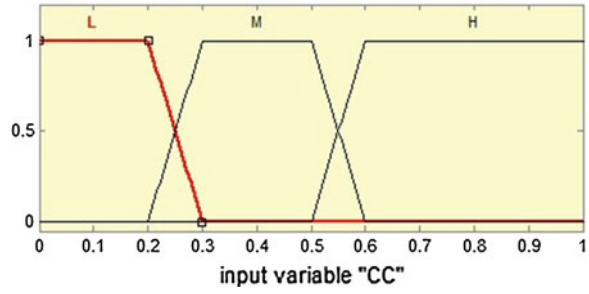


Fig. 7 Design review effectiveness

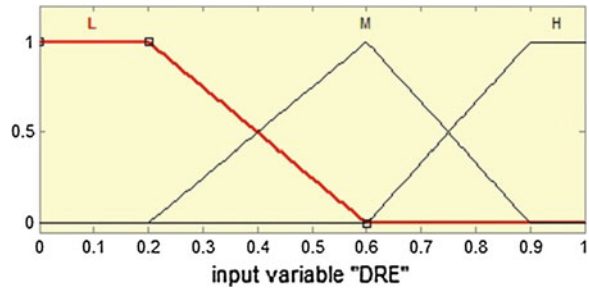


Fig. 8 Design phase defect density indicator

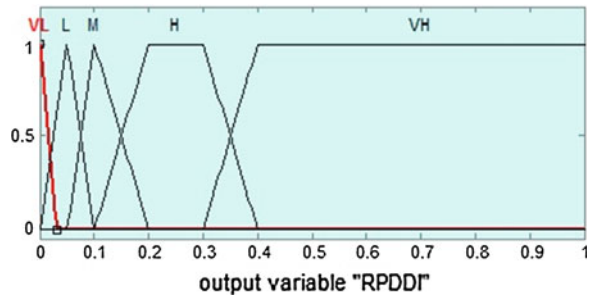


Fig. 9 Programmer capability

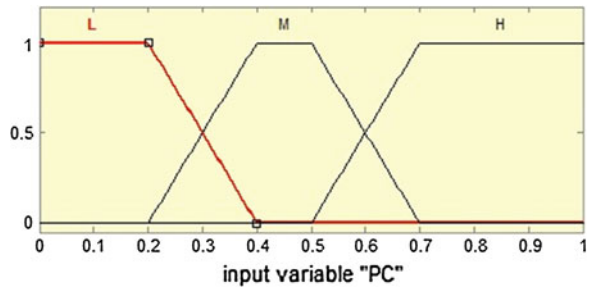


Fig. 10 Process maturity

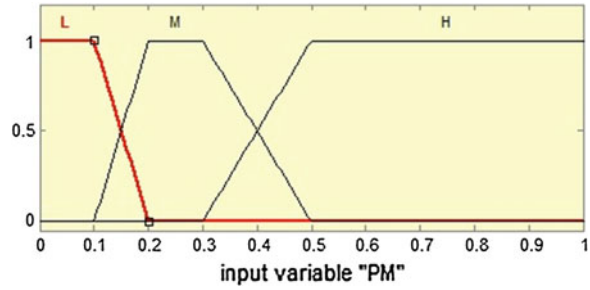


Fig. 11 Coding phase defect density indicator

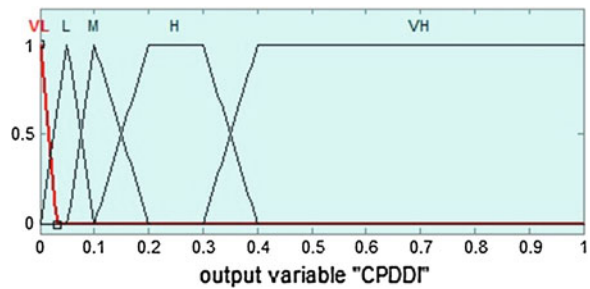


Fig. 12 Staff experience

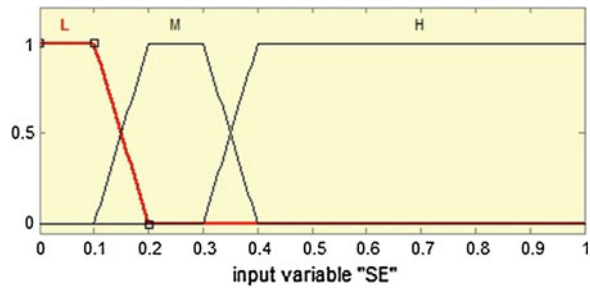


Fig. 13 Quality of documented test cases

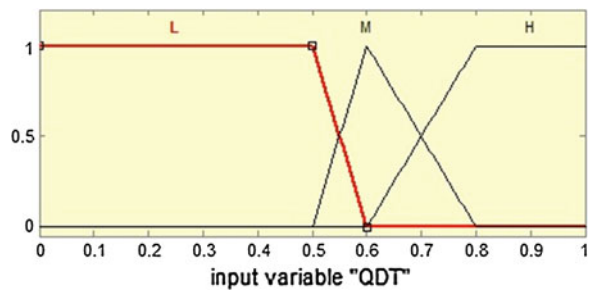
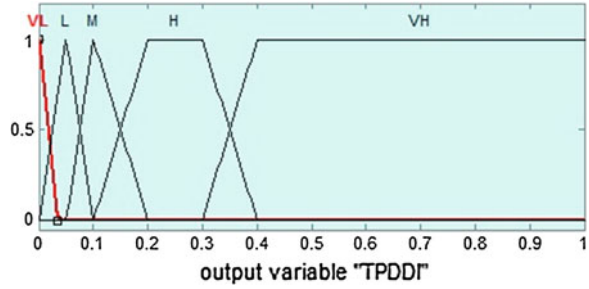


Fig. 14 Testing phase defect density indicator



- 1. If (RSCR is L) and (RFD is L) and (RMW is L) then (RPDDI is L) (1)
- 2. If (RSCR is L) and (RFD is L) and (RMW is M) then (RPDDI is VL) (1)
-
- 26. If (RSCR is H) and (RFD is H) and (RMW is M) then (RPDDI is H) (1)
- 27. If (RSCR is H) and (RFD is H) and (RMW is H) then (RPDDI is M) (1)

Fig. 15 Requirements phase fuzzy rule

- 1. If (CC is L) and (DRE is L) and (RPDDI is VL) then (DPDDI is VL) (1)
- 2. If (CC is L) and (DRE is L) and (RPDDI is L) then (DPDDI is VL) (1)
-
- 44. If (CC is H) and (DRE is H) and (RPDDI is H) then (DPDDI is VH) (1)
- 45. If (CC is H) and (DRE is H) and (RPDDI is VH) then (DPDDI is VH) (1)

Fig. 16 Design phase fuzzy rule

- 1. If (PC is L) and (PM is L) and (DPDDI is VL) then (CPDDI is VL) (1)
- 2. If (PC is L) and (PM is L) and (DPDDI is L) then (CPDDI is L) (1)
-
- 44. If (PC is H) and (PM is H) and (DPDDI is H) then (CPDDI is L) (1)
- 45. If (PC is H) and (PM is H) and (DPDDI is VH) then (CPDDI is H) (1)

Fig. 17 Coding phase fuzzy rule

- 1. If (SE is L) and (QDT is L) and (CPDDI is VL) then (TPDDI is VL) (1)
- 2. If (SE is L) and (QDT is L) and (CPDDI is L) then (TPDDI is L) (1)
-
- 44. If (SE is H) and (QDT is H) and (CPDDI is H) then (TPDDI is M) (1)
- 45. If (SE is H) and (QDT is H) and (CPDDI is VH) then (TPDDI is H) (1)

Fig. 18 Testing phase fuzzy rule

Table 1 Defect density indicator of various phases

Project no./ Fenton Project no.	Size (KLOC)	RPDDI	DPDDI	CPDDI	TPDDI	Predicted defect density	Actual defect density	Fenton (2008) defect density
1/29	11	0.00667	0.00805	0.00826	0.00829	8.29	8.27	10.54
2/3	53.9	0.0067	0.02550	0.00936	0.00847	8.47	3.88	4.71
3/21	22	0.05	0.00719	0.00813	0.00827	8.27	8.91	11.77
4/22	44	0.05	0.00719	0.00813	0.00827	8.27	4.18	11.38
5/7	21	0.00667	0.00805	0.00826	0.00829	8.29	9.71	12.47
6/13	49.1	0.05	0.05	0.00719	0.00813	8.13	2.63	10.5
7/8	5.8	0.00667	0.0255	0.00936	0.00847	8.47	9.14	8.3
8/9	2.5	0.00667	0.00805	0.00826	0.00829	8.29	6.8	22.8
9/10	4.8	0.00667	0.00805	0.00826	0.00829	8.29	6.04	4.2

3.3 Fuzzy Inference and Defuzzification

Fuzzy inference engine evaluate and combined result of each rule. Since fuzzy inference engine maps fuzzy set into a fuzzy set, but in many application crisp number must be obtained at the output of a fuzzy system. The defuzzification method such as centroid, max–min, and bisection etc. maps fuzzy set into crisp number [19–21]. Centroid method of defuzzification is used in this model.

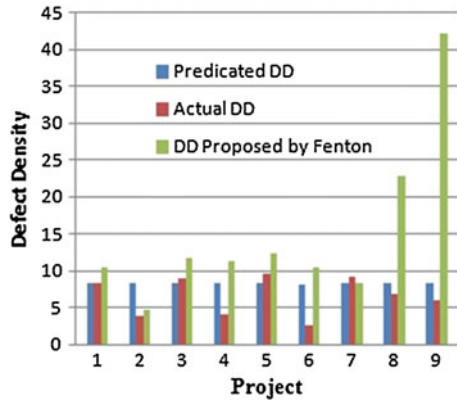
4 Result and Discussion

The proposed model is validated using actual defect density and prediction result of Fenton et al. [13]. Fenton proposed a Bayesian Net model for predicting the software defects for the same software projects. Table 1 show the actual defect density, predicated defect density indicator and defect density proposed by Fenton.

The prediction results are compared with the earlier work and it is clear that requirement phase is most important than all the other phases because, chances of getting fault is more than all the other phases. In literature, it is also shown that the 50 to 60 % of total defects is occurred due to ambiguous analysis at requirement stages, 15 to 30 % at design phase and 10 to 20 % defect occur at implementation phase [22].

From the Fig. 19 it is clear that predicated defect after testing phase is closer to actual defect than the defect predicated by Fenton et al.

Fig. 19 Compression of various projects



5 Conclusions

The proposed model considers only reliability relevant software metrics of each phase of SDLC. In this paper, a fuzzy logic based model is proposed for predicting software defects density indicator at each phase of SDLC. The predicted defect density for nine software projects are found very near to the actual defects detected during testing. The predicted defect density indicators are very helpful to analyze the defects severity in different artifacts of SDLC of a software project. This provides a guideline to the software manager for early identification of cost overruns, schedules mismatch, software development process issues, software resource allocation and release decision making etc.

References

1. ANSI/IEEE 729-1983, IEEE Standard Glossary of Software Engineering Terminology
2. Lyu, M.R.: Handbook of Software Reliability Engineering. McGraw-Hill, New York (1995)
3. Musa, J.D., Iannino, A., Okumoto, K.: Software Reliability: Measurement, Prediction, Application. McGraw-Hill Publishers, New York (1987)
4. Kaner, C.: Software engineering metrics: what do they measure and how do we know. In: 10th International Software Metrics Symposium 2004
5. Pham, H.: System Software Reliability. Reliability Engineering Series. Springer, London (2006)
6. Gaffney, J.E., Davis, C.F.: An approach to estimating software errors and availability. In: Proceedings of 11th Minnow Brook Workshop on Software Reliability (1988)
7. Gaffney, J.E., Pietrolewicz, J.: An automated model for software early error prediction (SWEEP). In: Proceedings of 13th Minnow Brook Workshop on Software Reliability (1990)
8. Methodology for Software Reliability Prediction and Assessment, Technical Report RL-TR-92-95, vol. 1-2. Rome Laboratory (1992)
9. Methodology for Software Prediction, RADAC-TR-87-171, USAF Systems Command (1987)

10. Agresti, W.W., Evanco, W.M.: Projecting software defects form analyzing Ada design. *IEEE Trans. Softw. Eng.* **18**(11), 988–997 (1992)
11. Zhang, X., Pham, H.: An analysis of factors affecting software reliability. *J. Syst. Softw.* **50**(1), 43–56 (2000)
12. Li, M., Smidts, C.: A ranking of software engineering measures based on expert opinion. *IEEE Trans. Softw. Eng.* **29**(9), 811–824 (2003)
13. Fenton, N., Neil, M., Marsh, W., Hearty, P., Radlinski, L., Krause, P.: On the effectiveness of early life cycle defect prediction with Bayesian Nets. *Empirical Softw. Eng.* **13**, 499–537 (2008)
14. Pandey, A.K., Goyal, N.K.: A fuzzy model for early software fault prediction using process maturity and software metrics. *Int. J. Electr. Eng.* **1**(2), 239–245 (2009)
15. Catal, Cagatay: Software fault prediction: a literature review and current trends. *Expert Syst. Appl.* **38**(4), 4626–4636 (2011)
16. Yadav, D.K., Charurvedi, S.K., Mishra, R.B.: Early software defects prediction using fuzzy logic. *Int. J. Perform. Eng.* **8**(4), 399–408 (2012)
17. Radjenovic, D., et al.: Software fault prediction metrics: a systematic literature review. *Inf. Softw. Technol.* **55**(8), 1397–1418 (2013)
18. Can, H., Jianchun, X., Ruide, Z., Juelong, L., Qiliang, Y., Liqiang, X.: A new model for software defect prediction using particle swarm optimization and support vector machine. *Control and Decision Conference (CCDC)*, 2013 25th Chinese, pp. 4106, 4110. 25–27 May 2013
19. Ross, T.J.: *Fuzzy logic with engineering applications*, 2nd edn. Wiley, London (2004)
20. Zadeh, L.A.: Knowledge representation in fuzzy logic. *IEEE Trans. Knowl. Data Eng.* **1**, 89–100 (1989)
21. Yadav, D.K., Chaturavedi, S.K., Mishra, R.B.: Forecasting time-between-failures of software using fuzzy time series approach. In: *IEEE Proceeding of North American Fuzzy Information Processing Society (NAFIPS)*, Berkley, USA, pp 1–8. 6–8 Aug 2012
22. Suma, V., Gopalakrishnan Nair T.R.: Defect Management Strategies in Software Development. In: Strangio, M.A. (ed.) *Recent Advances in Technologies* (2009)

Generalized Second-Order Duality for a Class of Nondifferentiable Continuous Programming Problems

Iqbal Husain and Santosh Kumar Srivastava

Abstract The research in this paper is organized in two parts. In the first part, a generalized second-order dual is formulated for a class of continuous programming problems with square roots of positive semi-definite quadratic forms, hence it is nondifferentiable. Under second-order pseudoinvexity and second-order quasi-invexity, various duality theorems are proved for this pair of dual nondifferentiable continuous programming problems. Lastly, it is pointed out that our duality results can be regarded as dynamic generalizations of those for a nondifferentiable nonlinear programming problem, already treated in the literature. In the second part, a generalized second-order dual is formulated for a continuous programming problem with support functions. Duality results similar to those of the first part are validated for this class of problems. Clearly the results of this part are more general in sense that a support function extends a square root of a positive semidefinite quadratic form.

Keywords Generalized second-order duality · Second-order pseudoinvexity · Second-order quasi-invexity · Continuous programming problems · Nonlinear programming problems

1 Introduction

A number of researchers have studied second-order duality in mathematical programming. A second-order dual to a nonlinear programming problem was first formulated by Mangasarian [1]. Subsequently, Mond [2] established various

I. Husain (✉) · S. K. Srivastava
Department of Mathematics, Jaypee University of Engineering and Technology, Guna,
Madhya Pradesh, India
e-mail: ihusain11@yahoo.com

S. K. Srivastava
e-mail: santosh_00147@rediffmail.com

duality theorems under a condition which is called second-order convexity, which is much simpler than that used by Mangasarian [1]. Mond and Weir [3] reformulated the second-order and higher-order duals to validate duality results. It is found that second-order dual to a mathematical programming problem offers a tighter bound and hence enjoys computational advantage over a first-order dual.

In the spirit of Mangasarian [1], Chen [4] formulated second-order dual for a constrained variational problem and established various duality results under an involved invexity-like assumption. Later, Husain et al. [5], studied Mond-Weir type second-order duality for the problem of [4], by introducing continuous-time version of second-order invexity and generalized second-order invexity, validated usual duality results. Subsequently, Husain and Masoodi [6] presented Wolf type duality while Husain and Srivastava [7] formulated Mond-Weir type dual for a class of continuous programming problems containing square root of a quadratic form to relax to the assumption of second-order pseudoinvexity and second-order quasi-invexity.

In this paper, in the first part, in order to combine dual formulations of [6] and [7], a generalized dual to the nondifferentiable continuous programming problem of Chandra et al. [8] is constructed and a number of duality results are proved under appropriate generalized second-order invexity assumption. A relation between our duality results and those of a nondifferentiable nonlinear programming problem is pointed out through natural boundary value variational problems.

Recently, Husain and Masoodi [9] presented Wolfe type second-order duality and Husain and Srivastava [10] presented Mond-Weir type second-order duality for a class of continuous programming problems containing support functions which are somewhat more general than the square root of certain positive semi-definite quadratic form. In the second part, a generalized second-order dual to the problem of [9, 10] is formulated and duality results are established under generalized second-order invexity conditions. Problems with natural boundary conditions are also constructed. It is also pointed out that our duality results are dynamic generalization of those of nonlinear programming problems with support functions treated by Husain et al. [11].

The popularity of the above mathematical programming problems seems to stem from the fact that, even though the objective functions and or constraint functions are differentiable, a simple formulation of the duals may be given. Nondifferentiable mathematical programming deals with much more general kinds of functions by using generalized sub-differentials and quasi-differentials.

2 Related Pre-requisites

Let $I = [a, b]$ be a real interval, $\phi: I \times R^n \times R^n \rightarrow R$ and $\psi: I \times R^n \times R^n \rightarrow R^m$ be twice continuously differentiable functions. In order to consider $\phi(t, x(t), \dot{x}(t))$, where $x: I \rightarrow R^n$ is differentiable with derivative \dot{x} denoted by ϕ_x and $\phi_{\dot{x}}$, the first order derivative of ϕ with respect to $x(t)$ and $\dot{x}(t)$ respectively. Denote by ϕ_{xx} , the $n \times n$ Hessian matrix of ϕ and ψ_x the $m \times n$ Jacobian

matrices respectively. Designate by X , the space of piecewise smooth functions $x:I \rightarrow R^n$ with the norm $\|x\| = \|x\|_\infty + \|Dx\|_\infty$ where the differentiation operator D is given by $u = Dx \Leftrightarrow x(t) = \int_a^t u(s)ds$, thus $\frac{d}{dt} = D$ except at discontinuities. Now, we incorporate the following definitions which are required in the subsequent analysis:

Definition 1 (Second-order Pseudoinvex): If the functional $\int_I \phi(t, x, \dot{x})dt$ satisfies $\int_I \{\eta^T \phi_x + (D\eta)^T \phi_{\dot{x}} + \eta^T G \beta(t)\}dt \geq 0 \Rightarrow \int_I \phi(t, x, \dot{x})dt \geq \int_I \{\phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2}\beta(t)^T G \beta(t)\}dt$, then $\int_I \phi(t, x, \dot{x})dt$ is said to be second-order pseudoinvex with respect to η .

Definition 2 (Second-order strictly pseudoinvex): If the functional $\int_I \phi(t, x, \dot{x})dt$ satisfies $\int_I \{\eta^T \phi_x + (D\eta)^T \phi_{\dot{x}} + \eta^T G \beta(t)\}dt \geq 0 \Rightarrow \int_I \phi(t, x, \dot{x})dt > \int_I \{\phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2}\beta(t)^T G \beta(t)\}dt$, then $\int_I \phi(t, x, \dot{x})dt$ is said to be second-order strictly pseudoinvex with respect to η .

Definition 3 (Second-order Quasi-invex): If the functional $\int_I \phi(t, x, \dot{x})dt$ satisfies $\int_I \phi(t, x, \dot{x})dt \leq \int_I \{\phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2}\beta(t)^T G \beta(t)\}dt \Rightarrow \int_I \{\eta^T \phi_x + (D\eta)^T \phi_{\dot{x}} + \eta^T G(t) \beta(t)\}dt \leq 0$, then $\int_I \phi(t, x, \dot{x})dt$ is said to be second-order quasi-invex with respect to η . If ϕ does not depend explicitly on t , then the above definitions reduce to those given in [2] for static cases.

3 Generalized Second-Order Duality for a Continuous Programming Problem with Square Root Functions

In this part, we consider the following class of non-differentiable continuous programming problem studied in [8]:

$$\begin{aligned} \text{(QP) : Minimize } & \int_I \left(f(t, x, \dot{x}) + (x(t)^T B(t) x(t))^{1/2} \right) dt \\ \text{subject to } & x(a) = 0 = x(b), \\ & g(t, x, \dot{x}) \leq 0, t \in I \end{aligned}$$

where (i) $I = [a, b]$ is a real interval. (ii) $f: I \times R^n \times R^n \rightarrow R$ and $g: I \times R^n \times R^n \rightarrow R^m$ are twice continuously differentiable function with respect

to its argument $x(t)$ and $\dot{x}(t)$. (iii) $x : I \rightarrow R^n$ is four times differentiable with respect to t and these derivatives are defined by $\dot{x}, \ddot{x}, \overset{\dots}{x}$ and $\overset{\dots}{\overset{\dots}{x}}$. (iv) $B(t)$ is a positive semi-definite $n \times n$ matrix with $B(\cdot)$ continuous on I .

Chandra et al. [8] derived the following optimality conditions for (QP):

Proposition 1 If (QP) attains an optimal solution at $x = \bar{x} \in X$, then there exist Lagrange multipliers $\tau \in R$ and piecewise smooth $y : I \rightarrow R^m$ not both zero and also piecewise smooth $\bar{\omega} : I \rightarrow R^n$ satisfying for all $t \in I$,

$$\begin{aligned} \tau [f_x(t, \bar{x}, \dot{\bar{x}}) + B(t)\bar{\omega}(t)] + y(t)^T g_x(t, \bar{x}, \dot{\bar{x}}) \\ - D(\tau f_{\dot{x}}(t, \bar{x}, \dot{\bar{x}}) + y(t)^T g_{\dot{x}}) = 0, \quad t \in I, \\ y(t)^T g(t, \bar{x}, \dot{\bar{x}}) = 0, \bar{x}(t)^T B(t)\bar{\omega}(t) = (\bar{x}(t)^T B(t)\bar{x}(t))^{1/2}, \\ \bar{\omega}(t)^T B(t)\bar{\omega}(t) \leq 1, \quad t \in I, \end{aligned}$$

The Fritz John necessary optimality conditions given in Proposition 1.1.1 become the Karush–Kuhn–Tucker type optimality conditions if $\tau = 1$. For $\tau = 1$, it suffices that the following Slater’s conditions [8] holds:

$g(t, \bar{x}, \dot{\bar{x}}) + g_x(t, \bar{x}, \dot{\bar{x}})v(t) + g_{\dot{x}}(t, \bar{x}, \dot{\bar{x}})\dot{v}(t) > 0$ for some $v(t) \in X$ and for all $t \in I$.

Husain et al. [5] formulated the following Wolf type second-order dual and established duality results for the pair of problems (QP) and (W-QD) under the second-order pseudoinvexity of $\int_I \{f(t, \cdot, \cdot) + (\cdot)^T B(t)w(t) + y(t)^T g(t, \cdot, \cdot)\} dt$ with respect to η .

$$(W - QD) : \text{Maximize } \int_I \left(f(t, u, \dot{u}) + u(t)^T B(t)w(t) + y(t)^T g(t, u, \dot{u}) - \frac{1}{2}\beta(t)^T L\beta(t) \right) dt$$

subject to $u(a) = 0 = u(b),$

$$f_u(t, u, \dot{u}) + w(t)^T B(t) + y(t)^T g_u(t, u, \dot{u}) - D(f_{\dot{u}} + y(t)^T g_{\dot{u}}) + L\beta(t) = 0, t \in I,$$

$$w(t)^T B(t)w(t) \leq 1, y(t) \geq 0, t \in I,$$

where $L = f_{uu}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_u - 2D[f_{u\dot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\dot{u}}]$
 $+ D^2[f_{u\dot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\dot{u}}] - D^3[f_{\dot{u}\dot{u}}(t, u, \dot{u}) + (y(t)^T g_{\dot{u}}(t, u, \dot{u}))_{\dot{u}}].$

Recently, Husain and Srivastava [7] investigated Mond-Weir type duality by constructing the following dual to (QP) under $\int_I (f(t, \cdot, \cdot) + (\cdot)^T B(t)w(t)) dt$ for all $w(t) \in R^n$ is second-order pseudoinvex and $\int_I y(t)^T g(t, \dots) dt$ is second-order quasi-invex with respect to the same η .

$$\begin{aligned}
 (M - WQD) : & \text{Maximize } \int_I \left\{ f(t, u, \dot{u}) + u(t)^T B(t)w(t) - \frac{1}{2} \beta(t)^T F \beta(t) \right\} dt \\
 & \text{subject to } \quad u(a) = 0 = u(b), \\
 & f_u(t, u, \dot{u}) + B(t)w(t) + y(t)^T g_u(t, u, \dot{u}) - D(f_{\dot{u}}(t, u, \dot{u}) + y(t)^T g_{\dot{u}}(t, u, \dot{u})) + (F + K) \beta(t) = 0, \quad t \in I, \\
 & \int_I \left(y(t)^T g(t, u, \dot{u}) - \frac{1}{2} \beta(t)^T K \beta(t) \right) dt \geq 0, w(t)^T B(t)w(t) \leq 1, y(t) \geq 0, \quad t \in I, \\
 & \text{where } \quad F(t, u, \dot{u}, \ddot{u}) = f_{uu} - 2Df_{u\dot{u}} + D^2f_{\dot{u}\dot{u}} - D^3f_{\dot{u}\ddot{u}}, \quad t \in I \\
 & \quad \text{and} \\
 & K(t, u, \dot{u}, \ddot{u}) = y(t)^T g_{uu} - 2Dy(t)^T g_{u\dot{u}} + D^2y(t)^T g_{\dot{u}\dot{u}} - D^3y(t)g_{\dot{u}\ddot{u}}, \quad t \in I.
 \end{aligned}$$

4 Generalized Second-Order Duality

Here, we formulate the following generalized second-order dual (QD) to (QP):

$$\begin{aligned}
 \text{Dual(QD)} : & \text{Maximize } \int_I \left[f(t, u, \dot{u}) + u(t)^T B(t)\omega(t) + \sum_{i \in I_0} y^i(t)g^i(t, u, \dot{u}) - \frac{1}{2} \beta(t)^T H^0 \beta(t) \right] dt \\
 & \text{subject to } \quad u(a) = 0 = u(b), \\
 & f_u(t, u, \dot{u}) + B(t)w(t) + y(t)^T g_u(t, u, \dot{u}) - D(f_{\dot{u}}(t, u, \dot{u}) + y(t)^T g_{\dot{u}}(t, u, \dot{u})) + H \beta(t) = 0, \quad t \in I, \\
 & \int_I \left(\sum_{i \in I_\alpha} y^i(t)g^i(t, u, \dot{u}) - \frac{1}{2} \beta(t)^T G^\alpha \beta(t) \right) dt \geq 0, \quad t \in I, \quad \alpha = 1, 2, 3, \dots, r, \\
 & w(t)^T B(t)w(t) \leq 1, y(t) \geq 0, \quad t \in I, \\
 \text{where } & (i) I_\alpha \subseteq M = \{1, 2, \dots, m\}, \quad \alpha = 0, 1, \dots, r \text{ with } \bigcap_{\alpha=0}^r I_\alpha = M \text{ and } I_\alpha \cap I_\beta = \emptyset, \text{ if } \alpha \neq \beta. \\
 & (ii) H^0 = f_{uu}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t)g^i(t, u, \dot{u}))_{uu} - 2D \left(f_{u\dot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t)g^i(t, u, \dot{u}))_{u\dot{u}} \right) \\
 & \quad + D^2 \left(f_{\dot{u}\dot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t)g^i_{\dot{u}\dot{u}}(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) - D^3 \left(f_{\dot{u}\ddot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t)g^i_{\dot{u}\ddot{u}}(t, u, \dot{u}))_{\dot{u}\ddot{u}} \right), \\
 & (iii) H = f_{uu}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_u - 2D(f_{u\dot{u}}(t, u, \dot{u}) + (y(t)^T g_{u\dot{u}}(t, u, \dot{u}))_{\dot{u}}) \\
 & \quad + D^2(f_{\dot{u}\dot{u}}(t, u, \dot{u}) + (y(t)^T g_{\dot{u}\dot{u}}(t, u, \dot{u}))_{\dot{u}\dot{u}}) - D^3(f_{\dot{u}\ddot{u}}(t, u, \dot{u}) + (y(t)^T g_{\dot{u}\ddot{u}}(t, u, \dot{u}))_{\dot{u}\ddot{u}}), \\
 & (iv) G^\alpha = \sum_{i \in I_\alpha} (y^i(t)g^i_u(t, u, \dot{u}))_u - 2D \left(\sum_{i \in I_\alpha} (y^i(t)g^i_{u\dot{u}}(t, u, \dot{u}))_{\dot{u}} \right) + D^2 \left(\sum_{i \in I_\alpha} (y^i(t)g^i_{\dot{u}\dot{u}}(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) \\
 & \quad - D^3 \left(\sum_{i \in I_\alpha} (y^i(t)g^i_{\dot{u}\ddot{u}}(t, u, \dot{u}))_{\dot{u}\ddot{u}} \right), \quad \alpha = 1, 2, \dots, r, \text{ and} \\
 & (v) G = \sum_{i \in M - I_0} (y^i(t)g^i_u(t, u, \dot{u}))_u - 2D \left(\sum_{i \in M - I_0} (y^i(t)g^i_{u\dot{u}}(t, u, \dot{u}))_{\dot{u}} \right) + D^2 \left(\sum_{i \in M - I_0} (y^i(t)g^i_{\dot{u}\dot{u}}(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) \\
 & \quad - D^3 \left(\sum_{i \in M - I_0} (y^i(t)g^i_{\dot{u}\ddot{u}}(t, u, \dot{u}))_{\dot{u}\ddot{u}} \right).
 \end{aligned}$$

Theorem 1 (Weak Duality): Let $\bar{x}(t)$ be feasible for (QP) and $(u(t), y(t), w(t), \beta(t))$ be feasible for (QD). If for all feasible $(x(t), u(t), y(t), w(t), \beta(t))$, $\int_I \{f(t, \dots) + \sum_{i \in I_0} y^i(t)g^i(t, \dots) + (\cdot)^T B(t)w(t)\}dt$ is second-order pseudoinvex and $\sum_{i \in I_x} \int_I y^i(t)g^i(t, \dots)dt$, $\alpha = 1, 2, \dots, r$ is second-order quasi-invex with respect to the same η , then

$$\inf.(\text{QP}) \geq \sup.(\text{QD}).$$

Theorem 2 (Strong Duality): If $\bar{x}(t)$ is a optimal solution of (QP) and normal [8], then there exist piecewise smooth $y : I \rightarrow R^m$ and $w : I \rightarrow R^n$ such that $(\bar{x}(t), y(t), w(t), \beta(t) = 0)$ is feasible for (QD) and the corresponding values of (QP) and (QD) are equal. If for all feasible $(\bar{x}(t), y(t), w(t), \beta(t))$, $\int_I \{f(t, \dots) + \sum_{i \in I_0} y^i(t)g^i(t, \dots) + (\cdot)^T B(t)w(t)\}dt$ is second-order pseudoinvex and $\int_I \sum_{i \in I_x} y^i(t)g^i(t, \dots)dt$, $\alpha = 1, 2, \dots, r$ is second-order quasi-invex, then $(\bar{x}(t), y(t), w(t), \beta(t))$ is an optimal solution of (QD).

Theorem 3 (Converse Duality): Let $(x(t), y(t), w(t), \beta(t))$ be an optimal solution of (QD) at which

(A₁) for all $\alpha = 1, 2, \dots, r$ either

$$(a) \int_I \beta(t)^T \left(G_\alpha + \sum_{i \in I_x} (y^i g_x^i) \right) \beta(t) dt > 0 \text{ and } \int_I \beta(t)^T \left(\sum_{i \in I_x} (y^i g_x^i) \right) dt \geq 0,$$

or

$$(b) \int_I \beta(t)^T \left(G^\alpha + \sum_{i \in I_x} (y^i g_x^i) \right) \beta(t) dt < 0 \text{ and } \int_I \beta(t)^T \left(\sum_{i \in I_x} (y^i g_x^i) \right) dt \leq 0,$$

(A₂) the vectors $H_j^0, G_j^\alpha, \alpha = 1, 2, \dots, r$ and $j = 1, 2, \dots, n$ are linearly independent, where H_j^0 is the j th row of the matrix H^0 and G_j^α is the j th row of the matrix G^α and

(A₃) the vectors $\sum_{i \in I_x} (y^i(t)g_x^i) - D \sum_{i \in I_x} (y^i(t)g_x^i), \alpha = 1, 2, \dots, r$ are linearly independent. If for all $(x(t), y(t), w(t), \beta(t))$, $\int_I \{f(t, \dots) + (\cdot)^T B(t)w(t) + \sum_{i \in I_0} y^i(t)g^i(t, \dots)\}dt$ is second order pseudoinvex and $\int_I \sum_{i \in I_x} y^i(t)g^i(t, \dots)dt, \alpha = 1, 2, \dots, r$ is second-order quasi-invex, then $x(t)$ is an optimal solution of (QP).

5 Special Cases

If $I_0 = M$, then (QD) becomes (W-QD) and from theorem 1-3, it follows that it is a second-order dual to (QP), if $\int_I \{f(t, \cdot, \cdot) + y^T(t)g(t, \cdot, \cdot) + (\cdot)^T B(t)w(t)\} dt$ is a second-order pseudoinvex.

If I_0 is empty set and $I_\alpha = M$ for some $\alpha \in \{1, 2, \dots, r\}$, then (QD) reduces to (M-WQD) which is a second-order dual to (QP), if $\int_I \{f(t, \cdot, \cdot) + (\cdot)^T B(t)w(t)\} dt$ is second-order pseudoinvex and $\int_I y(t)^T g(t, \cdot, \cdot) dt$ is second-order quasi-invex. If $B(t) = 0, t \in I$, then dual problem (QD) will reduce to the mixed type dual treated by Husain and Bilal [12].

6 Natural Boundary Values

In this part, we formulate a pair of nondifferentiable generalized dual variational problems with natural boundary values rather than fixed end points.

$$(QP_0) : \text{Minimize } \int_I \left\{ f(t, x, \dot{x}) + (x(t)^T B(t)x(t))^{\frac{1}{2}} \right\} dt$$

subject to $x(a) = 0 = x(b),$
 $g(t, x, \dot{x}) \leq 0, t \in I.$

$$(QD_0) : \text{Maximize } \int_I \left(f(t, u(t), \dot{u}(t)) + u(t)^T B(t)w(t) + \sum_{i \in I_0} y^i(t)g^i(t, u, \dot{u}) - \frac{1}{2} \beta(t)^T H^0 \beta(t) \right) dt$$

subject to $u(a) = 0 = u(b),$
 $f_u(t, u, \dot{u}) + B(t)w(t) + y(t)^T g_u(t, u, \dot{u}) - D(f_u(t, u, \dot{u}) + y(t)^T g_u(t, u, \dot{u})) + H \beta(t) = 0, t \in I,$
 $\int_I \left(\sum_{i \in I_\alpha} y^i(t)g^i(t, u, \dot{u}) - \frac{1}{2} \beta(t)^T G^\alpha \beta(t) \right) dt \geq 0, t \in I, \alpha = 1, 2, \dots, r$
 $w(t)^T B(t)w(t) \leq 1, y(t) \geq 0, t \in I,$
 $\sum_{i \in I_\alpha} y^i(t)g_u^i(t, u, \dot{u}) \Big|_{t=a}^{t=b} = 0 \text{ and } f_{\dot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} y^i(t)g_{\dot{u}}^i(t, u, \dot{u}) \Big|_{t=a}^{t=b} = 0, t \in I.$

7 Nondifferentiable Nonlinear Programming Problems

If all the functions of the problem (QP₀) and (QD₀) are independent of t as $b - a = 1$, then these problem will reduces to the following dual problem formulated by Zhang and Mond [13]:

$$\begin{aligned}
 (\mathbf{QP}_1): \text{ Minimize } & f(x) + (x^T Bx)^{1/2} \\
 \text{subject to } & g(x) \leq 0, \\
 (\mathbf{QD}_1): \text{ Maximize } & f(u) + \sum_{i \in I_0} y^i g^i(u) + u^T Bw - \frac{1}{2} \beta^T \hat{H}^0 \beta \\
 \text{subject to } & f_u(u) + Bw + y^T g_u(u) + y^T g_u(u) + \hat{H}\beta = 0, \\
 & \sum_{i \in I_z} y^i g^i(u) - \frac{1}{2} \beta^T \hat{G} \beta \geq 0, w^T Bw \leq 1, y \geq 0.
 \end{aligned}$$

where $\hat{H}^0 = f_{uu}(u) + \left(\sum_{i \in I_0} y^i g^i \right)_{uu}$, $\hat{H} = f_{uu}(u) + y^T g_{uu}(u)$ and $\hat{G} = \left(\sum_{i \in I_z} y^i g^i \right)_{uu}$.

In theorem 1.13, the symbols H^0, H_j^0, G^z , and G_j^z will respectively become as

$$\begin{aligned}
 H^0 &= f_{uu}(u) + \left(\sum_{i \in I_0} y^i g^i(u) \right)_{uu} = \nabla^2 \left(f(u) + \sum_{i \in I_0} y^i g^i(u) \right), H_j^0 = \left[\nabla^2 \left(f(u) + \sum_{i \in I_0} y^i g^i(u) \right) \right]'_j \\
 G^z &= \left(\sum_{i \in I_z} y^i g^i(u) \right)_{uu} = \nabla^2 \left(\sum_{i \in I_z} y^i g^i(u) \right) \text{ and } G_j^z = \left[\nabla^2 \left(\sum_{i \in I_z} y^i g^i(u) \right) \right]'_j.
 \end{aligned}$$

8 Generalized Second-Order Duality for a Continuous Programming Problem with Support Functions

In this part, we consider the following nondifferentiable continuous programming problems with support functions treated by Husain and Jabeen [14]:

$$\begin{aligned}
 (\mathbf{SP}) : \quad \text{Minimize } & \int_I \{f(t, x, \dot{x}) + S(x(t)|K)\} dt \\
 \text{subject to } & x(a) = 0 = x(b), \\
 & g^i(t, x, \dot{x}) + S(x(t)|C^i) \leq 0, i = 1, 2, \dots, m, t \in I.
 \end{aligned}$$

where f and g are continuously differentiable and each $C^i, i = 1, 2, \dots, m$ is a compact convex set in R^n .

Husain and Jabeen [14] derived the following optimality conditions for (SP):

Lemma 2 (Fritz-John necessary optimality conditions): If the problem (SP) attains a minimum at $x = \bar{x} \in X$, then there exist $r \in R$ and piecewise smooth

function $\bar{y} : I \rightarrow R^m$ with $\bar{y}(t) = (\bar{y}^1(t), \bar{y}^2(t), \dots, \bar{y}^m(t))$, $\bar{z} : I \rightarrow R^n$ and $\bar{w}^i : I \rightarrow R^n$, $i = 1, 2, \dots, m$ such that

$$r[f_x(t, \bar{x}, \dot{\bar{x}}) + \bar{z}(t)] + \sum_{i=1}^m \bar{y}^i(t) [g_x^i(t, \bar{x}, \dot{\bar{x}}) + \bar{w}^i(t)] = D[rf_{\dot{x}}(t, \bar{x}, \dot{\bar{x}}) + \bar{y}(t)^T g_{\dot{x}}(t, \bar{x}, \dot{\bar{x}})], t \in I,$$

$$\sum_{i=1}^m \bar{y}^i(t) [g^i(t, \bar{x}, \dot{\bar{x}}) + \bar{x}(t)^T \bar{w}^i(t)] = 0, t \in I,$$

$$\bar{x}(t)^T \bar{z}(t) = S(\bar{x}(t)|K), \bar{x}(t)^T \bar{w}^i(t) = S(\bar{x}(t)|C^i), t \in I,$$

$$w^i(t) \in C^i, i = 1, 2, \dots, m, \bar{z}(t) \in K, (r, \bar{y}(t)) \geq 0, (r, \bar{y}(t)) \neq 0, t \in I,$$

The minimum $\bar{x}(t)$ of (SP) may be described as normal, if $\bar{r} = 1$, so that the Fritz John optimality conditions reduce to Karush-Kuhn-Tucker optimality conditions. It suffices for $\bar{r} = 1$, that Slater's [14] condition holds at $\bar{x}(t)$.

Now, we review some well known facts about a support function for easy reference. Let K be a compact set in R^n , then the support function of K is defined by

$$S(x(t)|K) = \max \{x(t)^T v(t) : v(t) \in K, t \in I\}.$$

A support function, being convex everywhere finite, has a subdifferential in the sense of convex analysis then there exist $z(t) \in R^n$, $t \in I$ such that $S(y(t)|K) - S(x(t)|K) \geq (y(t) - x(t))^T z(t)$. From [15], the subdifferential of $S(x(t)|K)$ is given by $\partial S(x(t)|K) = \{z(t) \in K, t \in I | x(t)^T z(t) = S(x(t)|K)\}$. For any set $\Gamma \subset R^n$, the normal cone to Γ at a point $x(t) \in \Gamma$ is defined by $N_\Gamma(x(t)) = \{y(t) \in R^n | y(t)(z(t) - x(t)) \leq 0, z(t) \in \Gamma\}$. It can be verified that for a compact convex set C , $y(t) \in N_C(x(t))$, if and only if $S(y(t)|C) = x(t)^T y(t)$, $t \in I$.

9 Generalized Second-Order Duality

In this part, we present the formulation of a generalized differentiable second-order Mond-Weir type dual to (SP) which jointly represents Wolfe and Mond-Weir type duals to (SP). The second-order Wolfe type and Mond-Weir type to (SP) were treated in [9] and [10] as the following differentiable continuous programming problems.

$$(WSD): \text{ Maximize } \int_I \{f(t, u, \dot{u}) + u(t)^T z(t) + \sum_{i=1}^m y^i(t) (g^i(t, u, \dot{u}) + u(t)^T w^i(t))\} dt$$

subject to $u(a) = 0 = u(b),$

$$f_u(t, u, \dot{u}) + z(t) + \sum_{i=1}^m y^i(t)^T (g_u^i(t, u, \dot{u}) + w^i(t)) - D(f_u(t, u, \dot{u}) + y(t)^T g_u(t, u, \dot{u})) + H(t)p(t) = 0, t \in I,$$

$$z(t) \in K, w^i(t) \in C^i, i = 1, 2, \dots, m, y(t) \geq 0, t \in I,$$

$$H(t) = f_{uu}(t, u, \dot{u}) + (y(t)^T g_{uu}(t, u, \dot{u}))_u - 2D[f_{u\dot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\dot{u}}]$$

$$+ D^2[f_{\dot{u}\dot{u}}(t, u, \dot{u}) + (y(t)^T g_{\dot{u}\dot{u}}(t, u, \dot{u}))_{\dot{u}}].$$

$$\begin{aligned}
 (M - WSD) : \quad & \text{Maximize } \int_I \left(f(t, u, \dot{u}) + u(t)^T z(t) - \frac{1}{2} p(t)^T F p(t) \right) dt \\
 & \text{subject to} \quad u(a) = 0 = u(b), \\
 & f_{\dot{u}} + z(t) + \sum_{i=1}^m y^i(t) (g_u^i + w^i(t)) - D(f_{\dot{u}} + y(t)^T g_{\dot{u}}) + (F + G)p(t) = 0, \quad t \in I, \\
 & \int_I \left(\sum_{i=1}^m y^i(t) (g^i + u(t)^T \omega^i(t)) - \frac{1}{2} p(t)^T G p(t) \right) dt \geq 0, \quad t \in I, \\
 & z(t) \in K, w^i(t) \in C^i, t \in I, i = 1, 2, \dots, m, y(t) \geq 0, \quad t \in I,
 \end{aligned}$$

where (i) $p(t) \in R^n, t \in I$ (ii) $F = f_{uu} - 2Df_{u\dot{u}} + D^2f_{\dot{u}\dot{u}} - D^3f_{\dot{u}\dot{u}}, t \in I$
 (iii) $G = y(t)^T g_{uu} - 2D(y(t)^T g_u)_u + D^2(y(t)^T g_{\dot{u}\dot{u}}) - D^3(y(t)^T g_{\dot{u}\dot{u}}), t \in I$.

Using the second-order invexity conditions of

- (1) $\int_I \{f(t, \dots) + (\cdot)^T z(t)\} dt$ and $\sum_{i=1}^m \int_I \{y^i(t)(g^i(t, \dots) + (\cdot)w^i(t))\} dt$ or second-order pseudoinvexity of
- (2) $\int_I \left\{ f(t, \dots) + (\cdot)^T z(t) + \sum_{i=1}^m y^i(t)^T (g^i(t, \dots) + (\cdot)w^i(t)) \right\} dt,$

Husain and Masoodi [9] established various duality results between (SP) and (WSD) and Husain and Srivastava [10] validated duality theorems between (SP) and (M-WSD) under the assumptions that with respect to the same η ,

- (3) $\int_I (f(t, \dots) + (\cdot)^T z(t)) dt$ is second-order pseudoinvex and
- (4) $\int_I \sum_{i=1}^n (y^i(t) g^i(t, \dots)^T + (\cdot)^T w^i(t)) dt$ is second-order quasi-invex.

We now construct the following generalized second-order dual to (SP):

$$\begin{aligned}
 (SD) : \text{Maximize } & \int_I \left[f(t, u, \dot{u}) + u(t)^T z(t) + \sum_{i \in I_0} y^i(t) (g^i(t, u, \dot{u}) + u(t)^T w^i(t)) - \frac{1}{2} p(t)^T H^0 p(t) \right] dt \\
 & \text{subject to} \quad u(a) = 0 = u(b), \\
 & f_u(t, u, \dot{u}) + z(t) + \sum_{i=1}^m y^i(t)^T (g_u^i(t, u, \dot{u}) + w^i(t)) - D(f_u(t, u, \dot{u}) + y(t)^T g_u(t, u, \dot{u})) + H p(t) = 0, t \in I \\
 & \int_I \left(\sum_{i \in I_\alpha} y^i(t) (g^i(t, u, \dot{u}) + u(t)^T w^i(t)) - \frac{1}{2} p(t)^T G^\alpha p(t) \right) dt \geq 0, \alpha = 1, 2, \dots, r, t \in I, \\
 & z(t) \in K, w^i(t) \in C^i, i = 1, 2, \dots, m, y(t) \geq 0, t \in I,
 \end{aligned}$$

where (i) $I_\alpha \subseteq M = \{1, 2, \dots, m\}, \alpha = 0, 1, \dots, r$ with $\bigcup_{\alpha=0}^r I_\alpha = M$ and $I_\alpha \cap I_\beta = \emptyset$, if $\alpha \neq \beta$,

$$\begin{aligned}
 (ii) H^0 &= f_{uu}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t) g^i(t, u, \dot{u}))_{uu} - 2D \left(f_{u\dot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t) g^i(t, u, \dot{u}))_{u\dot{u}} \right) \\
 &+ D^2 \left(f_{\dot{u}\dot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t) g_{\dot{u}\dot{u}}^i(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) - D^3 \left(f_{\ddot{u}\ddot{u}}(t, u, \dot{u}) + \sum_{i \in I_0} (y^i(t) g_{\ddot{u}\ddot{u}}^i(t, u, \dot{u}))_{\ddot{u}\ddot{u}} \right), t \in I, \\
 (iii) H &= f_{uu}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_u - 2D(f_{u\dot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\dot{u}}) \\
 &+ D^2(f_{\dot{u}\dot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\dot{u}\dot{u}}) - D^3(f_{\ddot{u}\ddot{u}}(t, u, \dot{u}) + (y(t)^T g_u(t, u, \dot{u}))_{\ddot{u}\ddot{u}}), t \in I, \\
 (iv) G^\alpha &= \sum_{i \in I_\alpha} (y^i(t) g_u^i(t, u, \dot{u}))_u - 2D \left(\sum_{i \in I_\alpha} (y^i(t) g_u^i(t, u, \dot{u}))_{\dot{u}} \right) + D^2 \left(\sum_{i \in I_\alpha} (y^i(t) g_{\dot{u}\dot{u}}^i(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) \\
 &- D^3 \left(\sum_{i \in I_\alpha} (y^i(t) g^i(t, u, \dot{u}))_{\ddot{u}} \right), \alpha = 1, 2, \dots, r, t \in I \text{ and} \\
 (v) G &= \sum_{i \in M-I_0} (y^i(t) g_u^i(t, u, \dot{u}))_u - 2D \left(\sum_{i \in M-I_0} (y^i(t) g_u^i(t, u, \dot{u}))_{\dot{u}} \right) + D^2 \left(\sum_{i \in M-I_0} (y^i(t) g_{\dot{u}\dot{u}}^i(t, u, \dot{u}))_{\dot{u}\dot{u}} \right) \\
 &- D^3 \left(\sum_{i \in M-I_0} (y^i(t) g^i(t, u, \dot{u}))_{\ddot{u}} \right), \alpha = 1, 2, \dots, r, t \in I,
 \end{aligned}$$

The following theorems hold for the problem (SP):

Theorem 4 (Weak Duality): Let $x(t)$ be feasible for (SP) and $(u(t), y(t), z(t), w^1(t), w^2(t), \dots, w^m(t), p(t))$ be feasible for (SD). If for all feasible $(x(t), u(t), y(t), z(t), w^1(t), w^2(t), \dots, w^m(t), p(t))$, $\int_I (f(t, \dots) + (\cdot) z(t) + \sum_{i \in I_0} (g^i(t, \dots) + (\cdot)^T w^i(t))) dt$ is second-order pseudoinvex and $\sum_{i \in I_\alpha} \int_I (g^i(t, \dots) + (\cdot)^T w^i(t)) dt, \alpha = 1, 2, \dots, r$ is second-order quasi-invex with respect to the same η , then,

$$\inf. (SP) \geq \sup. (SD).$$

Theorem 5 (Strong Duality): If $\bar{x}(t)$ is an optimal solution of (SP) and normal [3], then there exist piecewise smooth functions $\bar{y} : I \rightarrow R^m, \bar{z} : I \rightarrow R^n$ and $\bar{w}^i : I \rightarrow R^n, i = 1, 2, \dots, m$ such that $(\bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{w}^1(t), \bar{w}^2(t), \dots, \bar{w}^m(t), \bar{p}(t) = 0)$ is feasible for (SD) and the corresponding values of (SP) and (SD) are equal. If for all feasible $(\bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{w}^1(t), \bar{w}^2(t), \dots, \bar{w}^m(t), \bar{p}(t))$, $\int_I \{f(t, \dots) + (\cdot)^T \bar{z}(t) + \sum_{i \in I_0} \bar{y}^i(t) (g^i(t, \dots) + (\cdot)^T \bar{w}^i(t))\} dt$ is second-order pseudoinvex and $\sum_{i \in I_\alpha} \int_I \bar{y}^i(t) (g^i(t, \dots) + (\cdot)^T \bar{w}^i(t)) dt, \alpha = 1, 2, \dots, r$ is second-order quasi-invex, then $(\bar{x}(t), \bar{y}(t), \bar{z}(t), \bar{w}^1(t), \bar{w}^2(t), \dots, \bar{w}^m(t), p(t))$ is an optimal solution of (SD).

Theorem 6 (Converse Duality): Let $(x(t), y(t), z(t), w^1(t), w^2(t), \dots, w^m(t), p(t))$ be an optimal solution of (SD) at which

(A₁): for all $\alpha = 1, 2, \dots, r$ either

$$(a) \int_I p(t)^T \left(G_\alpha + \sum_{i \in I_\alpha} (y^i g_x^i)_x \right) p(t) dt > 0 \text{ and } \int_I p(t)^T \left(\sum_{i \in I_\alpha} y^i (g_x^i + w^i(t)) \right) dt \geq 0,$$

or

$$(b) \int_I p(t)^T \left(G^\alpha + \sum_{i \in I_\alpha} (y^i g_x^i)_x \right) p(t) dt < 0 \text{ and } \int_I p(t)^T \left(\sum_{i \in I_\alpha} y^i (g_x^i + w^i(t)) \right) dt \leq 0,$$

(A₂): the vectors $H_j^0, G_j^\alpha, \alpha = 1, 2, \dots, r$ and $j = 1, 2, \dots, n$ are linearly independent, where H_j^0 is the j th row of the matrix H^0 and G_j^α is the j th row of the matrix G^α an

(A₃): the vectors $\sum_{i \in I_\alpha} (y^i(t)(g_x^i + w^i(t)) - D(y^i(t)g_x^i)), \alpha = 1, 2, \dots, r$ are linearly independent. If for all feasible $(x(t), y(t), z(t), w^1(t), w^2(t), \dots, p(t)), \int_I \{f(t, \dots) + (\cdot)^T z(t) + \sum_{i \in I_0} y^i(t)(g^i(t, \dots) + (\cdot)^T w^i(t))\} dt$ is second-order pseudoinvex and $\int_I \sum_{i \in I_\alpha} y^i(t)(g^i(t, \dots) + (\cdot)^T w^i(t)) dt, \alpha = 1, 2, \dots, r$ is second-order quasi-invex, then $x(t)$ is an optimal solution of (SP).

Theorem 7 (Strict Converse Duality): Assume that $\int_I f(t, \dots) + (\cdot)^T z(t) + \sum_{i \in I_0} y^i(t)(g^i(t, \dots) + (\cdot)^T w^i(t)) dt$ is second-order strictly pseudoinvex and $\sum_{i \in I_\alpha} \int_I y^i(t)(g^i(t, \dots) + (\cdot)^T w^i(t)) dt, \alpha = 1, 2, \dots, r$ is second-order quasi-invex with respect to same η . Assume also, that (SP) has an optimal solution $\bar{x}(t)$. If $(\bar{u}(t), y(t), z(t), w^1(t), w^2(t), \dots, w^m(t), p(t))$ is an optimal solution of (SD), then $\bar{u}(t)$ is an optimal solution of (SP) with $\bar{x}(t) = \bar{u}(t), t \in I$.

10 Special Cases

Let for $t \in I, A(t), B^i(t), i = 1, 2, \dots, m$ be positive semi definite matrices and continuous on I , then $(x(t)^T B^i(t)x(t))^{1/2} = S(x(t)|C^i)$ and $(x(t)^T A(t)x(t))^{1/2} = S(x(t)|K)$ where $K = \{A(t)z(t)|z(t)^T A(t)z(t) \leq 1\}, i = 1, 2, \dots, m, t \in I$ Replacing $S(x(t)|K)$ by $(x(t)^T A(t)x(t))^{1/2}$ and $S(x(t)|C^i)$ by $(x(t)^T B^i(t)x(t))^{1/2}$, we have the following problems:

$$(SD) : \text{Maximize } \int_I \left[f(t, u, \dot{u}) + u(t)^T A(t)z(t) + \sum_{i \in I_0} y^i(t)(g^i(t, u, \dot{u}) + u(t)^T B^i(t)w^i(t)) - \frac{1}{2}p(t)^T H^0 p(t) \right] dt$$

subject to $u(a) = 0 = u(b),$

$$f_u(t, u, \dot{u}) + A(t)z(t) + \sum_{i=1}^m y^i(t)(g_u^i(t, u, \dot{u}) + B^i(t)w^i(t)) - D(f_u(t, u, \dot{u}) + y(t)^T g_u(t, u, \dot{u})) + H p(t) = 0,$$

$$\int_I \left(\sum_{i \in I_\alpha} y^i(t)(g^i(t, u, \dot{u}) + u(t)^T B^i(t)w^i(t)) - \frac{1}{2}p(t)^T G^\alpha p(t) \right) dt \geq 0, \alpha = 1, 2, \dots, r, t \in I,$$

$$z(t)^T A(t)z(t) \leq 1, w^i(t)^T B^i(t)w^i(t) \leq 1, i = 1, 2, \dots, m, y(t) \geq 0, t \in I.$$

If $S(x(t)|C^i)$, $i = 1, 2, \dots, m$ are deleted from the constraints of (SP), then (SP) and (SD) form a pair of dual problems treated by Husain and Srivastava [16].

11 Problems with Natural Boundary Values

In this part, we formulate a pair of nondifferentiable dual variational problems with natural boundary values rather than fixed end points:

$$\begin{aligned}
 (SP_0) : \text{Minimize} \quad & \int_I \{f(t, x, \dot{x}) + S(x(t)|K)\} dt \\
 \text{subject to} \quad & x(a) = 0 = x(b), \\
 & g^i(t, x, \dot{x}) + S(x(t)|C^i) \leq 0, t \in I, i = 1, 2, \dots, m. \\
 (SD_0) : \text{Maximize} \quad & \int_I \left[f(t, u, \dot{u}) + u(t)^T z(t) + \sum_{i \in I_0} y^i(t) (g^i(t, u, \dot{u}) + u(t)^T w^i(t)) - \frac{1}{2} p(t)^T H^0 p(t) \right] dt \\
 \text{subject to} \quad & u(a) = 0 = u(b), \\
 & f_u(t, u, \dot{u}) + z(t) + \sum_{i=1}^m y^i(t)^T (g_u^i(t, u, \dot{u}) + w^i(t)) - D(f_u(t, u, \dot{u}) + y(t)^T g_u(t, u, \dot{u})) + H p(t) = 0, t \in I, \\
 & \int_I \left(\sum_{i \in I_2} y^i(t) (g^i(t, u, \dot{u}) + u(t) w^i(t)) - \frac{1}{2} p(t)^T G^2 p(t) \right) dt \geq 0, t \in I, \alpha = 1, 2, \dots, r \\
 & z(t) \in K, w^i(t) \in C^i, i = 1, 2, \dots, m, y(t) \geq 0, t \in I, \\
 & f_u(t, u, \dot{u}) = 0 \text{ at } t = a \text{ and } t = b, y^i(t) g_u^i(t, u, \dot{u}) = 0, i = 1, 2, \dots, m \text{ at } t = a \text{ and } t = b.
 \end{aligned}$$

12 Nonlinear Programming Problems

If all functions in the problems (SP₀) and (SD₀) are independent of t , then these problems will reduce to the following nonlinear programming problems studied by Husain et al. [11].

$$\begin{aligned}
 (SP_1) : \text{Minimize} \quad & f(x) + S(x|K) \\
 \text{subject to} \quad & g^i(x) + S(x|C^i) \leq 0, i = 1, 2, \dots, m.
 \end{aligned}$$

$$\begin{aligned}
 (\mathbf{SD}_1) : \quad & \text{Maximize } f(u) + u^T z + \sum_{i \in I_0} y^i (g^i + u w^i) - \frac{1}{2} p^T H^0 p \\
 & \text{subject to} \\
 & f_u(u) + z + \sum_{i=1}^m y^i (g_u^i(u) + w^i) + H p = 0, \sum_{i \in I_\alpha} y^i (g^i + u w^i) - \frac{1}{2} p^T G^\alpha p \geq 0, \alpha = 1, 2, \dots, r, \\
 & z \in K, w^i \in C^i, i = 1, 2, \dots, m, y \geq 0, \\
 & \text{where } H^0 = f_{uu} + \left(\sum_{i \in I_0} y^i g^i \right)_{uu}, H = f_{uu} + \left(\sum_{i \in I_\alpha} y^i g^i \right)_{uu} \text{ and } G^\alpha = \left(\sum_{i \in I_\alpha} y^i g^i \right)_{uu}, \alpha = 1, 2, r.
 \end{aligned}$$

13 Conclusions

In the first part, a generalized second-order dual is formulated for a class of continuous programming problems with square roots of positive semi-definite quadratic forms, hence it is nondifferentiable. Under second-order pseudoinvexity and second-order quasi-invexity, various duality theorems are proved for this pair of dual nondifferentiable continuous programming problems. Lastly, it is pointed out that our duality results can be regarded as dynamic generalizations of those for a nondifferentiable nonlinear programming problem, already treated in the literature. In the second part, a generalized second-order dual is formulated for a continuous programming problem with support functions. Duality results similar to those of the first part are validated for this class of problems. Clearly the results of this part are more general in sense that a support function extends a square root of a positive semidefinite quadratic form.

References

1. Mangasarian, O.L.: Second and higher-order duality in nonlinear programming. *J. Math. Anal. Appl.* **51**, 605–620 (1979)
2. Mond, B.: Second-order duality in nonlinear programming. *Opsearch* **11**, 90–99 (1974)
3. Mond, B., Weir, T.: Generalized convexity and higher-order duality. *J. Math. Anal. Appl.* **46**(1), 169–174 (1974)
4. Chen, X.: Second-order duality for variational problems. *J. Math. Anal. Appl.* **286**, 261–270 (2003)
5. Husain, I., Ahmad, A., Masoodi, M.: Second-order duality for variational problems. *Eur. J. Pure Appl. Math.* **2**(2), 278–295 (2009)
6. Husain, I., Masoodi, M.: Second-order duality for a class of nondifferentiable continuous programming problems. *Eur. J. Pure Appl. Math.* **5**(3), 390–400 (2012)
7. Husain, I., Srivastava, S.K.: On second-order duality in nondifferentiable continuous programming. *Am. J. Oper. Res.* **2**(3), 289–295 (2012)
8. Chandra, S., Craven, B.D., Husain, I.: A class of nondifferentiable continuous programming problems. *J. Math. Anal. Appl.* **107**, 122–131 (1985)

9. Husain, I., Masoodi, M.: Second-order duality for continuous programming containing support functions. *Appl. Math.* **1**, 534–541 (2010)
10. Husain, I., Srivastava, S.K., Shah, A.R.: On continuous programming with support functions. *Appl Math* **4**(10), 1441 (2013)
11. Husain, I., Ahmed, A., Masoodi, M.: Mashoob: second order duality in mathematical programming with support functions. *J. Inf. Math. Sci.* **1**(2), 165–182 (2009)
12. Husain, I., Bilal, Ahmad: Mixed type second-order duality for variational problems. *J. Inf. Math. Sci.* **5**, 1–13 (2013)
13. Zhang, J., Mond, B.: Duality for a nondifferentiable programming problem. *Bull. Australian Math. Sci.* **5**, 20–44 (2013)
14. Husain, I., Jabeen, Z.: Continuous programming containing support functions. *J. Appl. Math. Inf.* **26**(1–2), 75–106 (2008)
15. Mond, B., Schechter, M.: Nondifferentiable symmetric duality. *Bull. Australian Math. Soc.* **53**, 177–188 (1996)
16. Husain, I., Srivastava, S.K.: Mixed type second-order duality for a nondifferentiable continuous programming problem. *Theoretical Math. Appl.* **3**(1), 123–144 (2013)

A Low Cost Electrical Impedance Tomography (EIT) Instrumentation for Impedance Imaging of Practical Phantoms: A Laboratory Study

Tushar Kanti Bera and J. Nagaraju

Abstract A low cost Electrical Impedance Tomography (EIT) instrumentation is developed for studying the impedance imaging of practical phantoms. The EIT instrumentation is developed with a constant current injector, signal conditioner block and a DIP switch based multiplexer module. The constant current injector consists of a variable frequency Voltage Controlled Oscillator (VCO) and a modified Howland based Voltage Control Current Source (VCCS). The signal conditioner block is made up of an instrumentation amplifier (IAmp), a 50 Hz notch filter (NF) and a narrow band pass filter (NBPF) developed by cascading one lowpass filter and a highpass filter. The electrode switching module is developed using DIP switch based multiplexers to switch the electrodes sequentially for injecting current and measuring the boundary voltage data. Load response, frequency response and the Fast Fourier Transform (FFT) studies are conducted to evaluate the VCO, VCCS, IAmp, NF and NBPF performance. A number of practical phantoms are developed and the resistivity imaging is studied in EIDORS to evaluate the instrumentation. Result shows that the instrumentation is suitable for laboratory based practical phantom imaging studies.

Keywords Electrical impedance tomography (EIT) · Low cost EIT instrumentation · VCO · VCCS · Practical phantom · Boundary potentials · Reconstructed image · EIDORS

T. K. Bera · J. Nagaraju
Department of Instrumentation and Applied Physics, Indian Institute of Science,
Bangalore, India
e-mail: tkbera77@gmail.com

1 Introduction

Electrical Impedance Tomography (EIT) [1–10] is an imaging modality in which the spatial distribution of electrical conductivity or resistivity of a closed conducting domain (Ω) is reconstructed from the boundary potentials developed by a constant current injecting through the surface electrodes attached to the boundary ($\partial\Omega$) of the domain to be imaged. The reconstructed image quality in EIT mainly depends on the boundary data accuracy [11] and the performance of the reconstruction algorithm [12–18]. The boundary data profile and its SNR depend on the EIT-instrumentation [19–24] as well as the practical phantom [25–33] under experiment. Hence, in EIT, the practical phantom study with its associated instrumentation is essential for better understanding and analyzing the imaging system so that the boundary data profile can be improved for better image reconstruction. A thorough practical phantom study is also essential for troubleshooting the instrumentation problem or phantom design deficiency. Moreover, a medical EIT system must be tested, calibrated and assessed by a rigorous experimentation with practical phantoms prior to the testing it on patients. Hence the practical phantom study is essentially to be conducted for certifying a system as medically safe.

The instrumentation used for EIT has a great impact on the boundary data accuracy which depends on the current injector, signal conditioner, and a data acquisition. The higher boundary data accuracy in modern EIT systems is generally achieved with a high precision current injector, improved signal conditioner and precision measurement system with high SNR. Also, a modern digital electronic instrumentation [22, 34, 35] and PC based data acquisition system [36] (Fig. 1a) are generally preferred for high speed data collection from a patient under test in real time imaging. Practical phantoms are essentially studied by an EIT system for comparison, calibration and evaluation purposes. An EIT instrumentation is required for impedance imaging studies of practical phantoms. In this direction a low cost, portable, simple, fully controllable analog instrumentation is developed (Fig. 1b) and the resistivity reconstruction of practical phantoms [30] is studied for different phantom configurations and different current injection patterns [1, 2, 6, 37, 38] using EIDORS [39, 40] with different regularization methods [41, 42]. The reconstructed images are evaluated by the image parameters [43, 44] to assess the instrumentation.

2 Materials and Methods

2.1 Instrumentation

In EIT, an electronic instrumentation is essentially required to inject a constant current and measure the boundary potentials required for image reconstruction.

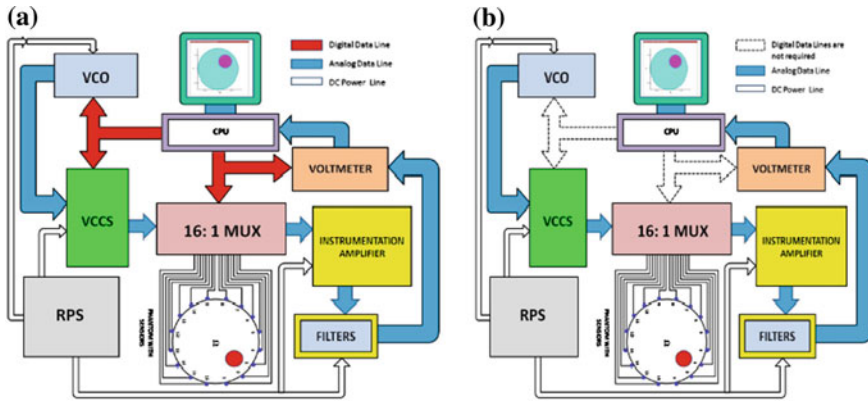


Fig. 1 Comparison with DIP based DAS and PC based DAS. **a** A PC based EIT system schematic. **b** DIP switched based low cost EIT system schematic: digital data lines (shown by the dotted lines in **b**) are not required

A basic EIT instrumentation consists of a constant current injector, signal conditioner, electrode switching module and a data acquisition system as discussed in the next sections. In this paper, a simple EIT instrumentation consisting of consists of a constant current injector, signal conditioner, a simple electrode switching module and a digital voltmeter, is developed. To evaluate the analog instrumentation a detail practical phantom study is conducted in laboratory by analyzing the boundary data profile and the resistivity imaging.

2.1.1 Constant Current Injector

The constant current injector consists of a variable frequency Voltage Controlled Oscillator (VCO) (Fig. 2) and a modified Howland Voltage Control Current Source (VCCS). VCO generates the sinusoidal voltage signal which is fed to the VCCS input for voltage to current conversion. The VCO is developed with is a high-frequency, precision function generator IC, MAX038 [45] as shown in Fig. 2. The VCO is developed to generate a pure sinusoidal signal with an adjustable frequency which is adjusted by changing the value of frequency controller capacitance (C_f) and resistance (R_{in}) and the supply voltage V_{in} as shown in Eq. 1 [45]. The wave form analysis of the VCO output is conducted in a digital storage oscilloscope (DSO) to study its waveform accuracy. The FFT of the VCO output is studied to evaluate the 50 kHz signal and identify the unwanted harmonics presented (if any) in the signal.

$$f = \frac{V_{in}}{R_{in}C_f} \tag{1}$$

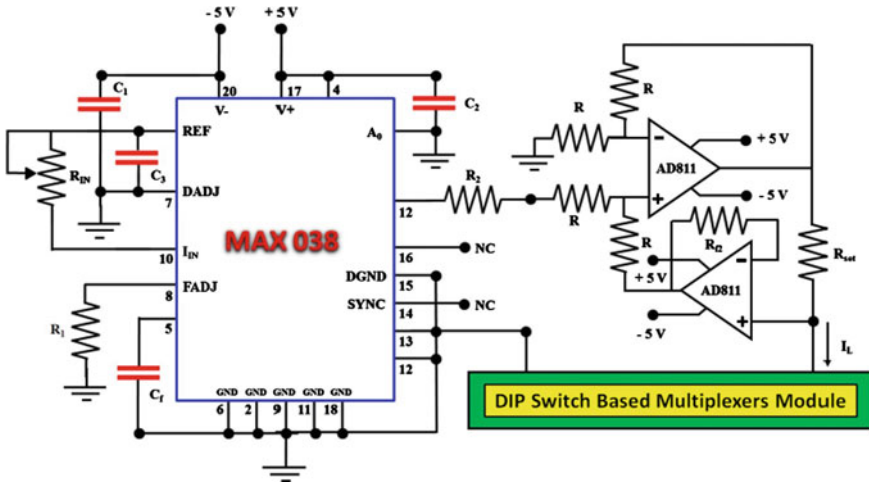


Fig. 2 Schematic of the constant current injector developed with VCO and VCCS

The VCCS is developed with a modified Howland based voltage controlled current source developed with two AD811 Op-Amps [46]. The current amplitude of VCCS is controlled by different values of resistors connected through a selector switch. The wave form accuracy of the VCCS output signal is studied in a DSO. The FFT of the VCCS output is studied to evaluate the 50 kHz current signal and identify the unwanted harmonics presented (if any). The frequency responses of the VCCS for different load are studied to evaluate its load response.

2.1.2 Signal Conditioner

A signal conditioner block (Fig. 3) is developed with an instrumentation amplifier, a 50 Hz notch filter and a narrow band pass filter. The narrow band pass filter developed by cascading one lowpass filter and a highpass filter. The instrumentation amplifier is used to increase the signal to noise ratio (SNR) of the boundary potential and common mode rejection ratio (CMMR) of the circuit. An instrumentation amplifier (Fig. 3) with variable gain is developed using three AD811 Op Amps and the output (V_o) of the amplifier is controlled by varying the value of R_g [47]. The 50 Hz notch filter (Fig. 3) is developed with a single AD811 Op Amps to filter out the 50 Hz noise in the boundary potential. A Band Pass Filter (BPF) is developed (Fig. 3) by cascading a 3rd order Low Pass Filter (LPF) with a 3rd order High Pass Filter (HPF) using AD811 ICs (Fig. 3). All the filter performance is studied in DSO by analyzing their output signals developed by feeding them with a standard sinusoidal voltage signal. The frequency responses of the notch filter and band pass filter are studied by varying the input signal frequency. The entire instrumentation is developed in general purpose printed circuit board (PCB).

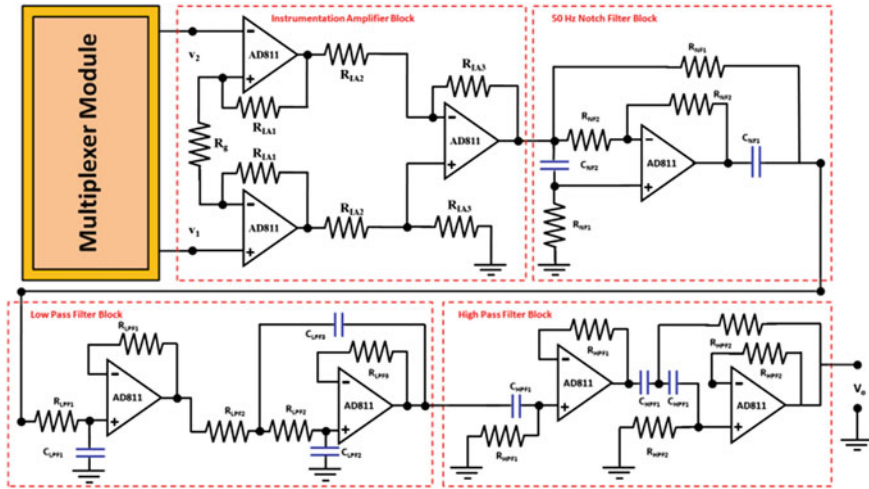


Fig. 3 Schematic of the signal conditioner block

The high band width of VCO and VCCS made them suitable for nonbiological material imaging with capacitance tomography [16, 48–51] and electrical resistance tomography (ERT) [52–56] respectively.

2.1.3 DIP Switch Based Multiplexer Module

In EIT, a sinusoidal current signal of constant amplitude is injected to the boundary of the object under test using a particular current injection protocol [1, 2, 6, 37, 38] and the boundary potentials are measured. Hence the surface electrodes are essentially to be switched which needs a programmable electrode switching module consisting four analog multiplexers interfacing the practical phantom with the data acquisition system. Extra software is required to operate and control the module for accurate switching of current and voltage electrodes on the phantom boundary. In this context a simple electrode switching module is developed using DIP switch [57] based multiplexers which does require any electronics and software to operate and allows us to inject a current signal of suitable magnitude, frequency for the time duration required to troubleshoot the connection problems, faults in instrumentation or improper electrode instrumentation interface. DIP switch is a slide actuated electronic switch (Fig. 4a) having low contact resistance and good mechanical life. Gold plated single pole single throw (SPST) slide actuated DIP switches having 50 mΩ dc contact resistance, 1000 MΩ insulation resistance with a mechanical life of 2,000 operation are used [55]. An 8-point DIP switch (Fig. 4b) has eight parallel switches (Fig. 4c–d) which can be individually or simultaneously operated for single or multiple circuit connections. An 8-point DIP switch can be used as an 8:1 MUX (Fig. 4e–f) by shorting all the eight pins at one side keeping other eight pins open. Similarly two 8-point DIP switches can be

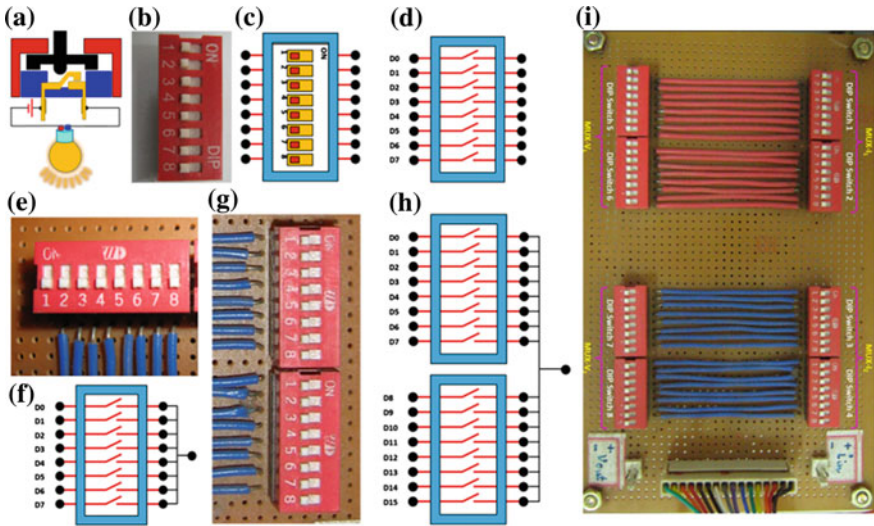


Fig. 4 DIP switch and its applications as the analog multiplexers. **a** a DIP switch schematic. **b** An 8-pin DIP switch. **c** An 8-pin DIP schematic. **d** An 8-pin DIP switch having 8 parallel switches. **e** An 8-pin DIP switch connected as a 8: 1 MUX. **f** Schematic of an 8-pin DIP switch based 8:1 MUX. **g** Two 8-pin DIP switches connected as a 16: 1 MUX. **h** schematic of two 8-pin DIP switch based 16:1 MUX. **i** complete PCB of the DIP switch based electrode switching module

used as a 16:1 MUX (Fig. 4g–h) by shorting all the sixteen pins at one side keeping other sixteen pins open.

In present study, four 16:1 MUXs are used to develop the electrode switching module (Fig. 4i) by shorting all the similar pins of four MUXs and connecting the common point to the corresponding electrodes. Hence, in the electrode switching module, total eight DIP switches are used for developing the electrode switching MUX board consisting of four 16:1 MUXs each of which is made up of two DIP switches. That means, four 1-numbered pins from each of the four MUXs are shorted and connected to the electrode number 1 (Fig. 4i).

Similarly four 2-numbered pins from all the MUXs are shorted and connected to the electrode number 2 and so on. The entire electrodes of the phantom are connected in similar way as shown in the Fig. 4i. Two 16:1 MUXs are used for current electrode switching and other two 16:1 MUXs are used for voltage electrode switching (Fig. 4i). When the pin number 1 of the MUX-I₁ is switched on, the electrode 1 is connected to the positive terminal of the current source and when the pin number 2 of the MUX-I₂ is switched on then the electrode 2 is connected to the negative terminal of the current source (Fig. 4i). Similarly by switching on the pin number 3 of the MUX-V₁ and pin number 4 of the MUX-V₂, electrodes 3 and 4 can easily be connected to the positive and negative terminal of the voltmeter respectively (Fig. 4i).

All the circuit blocks are designed and developed separately to facilitate their individual testing and trouble shooting. All the analog circuits are separately tested

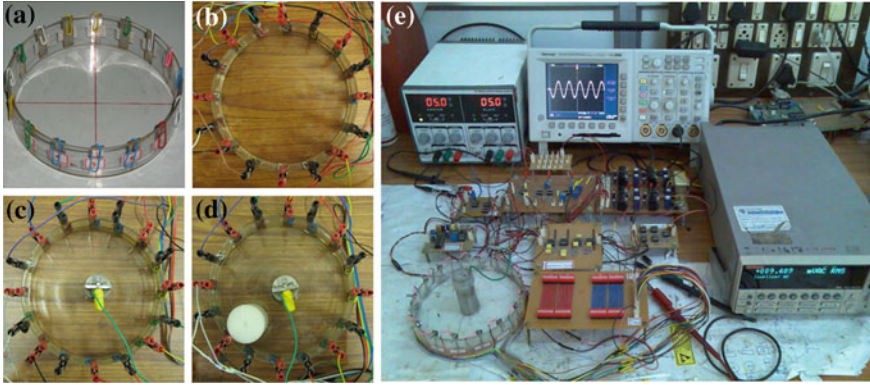


Fig. 5 Phantom development and boundary data collection. **a** A glass tank phantom with stainless steel electrodes fixed with steel clips. **b** SS electrodes connected by steel crocodile clips. **c** Phantom with CME. **d** Phantom with nylon cylinder near electrode 11. **e** Boundary voltage measurement in EIT using Keithley Multimeter (Model 2000)

by studying their output responses in no load as well as loaded conditions. Waveform response, load response, frequency response, noise status and Fast Fourier Transform (FFT) are studied to evaluate the analog circuit performances.

2.2 Image Reconstruction of Practical Phantoms

Sixteen electrode phantoms are developed with a glass tank of 150 mm dia. (Fig. 5a) for resistivity imaging in EIT. Arrays of 16 stainless steel square electrodes are placed on the inner surface of the phantom tanks (Fig. 5b). The electrode arrays are developed by fixing identical and equally spaced squared (10×10 mm) stainless steel (SS) electrodes on the inner wall of the phantom tanks using steel crocodile clips (Fig. 5a). A common mode electrode (CME) made up of stainless steel cylinder (25 mm dia.) is put at the phantom tank center (Fig. 5c) for reducing the common mode error of the system. As the inhomogeneities of the phantoms, the nylon cylinders (35 mm dia.) are put inside NaCl solution at different electrode positions and different phantom configurations are obtained. 1 mA, 50 kHz sinusoidal constant current is injected to the phantom boundary and the potentials are measured for different phantom configurations. Boundary data are also collected for different current injection protocols [1, 2, 6, 37, 38] for the developed sixteen electrode phantoms. In the present study, the boundary potentials of all the electrodes are measured to obtain the greatest sensitivity to the resistivity changes in the domain under test [58].

Resistivity images are reconstructed from the boundary data using Electrical Impedance and Diffuse Optical Reconstruction Software (EIDORS) which is an open source software that aims to reconstruct the 2D-images from electrical or

diffuse optical data. In the present study of the image reconstruction in EIDORS, the circular domains (diameter 150 mm) are discretized with a triangular element mesh using finite element method (FEM) [59]. The image parameters [43, 44] are calculated from the resistivity profiles of the reconstructed images and all the images are analyzed to assess the reconstruction.

3 Results and Discussion

The output signal of the VCO and VCCS are studied across a 1 k Ω resistive load using DSO and it is observed that the outputs of the VCO and VCCS are perfect sinusoidal signals with good accuracy. Multifrequency response are also studied for VCO and VCCS across 1 k Ω resistive load and the results show that the VCO suitably generates sinusoidal voltage signals from 10 Hz to 10 MHz. VCCS was also found suitable for a constant current signal generation with negligible noise from 10 Hz to 1 MHz and hence it is suitable for multifrequency EIT study. The Fast Fourier Transforms (FFT) of the VCO output (voltage) signal and VCCS output (current) signal across 1 k Ω resistive load are studied at different frequency levels. Results show that the signals are free from undesired harmonics which is essential in EIT as the injected current signal and measured voltage signal play a significant role in EIT image quality. The load responses of the VCCS (Fig. 6a) are studied at different frequency levels across different loads from 1 to 7.7 k Ω (Fig. 6a). For all the loads the VCCS generates constant current within the frequency range of 10 Hz to 1 MHz and hence it is found suitable for EIT studies up to 1 MHz.

The band pass filter response shows that it eliminates low and high frequencies from the boundary potentials by allowing the frequencies within a narrow pass band of center frequency of 55 kHz. The notch filter is fed by a standard voltage signal and the filtered output is found free from 50 Hz noise.

Switching of current and voltage electrodes through DIP switch-based multiplexers is found advantageous for conducting the preliminary studies on EIT system. It is also possible to inject current signal with opposite or any other current injection protocols without implementing an extra software and electronic hardware (electronic relays or analog multiplexers). It also injects a current signal of suitable magnitude, frequency for the time duration required for troubleshooting of any connection problems, faults in instrumentation or improper electrode-instrumentation interface. Also, DIP switch-based multiplexer eliminates the effect of stray capacitance of programmable multiplexer hardware. Hence the DIP switch based multiplexers can be suitably used for electrode switching in EIT.

The potential developed at the phantom boundary is also studied in DSO before and after filtering and it is observed that the boundary potential is a noisy sinusoidal signal. It is also noticed that the noise level of the boundary potential is remarkably reduced after filtering (Fig. 6b). Figure 6b shows the boundary

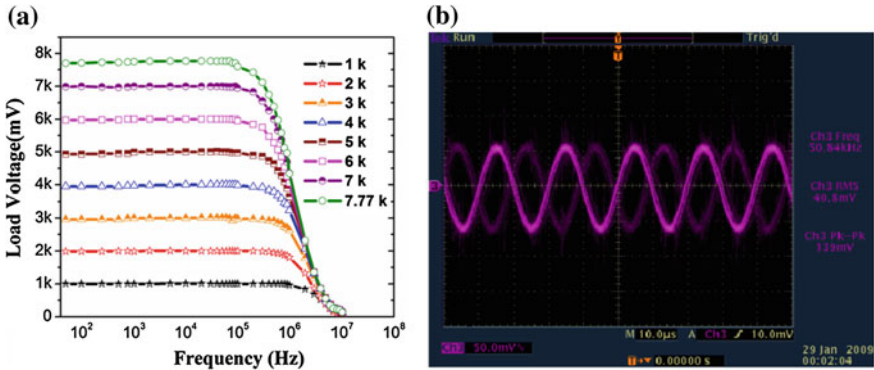


Fig. 6 Current injection performance and boundary voltage data quality. **a** Load response of VCCS. **b** Filtered boundary potential of the practical phantom

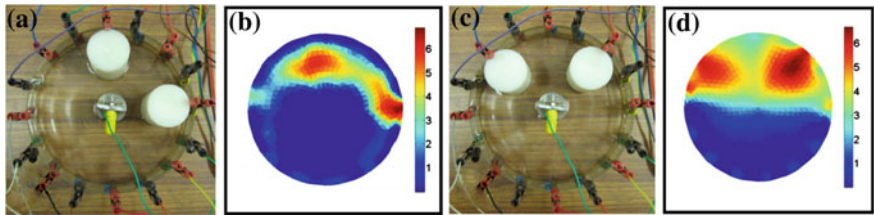


Fig. 7 Resistivity imaging of double inhomogeneity phantoms. **a** Phantom with inhomogeneity at electrode No. 1 and 5. **b** Resistivity image of the phantom of **a**. **c** Reconstructed image at electrode No. 3 and 7. **d** Resistivity image of the phantom of **c**

potential across electrode 3 and 4 generated for current injection through electrode 1 and 2 after filtering.

Resistivity images are reconstructed from the boundary data profiles of the double inhomogeneity phantoms (Fig. 7a, c) but the image quality is poor (Fig. 7b, d). Two peaks of the voltages peaks (V_{ei1} and V_{ei2}) are obtained (due to two inhomogeneities) in the boundary data profiles of double inhomogeneity phantoms. It is observed that for the phantom with two nylon cylinders (35 mm diameter) near the electrode no. 1 and electrode no. 5 (Fig. 7a), V_{ei1} and V_{ei2} are 84.32 mV and 81.96 mV respectively. For the phantom with two similar nylon cylinders near the electrode no. 3 and electrode no. 7 (Fig. 7c), the V_{ei1} and V_{ei2} are 74.97 mV and 74.98 mV respectively (Fig. 7b).

4 Conclusion

A low cost electrical impedance tomography (EIT) instrumentation is developed for impedance imaging of practical phantoms. Frequency response, DSO study and FFT analysis show that the VCO generates a noise free sinusoidal signal with

desired amplitude and frequency suitable for EIT. The DSO study, frequency response, load response and FFT analysis show that the VCCS output is free from unwanted noises and maintains a constant amplitude through a large resistive load up to 1 MHz. Signal conditioner block successfully purify the boundary potential by eliminating the power line noise (50 Hz) and filtering other unwanted noisy signals contributed by the analog circuits. DIP switch-based multiplexer injects a constant current signal of suitable magnitude, frequency in different current injection protocols for the suitable time duration required to troubleshoot the connection problems, faults in instrumentation or improper electrode instrumentation interface. Hence the switching of current and voltage electrodes through DIP MUX module is found simple and advantageous for conducting the preliminary studies of EIT using practical phantoms in laboratory. It is observed that the EIT instrumentation successfully injects a constant current and measures the boundary data with different current injection protocols for resistivity reconstruction. Results show that the SNR is more for the boundary potential data collected from the electrodes nearer to the high current density compared to the data collected from the electrodes in low current density regions. Results also show that the boundary data SNR are more in opposite method compared to the neighbouring method. It is also observed that the phantom domains are reconstructed with the proper resistivity profiles for single and multiple inhomogeneities in both the current patterns. Shape and resistivity of the inhomogeneity and the background resistivity are successfully reconstructed from the potential data for single or double inhomogeneity phantoms. CNR, PCR COC and DRP of the reconstructed images are found high for the inhomogeneities near all the electrodes arbitrarily chosen for the entire study. Hence, the analog instrumentation developed with low cost external components is simple and found suitable for experimental studying of resistivity imaging in EIT.

References

1. Webster, J.G.: *Electrical Impedance Tomography*. Adam Hilger Series of Biomedical Engineering. Adam Hilger, New York (1990)
2. Holder, D.S.: *Electrical Impedance Tomography: Methods, History And Applications*. Series in Medical Physics and Biomedical Engineering. Institute of Physics Publishing Ltd. (2005)
3. Bera, T.K., Nagaraju, J.: *Electrical Impedance Tomography (EIT): A Harmless Medical Imaging Modality*. Research Developments in Computer Vision and Image Processing: Methodologies and Applications, Chap. 13, pp 224–262, IGI Global, USA
4. Bera, T.K., Nagaraju, J.: *Sensors for electrical impedance tomography*. In: Webster, J.G. (ed.) *The Measurement, Instrumentation, and Sensors Handbook*, 2nd Edn, Chap. 61, pp 61-1–61-30. CRC Press, Boca Raton (2014)
5. Cheney, M., Isaacson, D., Newell, J.C.: *Electrical impedance tomography*. *SIAM Rev.* **41**(1), 85–101 (1999)
6. Bera, T.K., Nagaraju, J.: *Studying the resistivity imaging of chicken tissue phantoms with different current patterns in electrical impedance tomography (EIT)*. *Measurement* **45**, 663–682 (2012)

7. Bera, T.K., Nagaraju, J.: Studies on the thin film based flexible gold electrode arrays for resistivity imaging in electrical impedance tomography. *Measurement* **47**, 264–286, Impact Factor: 1.130 (2014)
8. Bera, T.K., Nagaraju, J.: A multifrequency electrical impedance tomography (EIT) system for biomedical imaging. In: International Conference on Signal Processing and Communications (SPCOM 2012), India, IISc, Bangalore, Karnataka, India, pp. 1–5
9. Seo, J.K., Woo, E.J.: Nonlinear inverse problem in imaging. Wiley, New York (2012)
10. Brown, B.H.: Electrical impedance tomography (EIT): a review. *J. Med. Eng. Technol.* **27**(3), 97–108 (2003)
11. Bera, T.K., Nagaraju, J.: Studying the boundary data profile of a practical phantom for medical electrical impedance tomography with different electrode geometries. In: Dössel, O., Schlegel, W.C. (eds.) Proceedings of the World Congress on Medical Physics and Biomedical Engineering. Munich, Germany, WC 2009, IFMBE Proceedings 25/II, pp. 925–929 (2009)
12. Yorkey, T.J.: Comparing reconstruction methods for electrical impedance tomography. PhD thesis, University of Wisconsin at Madison, Madison, WI 53706 (1986)
13. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving conductivity image quality using block matrix-based multiple regularization (BMMR) technique in EIT: a simulation study. *J. Electr. Bioimpedance* **2**, 33–47 (2011)
14. Lionheart, W.R.B.: EIT reconstruction algorithms: pitfalls, challenges and recent developments. *Physiol. Meas.* **25**, 125–142 (2004). Review article
15. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving image quality in electrical impedance tomography (EIT) using projection error propagation-based regularization (PEPR) technique: a simulation study. *J. Electr. Bioimpedance* **2**, 2–12 (2011)
16. Jing, L., Liu, S., Zhihong, L., Meng, S.: An image reconstruction algorithm based on the extended Tikhonov regularization method for electrical capacitance tomography. *Measurement* **42**(3), 368–376 (2009)
17. Bera, T.K., Nagaraju, J.: A simple instrumentation calibration technique for electrical impedance tomography (EIT) using a 16 electrode phantom. In: The fifth annual IEEE conference on automation science and engineering (IEEE CASE), India, pp. 347–352 (2009)
18. Teniou, S., Meribout, M.: A new hierarchical reconstruction algorithm for electrical capacitance tomography using a relaxation region-based approach. *Measurement* **45**(4), 683–690 (2012)
19. Bera, T.K., Nagaraju, J.: A multifrequency constant current source for medical electrical impedance tomography. In: Proceedings of the IEEE international conference on systems in medicine and biology. Kharagpur, India, pp. 278–283 (2010)
20. Soleimani, M.: Electrical impedance tomography system: an open access circuit design. *BioMed. Eng. (OnLine)* **5**(28), 1–8 (2006)
21. Bera, T.K., Nagaraju, J.: Surface electrode switching of a 16-electrode wireless EIT system using RF-based digital data transmission scheme with 8 channel encoder/decoder ICs. *Measurement* **45**, 541–555 (2012)
22. Oh, T.I., Wi, H., Kim, D.Y., Yoo, P.J., Woo, E.J.: A fully parallel multi-frequency EIT system with flexible electrode configuration: KHU Mark2. *Physiol. Meas.* **32**, 835
23. Bera, T.K., Nagaraju, J.: Switching of a sixteen electrode array for wireless eit system using a RF-based 8-bit digital data transmission technique. In: Communications in Computer and Information Science, Part I, CCIS 269, pp. 202–211. Springer, Berlin (2012)
24. Bera, T.K., Nagaraju, J.: Switching of the surface electrodes array in A 16-electrode EIT system using 8-bit parallel digital data. In: IEEE World Congress on Information and Communication Technologies, Mumbai, India, pp. 1288–1293 (2011)
25. Holder, D.S., Hanquan, Y., Rao, A.: Some practical biological phantoms for calibrating multifrequency electrical impedance tomography. *Physiol. Meas.* **17**, A167–A177 (1996)
26. Bera, T.K., Nagaraju, J.: A chicken tissue phantom for studying an electrical impedance tomography (EIT) system suitable for clinical imaging. *Sens. Imaging: Int. J.* **12**(3–4), 95–116

27. Sperandio, M., Guermandi, M., Guerrieri, R.: A four-shell diffusion phantom of the head for electrical impedance tomography. *IEEE Trans. Biomed. Eng.* **59**(2), 383–389 (2012)
28. Bera, T.K., Nagaraju, J.: Resistivity imaging of a reconfigurable phantom with circular inhomogeneities in 2D-electrical impedance tomography. *Measurement* **44**(3), 518–526 (2011)
29. Kimoto, A., Shida, K.: Imaging of temperature-change distribution in the brain phantom by means of capacitance measurement. *IEEE Trans. Instrum. Meas.* **49**(3) (2000)
30. Bera, T.K., Nagaraju, J.: A reconfigurable practical phantom for studying the 2D electrical impedance tomography (EIT) using a fem based forward solver. In: 10th International Conference on Biomedical Applications of Electrical Impedance Tomography (EIT 2009), Manchester, UK (2009)
31. Bera, T.K., Nagaraju, J.: A study of practical biological phantoms with simple instrumentation for electrical impedance tomography (EIT). In: Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC2009), Singapore, pp. 511–516 (2009)
32. Bera, T.K., Nagaraju, J.: A simple instrumentation calibration technique for electrical impedance tomography (EIT) using a 16 electrode phantom. In: 5th Annual IEEE Conference on Automation Science and Engineering, Bangalore, pp. 347–352 (2009)
33. Bera, T.K., Nagaraju, J.: A stainless steel electrode phantom to study the forward problem of electrical impedance tomography (EIT). *Sens. Transducers J.* **104**(5), 33–40 (2009)
34. Wang, M., Ma, Y., Holliday, N., Dai, Y., Williams, R.A., Lucas, G.: A high-performance EIT system. *IEEE Sens. J.* **5**(2), 289 (2005)
35. Goharian, M., Soleimani, M., Jegatheesan, A., Chin, K., Moran, G.R.: A DSP based multi-frequency 3D electrical impedance tomography system. *Ann. Biomed. Eng.* **36**(9), 1594–1603 (2008)
36. Casas, O., Rosell, J., Brag'os, R., Lozano, A., Riu, P.J.: A parallel broadband real-time system for electrical impedance tomography. *Physiol. Meas.* **17**, A1–A6 (1996)
37. Malmivuo, J., Plonsey, R.: *Bioelectromagnetism: principles and applications of bioelectric and biomagnetic fields*, Chap. 26, Sect. 26.2.1. Oxford University Press, New York (1995)
38. Bera, T.K., Nagaraju, J.: Studying the 2D resistivity reconstruction of stainless steel electrode phantoms using different current patterns of electrical impedance tomography (EIT). In: *Biomedical Engineering*, Narosa Publishing House, Proceeding of the International Conference on Biomedical Engineering 2011 (ICBME-2011), Manipal, pp. 163–169 (2011)
39. Polydorides, N., Lionheart, W.R.B.: A Matlab toolkit for three-dimensional electrical impedance tomography: a contribution to the electrical impedance and diffuse optical reconstruction software project. *Meas. Sci. Technol.* **13**, 1871–1883 (2002)
40. Bera, T.K., Nagaraju, J.: A MATLAB based boundary data simulator for studying the resistivity reconstruction using neighbouring current pattern. *J. Med. Eng.* **15** (2013) (Article ID 193578)
41. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Image reconstruction in electrical impedance tomography (EIT) with projection error propagation-based regularization (PEPR): a practical phantom study. In: *Lecture Notes in Computer Science*, vol. 7135/2012, pp. 95–105. Springer, Berlin (2012)
42. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving the image reconstruction in electrical impedance tomography (EIT) with block matrix-based multiple regularization (BMMR): a practical phantom study. In: *IEEE World Congress on Information and Communication Technologies*, India, pp. 1346–1351 (2011)
43. Bera, T.K., Nagaraju, J.: Studying the elemental resistivity profile of electrical impedance tomography (EIT) images to assess the reconstructed image quality. In: 5th International Conference on Information Processing (ICIP 2011). *Communications in Computer and Information Science*, vol. 157 (CCIS-157), India, pp. 621–630 (2011)
44. Bera, T.K., Nagaraju, J.: Gold electrode sensors for electrical impedance tomography (EIT) studies. In: *IEEE Sensors Application Symposium*, USA, pp. 24–28 (2011)

45. Data Sheet, MAX038-high-frequency waveform generator, Maxim Integrated Prod., Inc. USA
46. Data Sheet, AD811—120 MHz, High Speed, Low Noise Video Op Amp, Analog Devices, Inc
47. Bell, A.: *Operational Amplifiers: Applications, Design and Troubleshooting*, 2nd Edn, pp. 272–276. Prentice-Hall, Englewood Cliffs, Chap. 11, 9 Jan 1990
48. Wang, P., Guo, B., Li, N.: Multi-index optimization design for electrical resistance tomography sensor. *Measurement* **46**(8), 2845–2853 (2013)
49. Li, Y., Soleimani, M.: Imaging conductive materials with high frequency electrical capacitance tomography. *Measurement* **46**, 3355–3361 (2013)
50. Gao, H., Xu, C., Fu, F., Wang, S.: Effects of particle charging on electrical capacitance tomography system. *Measurement* **45**(3), 375–383 (2012)
51. Fan, Z., Gao, R.X., Wang, J.: Virtual instrument for online electrical capacitance tomography, practical applications and solutions using LabVIEW, Edited by Folea Silviu, ISBN 978-953-307-650-8, Published: Aug 1 2011 under CC BY-NC-SA 3.0 license
52. Dickin, F., Wang, M.: Electrical resistance tomography for process applications. *Meas. Sci. Technol.* **7**, 247–260 (1996)
53. Daffy, W., Ramlrez, A.: Electrical resistance tomography during in situ trichloroethylene remediation at the Savannah River Site. *J. Appl. Geophys.* **33**, 239–249 (1995)
54. Beauvais, A., Ritz, M., Parisot, J.-C., Dukhan, M., Bantsimba, C.: Analysis of poorly stratified lateritic terrains overlying a granitic bedrock in West Africa, using 2-D electrical resistivity tomography. *Earth Planet. Sci. Lett.* **173**, 413–424 (1999)
55. Kemna, A., Vanderborght, J., Kulesa, B., Vereecken, H.: Imaging and characterisation of subsurface solute transport using electrical resistivity tomography (ERT) and equivalent transport models. *J. Hydrol.* **267**, 125–146 (2002)
56. Beauvais, A., Ritz, M., Parisot, J.-C., Bantsimba, C., Dukhan, M.: Combined ERT and GPR methods for investigating two-stepped lateritic weathering systems. *Geoderma* **119**, 121–132 (2004)
57. Data Sheet, DIP Switches, EDG/EDS type, Excel Cell Electronic Co. Ltd., No. 20, 25th Rd., Taichung Industrial Park, Taichung, Taiwan
58. Cheng, K.S., Simske, S.J., Isaacson, D., Newell, J.C., Gisser, D.G.: Errors due to measuring voltage on current-carrying electrodes in electric current computed tomography. *IEEE Trans. Biomed. Eng.* **37**(60), 60–65 (1990)
59. Bera, T.K., Nagaraju, J.: A FEM-based forward solver for studying the forward problem of electrical impedance tomography (EIT) with a practical biological phantom. In: *Proceedings of IEEE International Advance Computing Conference' 2009 (IEEE IACC 2009)*, 6–7th Mar 2009, Patiala, Punjab, India, pp. 1375–1381

A Broyden's Method Based High Speed Jacobean Matrix Calculator (JMC) for Electrical Impedance Tomography (EIT)

Tushar Kanti Bera, Samir Kumar Biswas, K. Rajan and J. Nagaraju

Abstract Electrical Impedance Tomography (EIT) essentially needs the Jacobean matrix to reconstruct the conductivity distribution of the domain under test. A Broyden's method based high speed Jacobean matrix (J) calculator is proposed for Electrical Impedance Tomography (EIT). The Gauss-Newton-based EIT image reconstruction algorithm repetitively calculates the Jacobian matrix (J) which needs a lot of computation time and cost. Broyden's method based high speed Jacobean matrix calculator (JMC) makes explicit use of secant and adjoint information that can be obtained from the forward solution of the EIT. The Broyden's method based high speed Jacobean matrix calculator (JMC) approaches reduce the computational time remarkably by approximating the system Jacobian (J) successively through low-rank updates. The performance of the JMC is studied with simulated EIT data and the results are compared with Gauss-Newton method based EIT reconstruction. Simulated results show that the Broyden's method based image reconstruction algorithm accelerates the reconstruction speed remarkably.

Keywords Electrical impedance tomography · Broyden's method · Jacobean matrix calculator (JMC) · Image reconstruction · Gauss-Newton's method · MoBIIR

T. K. Bera (✉) · J. Nagaraju
Department of Instrumentation and Applied Physics, Indian Institute of Science,
Bangalore, India
e-mail: tkbera77@gmail.com

S. K. Biswas · K. Rajan
Department of Physics, Indian Institute of Science, Bangalore, India

1 Introduction

Electrical Impedance Tomography (EIT) [1–10] reconstructs the conductivity distribution of an object under test from the boundary voltage-current data using an image reconstruction algorithm. The EIT image reconstruction [1, 4, 8, 9, 11–14] essentially needs to develop a Jacobean matrix [1, 4, 8, 9, 11–14] to reconstruct the conductivity distribution of the domain under test from its boundary measurements. The conventional Gauss-Newton-based EIT image reconstruction algorithm [1, 4, 8, 9, 11–14] repetitively calculates the Jacobian matrix (J) which consumes a lot of computation time and computational cost. The better image reconstruction, generally, requires a fine FEM mesh [15, 16] containing a huge number of finite elements. But, the EIT reconstruction with a domain with FEM mesh with larger number of finite elements needs a lot of time to solve the EIT problem. Also, the 3D EIT reconstruction [9] needs a 3d FEM mesh of the 3D domain under test which essentially consists of a large number of finite elements. Therefore, a better image reconstruction in 2D or 3D with a large FEM mesh needs a lot of computation time for Newton based or Gauss-Newton based algorithms [11–14] due their repetitive calculation of Jacobian matrix (J).

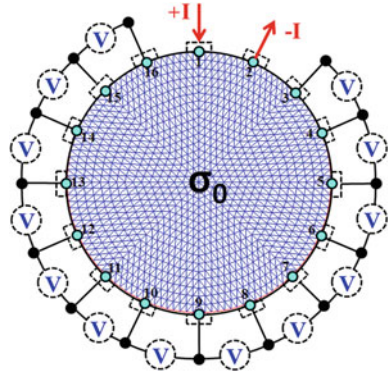
In this direction, a Broyden's method [17] based high speed Jacobean matrix calculator (JMC) is proposed for 2D EIT reconstruction. The Broyden's method based high speed JMC makes explicit use of secant and adjoint information that can be obtained from the EIT forward solution. The Broyden's method based Jacobean matrix calculator (JMC) computes the Jacobian matrix (J) very fast by approximating the system Jacobian (J) successively through low-rank updates. The Broyden's method based JMC has been implemented a model based iterative image reconstruction (MoBIIR) developed in Matlab and the performance of the JMC is studied with simulated EIT data. The conductivity image reconstruction is conducted in MoBIIR with Broyden's method based JMC for different simulated phantom configurations and the results are compared with the Gauss-Newton method based EIT reconstruction. The results obtained from the simulation studies demonstrate that the Broyden's method based JMC accelerates the Jacobian computation and reduce the reconstruction time remarkably.

2 Methods

2.1 Electrical Impedance Tomography (EIT)

Electrical impedance tomography (EIT) is a nonlinear illposed inverse problem [11–14] in which the electrical conductivity distribution of a closed domain under test (DUT) is reconstructed from the boundary potentials developed by injecting a constant current through the surface electrodes [18] attached to the boundary (Fig. 1) using an EIT instrumentation. Using EIT instrumentation [19–22], voltage

Fig. 1 Schematic of the constant current injection and the boundary data collection in electrical impedance tomography (EIT) for a circular domain



measurements are made for a set of current injections, called the current projections, through the different pairs of electrodes. The conductivity distribution is reconstructed from the boundary voltage data sets using an image reconstruction algorithm. The reconstruction algorithm is a computer program which computes the boundary potential data by simulating a known current injection to the domain under test. The current simulation and voltage computation is made by simulating all the current projections through different electrode positions in the domain under test. The reconstruction algorithm uses the boundary current data matrix [C] and the initial (guessed) conductivity distribution [σ_0] and repetitively computes the conductivity distribution by updating the initial conductivity distribution through a number of iterations.

Through the iterative computation of the conductivity distribution, the image reconstruction algorithm actually tries to obtain the original conductivity distribution of the object for which the difference between the measured boundary potential (V_m) and the calculated potential (V_c) is minimum. Thus, EIT image reconstruction algorithm is developed with two parts: forward solver (FS) and inverse solver (IS) as explained in Fig. 2. FS and IS are used to solve the forward problem (FP) and inverse problem (IP) of the EIT respectively.

The forward solver (FS) of EIT image reconstruction algorithm computes the boundary potential data for a known current injection (current simulation) at the boundary and known conductivity distribution of the domain under test (DUT). The inverse solver (IS) is constructed with a minimization algorithm to compute the conductivity update vector for iteratively modifying the initial (guessed) conductivity distribution and to find out an optimum conductivity distribution for which the difference between the measured and calculated data ($\Delta V = V_m - V_c$) becomes minimum. In EIT reconstruction algorithm the FS applies a numerical technique such as finite element method (FEM) to derive the forward model (FM) of DUT to solve the EIT governing equation [1, 4, 8, 9, 11–14] as shown in Eq. 1.

$$\nabla \cdot \sigma \nabla \varphi = 0 \tag{1}$$

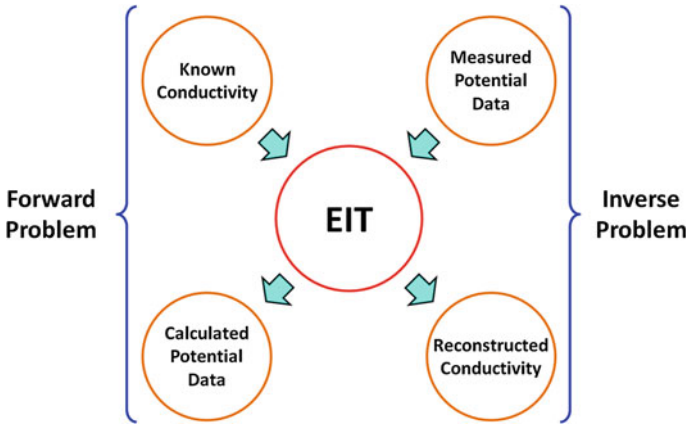


Fig. 2 Schematic of the forward and inverse problem in EIT

The above partial differential equation, EIT governing equation, can be derived from the Maxwell’s equations [4, 7] applicable to the DUT and it relates the elemental conductivities (σ) of all the finite elements of the FEM mesh to their corresponding nodal potential values (Φ).

Using the FEM formulation [3, 9] applied to the EIT governing equation and by utilizing the elemental conductivity values, the nodal coordinates and boundary conditions, the FM of EIT is developed for the DUT and its associated boundary conditions [3, 9]. The FM is developed from the EIT governing equation in the form of a matrix equation as shown by Eq. 2 which represents the relationship between the elemental conductivities of the DUT and the nodal potentials [3, 9]. The FS solves the forward problem (FP) using FM (Eq. 2) and computes the potential distribution with an initial guessed conductivity distribution $[\sigma_0]$ and a known constant current simulation $[C]$. The boundary potential data $[V_c]$, which are the potential of the nodes under the electrode positions at the domain boundary, is then extracted from the nodal potential matrix $[\Phi]$

$$[\Phi] = [K(\sigma)]^{-1}[C] \tag{2}$$

In inverse solver of the EIT reconstruction algorithm, the boundary data mismatch vector ($[\Delta V] = [V_m] - [V_c]$) is minimized by a minimization algorithm such as Gauss-Newton Method based minimization algorithm (GNMMA) [3, 11–14, 24–27]. GNMMA iteratively calculates the conductivity update vector ($[\Delta\sigma]$) to update the initial guessed conductivity distribution ($[\sigma_0]$) [3]. The inverse solver of the image reconstruction algorithm iteratively update vector ($[\Delta\sigma]$) and find the optimum conductivity distribution for which the ΔV is minimum using a iterative technique such as Newton-Raphson Iteration Technique (NRIT) [3]. NRIT iteratively update the initial conductivity distribution to a new conductivity distribution ($[\sigma_1] = [\sigma_0] + [\Delta\sigma]$) for which the forward solver calculates the boundary potentials

more similar to the measured value [3]. Therefore, the inverse solver calculates the conductivity distribution vector ($[\Delta\sigma]$) using Jacobian matrix ($[J]$), regularization parameters and boundary voltage difference ($[\Delta V]$).

2.2 Gauss-Newton Method Based MoBIIR

The Gauss-Newton method [3, 11–14, 24–27] based MoBIIR (GN-MoBIIR) algorithm [3] is developed with a Finite Element Method (FEM) based flexible forward solver (FFS) [3] to solve the forward problem (FP) and Gauss-Newton method based inverse solver (GNIS) working with a modified Newton-Raphson iterative technique (NRIT) to solve the inverse solver (IP). FFS in GN-MoBIIR calculates the boundary potential (V_c) data for a known conductivity distribution $[\sigma_0]$ and current injection and the GNIS computes the domain conductivity distribution for a known current injection and measured boundary potentials (V_m) by utilizing the Jacobian matrix $[J]$.

In GN-MoBIIR, the Jacobian matrix $[J]$ is computed by the adjoint method [3] and conductivity update vector $[\Delta\sigma]$ is calculated by conjugate gradient search using a very low tolerance and large number of iterations. The FFS is developed with a flexible, fast and direct mesh generation code running on pdeTool in Matlab [3] and hence the conductivity imaging is found suitable with a symmetric mesh containing triangular finite elements and nodes.

If V_m is the measured voltage matrix and f is a function mapping an N_e dimensional (N_e is the number of element in the FEM mesh) impedance distribution into a set of M (number of the measured data) approximate measured voltage, then, the GN-MoBIIR [3, 11–14, 24–27] tries to find a least square solution [27, 28] of the minimized object function s_r defined as [3, 11–14, 24–27]:

$$s_r = \frac{1}{2} \|V_m - f\|^2 + \frac{1}{2} \lambda \|\sigma\|^2 = \frac{1}{2} (V_m - f)^T (V_m - f) + \frac{1}{2} \lambda (\sigma)^T (\sigma) \quad (3)$$

where, s_r is the constrained least-square error of the reconstructions, matrix G is the matrix incorporated for regularization process [11–14, 29] called regularization matrix and λ is a positive scalar and called as the regularization coefficient.

Due to the inherent ill-posed nature of EIT, small measurement errors make the solution of Eq. 1 with large error and hence the regularization technique [3, 11–14] is incorporated in GN-MoBIIR in terms of regularization matrix G and regularization operator λ to make the inverse problem well-posed [3, 11–14]. Now, using the Gauss-Newton (GN) method [3, 11–14],

$$\Delta\sigma = \frac{s'_r}{s''_r} = \frac{(J^T)(V_m - f) - \lambda(G)^T(G\sigma)}{(J^T)(J) - (f'')^T(V_m - f) + \lambda G^T G} \quad (4)$$

Neglecting higher terms and replacing $\mathbf{G}^T\mathbf{G}$ by \mathbf{I} (Identity matrix) in Eq. 4, the update vector reduces to reduces to:

$$\Delta\sigma = \frac{\mathbf{J}^T(\mathbf{V}_m - \mathbf{f}) - \lambda I\sigma}{\mathbf{J}^T\mathbf{J} + \lambda I} \quad (5)$$

Thus for k th iteration (k is a positive integer), the GN-MoBIIR gives a regularized solution of the conductivity distribution as:

$$\sigma_{k+1} = \sigma_k + \left[[\mathbf{J}_k]^T[\mathbf{J}_k] + \lambda\mathbf{I} \right]^{-1} \left[[\mathbf{J}_k]^T(\Delta\mathbf{V})_k - \lambda\mathbf{I}\sigma_k \right] \quad (6)$$

Where \mathbf{J}_k and $[\Delta\mathbf{V}]_k$ are the Jacobian and voltage difference matrix respectively at the k th iteration.

2.3 Adjoint Method Based Jacobian Calculation in GN-MoBIIR

In GN-MoBIIR a finite-element model is established and the inverse problem is solved by means of a Gauss-Newton method, which requires the Jacobian matrix (\mathbf{J}). In the EIT, the Jacobian describes the mapping between voltage distribution on the surface of DUT and its internal conductivity distribution. For better reconstruction in EIT, the Jacobian computation has to meet an accuracy requirement as well as to maintain a minimal computation time requirement. In Gauss-Newton method based EIT reconstruction algorithm such as GN-MoBIIR, the Jacobian matrix (\mathbf{J}) is computed using the adjoint method [3, 7, 9, 30–34] using the following Eq. 7:

$$\mathbf{J} = \oint_{\Omega} \nabla\Phi_s \cdot \nabla\Phi_d d\Omega \quad (7)$$

Where, Φ_s and Φ_d denotes the potential data calculated for the current injection through a particular position called source position and the adjoint position.

Thus, in GN-MoBIIR algorithm, the nodal potential matrix (Φ) is calculated in FS for a current injection through a particular electrode (working as a source) with a voltage measurement on a separate electrode (working as a detector) and the potential matrix is denoted by Φ_s . After that, the same procedure is repeated in FS by interchanging the source positions to detector positions and detector positions to the source positions and hence a new potential matrix is obtained which is denoted as Φ_d .

2.4 Broyden's Method Based Jacobean Matrix Calculator (JMC)

A novel high speed Jacobean matrix calculator (JMC) is developed using the Broyden's method [17, 35–37]. In proposed Broyden's method based JMC, the Broyden's method based accelerated scheme computes the Jacobian (J) matrix and the Jacobian is then combined with conjugate gradient scheme (CGS) in MoBIIR for fast reconstruction in EIT. The Broyden-based JMC makes explicit use of secant and adjoint information which is obtained from forward solution of the EIT governing equation. The Broyden-based JMC calculation scheme reduces the computational time remarkably by approximating the system Jacobian (J) successively through low-rank updates.

The Broyden's method improves Newton's method with respect to storage and estimation of the Jacobian. This Broyden method computes the Jacobian (J_{i-1}) at the starting step of the algorithm and then it uses J_{i-1} and improves it using the solution of the secant equation that is a minimal modification to J_{i-1} (minimal in the sense of minimizing the Frobenius norm $\|J_i - J_{i-1}\|_{\text{Frob}}$). The whole system Jacobian $J(\sigma_k)$ is then updated approximately to $J(\sigma_{k+1})$ through rank-1 updates [17, 37–40] using the least change secant based Broyden update method by assuming a $J(\sigma_k)$ as nonsingular matrix. The forward solution can be expressed in terms of derivatives ($A(\sigma) = F'$) of the forward solution. By Taylor expansion we can write:

$$F(\sigma_k) = F(\sigma_{k+1}) + A(\sigma_{k+1})[\sigma_k - \sigma_{k+1}] \quad (8)$$

We wish to form an approximation of $J(\sigma_{k+1})$ from $A(\sigma_{k+1})$, and is given as:

$$F(\sigma_k) = F(\sigma_{k+1}) + J(\sigma_{k+1})[\sigma_k - \sigma_{k+1}] \quad (9)$$

The secant condition [17, 37–40] equation is:

$$J_{k+1}\Delta\sigma_k = \Delta V_k \quad (10)$$

where,

$$\Delta\sigma_k = \sigma_{k+1} - \sigma_k \text{ and } \Delta V_k = F(\sigma_{k+1}) - F(\sigma_k) \quad (11)$$

Now, as the only new information is available in the direction $\Delta\sigma_k$ the J_i and J_{i+1} both must agree on all vectors orthogonal to $\Delta\sigma_k$ [17, 37–40]. Hence, for any direction u , we can write the following relation:

$$p \cdot \Delta\sigma_k = 0 \Rightarrow J_{k+1}p = J_k p \quad (12)$$

Now, as the equations Eqs. 11 and 12 tell that the matrix $[J_{k+1} - J_k]$ has a null space of dimension $(n - 1)$ (because there are $n - 1$ independent vectors orthogonal to $\Delta\sigma_k$) [17, 37–40], J_{i+1} is chosen based on Eqs. 11 and 12. Now, for

any matrix $[A_m]$, its rank plus the dimension of its null space is equal to the number of columns [17, 37–40] in $[A_m]$. Therefore, the rank of the matrix $[J_{k+1} - J_k]$ must be 1, and hence, each column of $[J_{k+1} - J_k]$ must be a multiple of a common vector u and the matrix can be written as [17, 37–40] :

$$[J_{k+1} - J_k] = [v_1u|v_2u|v_3u|v_4u| \dots v_{n-2}u|v_{n-1}u|v_nu] \tag{13}$$

The matrix on the right hand side can be written as uv^T , where u , w and v_n are vectors and are regarded as matrixes. Now, applying the fundamental matrix theories, we can write:

$$uv^T w = (v \cdot w)u \text{ for all } \in \mathbb{R}^n \tag{14}$$

Thus J_{k+1} is required to be chosen as a rank-1 update of J_k [17, 37–40]. Using equation Eq. 13, we get:

$$J_{k+1} = J_k + uv^T \tag{15}$$

Now assuming $v = \Delta\sigma_k$ we get:

$$J_{k+1}\Delta\sigma_k = J_k\Delta\sigma_k + [\Delta\sigma_k \cdot \Delta\sigma_k]u \tag{16}$$

$$\Delta V_k = J_k\Delta\sigma_k + [\Delta\sigma_k \cdot \Delta\sigma_k]u \tag{17}$$

$$u = \frac{[\Delta V_k - J_k\Delta\sigma_k]}{[\Delta\sigma_k \cdot \Delta\sigma_k]} \tag{18}$$

Therefore, the expression for the Jacobian update using Broyden’s method becomes:

$$J_{k+1} = J_k + \frac{[\Delta V_k - J_k\Delta\sigma_k][\Delta\sigma_k]^T}{[\Delta\sigma_k \cdot \Delta\sigma_k]} \tag{19}$$

$$J_{k+1} = J_k + \Delta J \tag{20}$$

where,

$$\Delta J = \frac{[\Delta V_k - J_k\Delta\sigma_k][\Delta\sigma_k]^T}{[\Delta\sigma_k \cdot \Delta\sigma_k]} \tag{21}$$

The Eq. 21 is referred to as the Broyden’s Jacobian update equation. As in the Broyden’s method there is no clue for the initial Jacobian estimation [39], in HSMoBIIR the initial Jacobian $[J(\sigma_0)]$ is calculated through analytical method in which the solution domain is transformed into the trust region and the Jacobian is updated through Broyden’s method. If the initial guessed conductivity distribution σ_0 is sufficiently close to the actual conductivity distribution, then the $J(\sigma_0)$ becomes very close to $A(\sigma_0)$ and the solution converges q-superlinearly to σ^* .

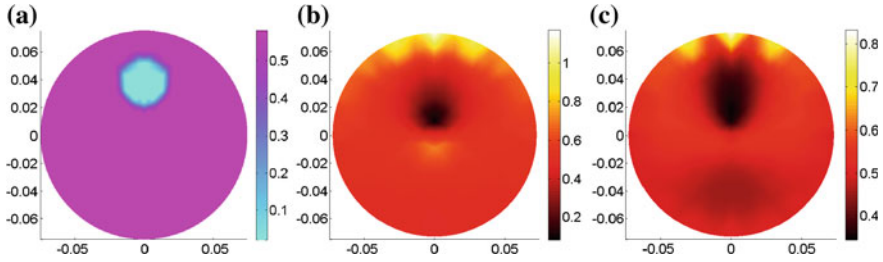


Fig. 3 Conductivity imaging of the simulated phantom with GN-MoBIIR and BM-MoBIIR. **a** Simulated phantom, **b** image with GN-MoBIIR, **c** image with BM-MoBIIR

2.5 Simulated Data Generation

Boundary data are generated from a simulated domain created in computer and by simulating a constant current and calculating the boundary potentials [5, 41]. A number of boundary data sets are generated in MatLAB-based program for a number of phantom configurations. Circular domains are created in the computer program and single and multiple inhomogeneities are generated to define different phantom configurations. The background domain and the inhomogeneities are assigned with two different conductivities and a constant current is simulated to the domain and the boundary data are generated to study the conductivity reconstruction.

2.6 Conductivity Reconstruction

Boundary data sets are generated for a number of phantom configurations and the conductivity images are reconstructed in GN-MoBIIR and with the Broyden's method based MoBIIR (BM-MoBIIR). The results obtained with GN-MoBIIR and BM-MoBIIR are compared to assess the performance of the Broyden's method based Jacobean matrix calculator (JMC).

3 Results and Discussion

Results obtained from the simulation studies show that that the Broyden's method based Jacobean calculator of BM-MoBIIR computes the Jacobian matrix more faster compared to the Jacobian calculator in GN-MoBIIR. Consequently, the BM-MoBIIR reduces the conductivity reconstruction time remarkably. It is observed that the BM-MoBIIR reconstruct the conductivity images from the simulated data five times faster than the GN-MoBIIR. Figure 3 shows the conductivity imaging of a simulated phantom (Fig. 3a) and its reconstructed images by GN-MoBIIR

(Fig. 3b) and BM-MoBIIR (Fig. 3c). The simulated phantom is developed with a circular domain (diameter 150 mm, background conductivity = 0.58 S/m) containing a circular inhomogeneity (diameter 35 mm, conductivity = 0.02 S/m) placed at a polar coordinate of 37.5 mm, 90°. Results show that the GN-MoBIIR takes 178 s where as the BM-MoBIIR takes 25.02 s.

4 Conclusion

Conductivity reconstruct in Electrical Impedance Tomography (EIT) essentially needs the Jacobean matrix. The Gauss-Newton-based EIT image reconstruction algorithm repetitively calculates the Jacobian matrix (J) which needs a lot of computation time and cost. In this paper a Broyden's method based high speed Jacobean matrix (J) calculator is proposed for EIT to accelerate image reconstruction with fast Jacobian computation. Broyden's method based high speed JMC uses the secant and adjoint information obtained from the forward solution and reduce the Jacobian computation time remarkably by approximating the system Jacobian (J) successively through low-rank updates. The performance of the Broyden's method based high speed JMC is studied with simulated data in a model based iterative image reconstruction (MoBIIR) algorithm. The results obtained with Broyden's method based high speed JMC implemented MoBIIR (BM-MoBIIR) are compared with based MoBIIR(GN-MoBIIR). Results demonstrated that the Broyden's method based Jacobian calculator of BM-MoBIIR computes the Jacobian matrix more faster compared to the Jacobian calculator in Gauss-Newton algorithm. Conductivity reconstruction studies show that the BM-MoBIIR reconstructs the conductivity images from the simulated data five times faster than the GN-MoBIIR.

References

1. Webster, J.G.: *Electrical Impedance Tomography*. Adam Hilger Series of Biomedical Engineering. Adam Hilger, New York (1990)
2. Bera T. K., Nagaraju J.: Switching of a sixteen electrode array for wireless EIT system using a RF-based 8-bit digital data transmission technique. In: *Communications in Computer and Information Science*, Part I, CCIS 269, ObCom 2011, pp. 202–211, Springer (2012)
3. Bera T.K.: Studies on multi frequency electrical impedance tomography (EIT) to improve the impedance imaging for biomedical applications. Ph.D. Thesis, Department of Instrumentation and Applied Physics, Indian Institute of Science, Bangalore
4. Seo, J.K., Woo, E.J.: *Nonlinear Inverse Problems in Imaging*, 1st edn. Wiley (2012)
5. Bera T.K., Nagaraju J.: A MATLAB based boundary data simulator for studying the resistivity reconstruction using neighbouring current pattern. *J. Med. Eng.* **2013**, 15 (2013). Article ID 193578
6. Bera T.K., Nagaraju J.: Switching of the surface electrodes array in a 16-electrode EIT system using 8-bit parallel digital data. In: *IEEE World Congress on Information and*

- Communication Technologies 2011(WICT-2011), pp. 1288–1293, University of Mumbai, Mumbai (2011)
7. Bera, T.K., Nagaraju, J.: Studies on the thin film based flexible gold electrode arrays for resistivity imaging in electrical impedance tomography. *Measurement* **47**, 264–286 (2014)
 8. Bera T. K., Nagaraju J.: Electrical impedance tomography (EIT): a harmless medical imaging modality. In: *Research Developments in Computer Vision and Image Processing: Methodologies and Applications*, Chap. 13, pp. 224–262, IGI Global USA
 9. Graham, B.M.: Enhancements in Electrical Impedance Tomography (EIT) Image Reconstruction for 3D Lung Imaging, p. 25. Ph.D. thesis, University of Ottawa (April 2007)
 10. Bera, T.K., Nagaraju, J.: Surface electrode switching of a 16-electrode wireless EIT system using RF-based digital data transmission scheme with 8 channel encoder/decoder ICs. *Measurement* **45**, 541–555 (2012)
 11. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving conductivity image quality using block matrix-based multiple regularization (BMMR) technique in EIT: a simulation study. *J. Electr. Bioimpedance* **2**, 33–47 (2011)
 12. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J.: Improving image quality in electrical impedance tomography (EIT) using projection error propagation-based regularization (PEPR) technique: a simulation study. *J. Electr. Bioimpedance* **2**, 2–12 (2011)
 13. Bera T.K., Biswas S.K., Rajan K., Nagaraju J.: Image reconstruction in electrical impedance tomography (eit) with projection error propagation-based regularization (PEPR): a practical phantom study. In: *Lecture Notes in Computer Science, ADCONS 2011*, vol. 7135/2012, pp. 95–105. Springer (2012)
 14. Bera, T.K., Biswas, S.K., Rajan, K., Nagaraju, J., Improving the image reconstruction in electrical impedance tomography (EIT) with block matrix-based multiple regularization (BMMR): a practical phantom study. In: *IEEE World Congress on Information and Communication Technologies 2011(WICT-2011)*, pp. 1346–1351, University of Mumbai, Mumbai (2011)
 15. Bera T.K., Nagaraju J.: A FEM-based forward solver for studying the forward problem of electrical impedance tomography (EIT) with a practical biological phantom. In: *Advance Computing Conference (IACC 2009)*, pp. 1375–1381. IEEE, Patiala (2009)
 16. Reddy, J.N.: *An Introduction to the Finite Element Method*, 3rd edn., 2nd Reprint. TATA McGraw-Hill Pub. Co. Ltd, New York (2006)
 17. Biswas, S.K., Rajan, K., Vasu, R.M.: Accelerated gradient based diffuse optical tomographic image reconstruction. *Med. Phys.* **38**, 539 (2011)
 18. Bera T.K., Nagaraju, J.: Sensors for Electrical Impedance Tomography, The Measurement, Instrumentation, and Sensors Handbook, 2nd edn. In: Webster, J.G. (eds.) Chap. 61, pp. 61–1–61–30. CRC Press, Boca Raton (2014)
 19. Bera T.K., Nagaraju, J.: A multifrequency constant current source for medical electrical impedance tomography. In: *Proceedings of the IEEE International Conference on Systems in Medicine and Biology 2010*, pp. 290–295. IIT Kharagpur, India 16–18 Dec (2010)
 20. Bera T.K., Nagaraju, J.: A study of practical biological phantoms with simple instrumentation for electrical impedance tomography (EIT). In: *Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC2009)*, pp. 511–16. Singapore (2009)
 21. Bera, T.K., Saikia, M., Nagaraju, J.: A battery-based constant current source (Bb-CCS) for biomedical applications. In: *International Conference on Computing, Communication and Networking Technologies (ICCCNT 2013)* Germany, 22–25 April 2013
 22. Bera, T.K., Nagaraju, J.: A Lab VIEW based multifunction multifrequency electrical impedance tomography (Mfmf-EIT) instrumentation for flexible and versatile impedance imaging. In: *15th International Conference on Electrical Bio-Impedance (ICEBI) and 14th Conference on Electrical Impedance Tomography (EIT)*, Germany, vol. 216, 22–25 April 2013
 23. Bera, T.K., Nagaraju, J.: A battery based multifrequency electrical impedance tomography (BbMf-EIT) system for impedance, imaging of human anatomy. In: *15th International*

- Conference on Electrical Bio-Impedance (ICEBI) and 14th Conference on Electrical Impedance Tomography (EIT), Germany, vol. 216, 22–25 April 2013
24. Yorkey, T.J.: Comparing reconstruction methods for electrical impedance tomography. Ph.D. Thesis, University of Wisconsin, Madison, 1986 (WI 53706)
 25. Vauhkonen, M.: Electrical Impedance Tomography and Prior Information. Kuopio University Publications, Natural and Environmental Sciences (1997)
 26. Grootveld, C.J.: Measuring and modeling of concentrated settling suspensions using electrical impedance tomography. Ph.D. Thesis, Delft University of Technology, The Netherlands, (1996)
 27. Arridge, S.R.: Optical tomography in medical imaging. *Topical Rev. Inverse Prob.* **15**, R41–R93 (1999)
 28. Chan, T.F., Tai, X.C.: Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients. *J. Comput. Phys.* **193**(1), 40–66 (2004)
 29. Jing, Lei, Liu, Shi, Zhihong, Li, Meng, Sun: An image reconstruction algorithm based on the extended Tikhonov regularization method for electrical capacitance tomography. *Measurement* **42**(3), 368–376 (2009)
 30. Lionheart, W.R.B.: EIT reconstruction algorithms: pitfalls, Review Article, challenges. *Physiol. Meas.* **25**, 125–142 (2004)
 31. Brandstätter, B.: Jacobian calculation for electrical impedance tomography based on the reciprocity Principle. *IEEE Trans. Magn.* **39**(3), 1309–1312 (2003)
 32. Arridge, S.R., Schweiger, M.: A gradient-based optimisation scheme for optical tomography. *Opt. Express* **2**, 213–226 (1998)
 33. Paulsen, K.D., Meaney, P.M., Gilman, L.C.: Alternative Breast Imaging: Four Model-Based Approaches, The Springer International Series in Engineering and Computer Science, Chapter-5, 1st edn., p. 92. Springer (2004)
 34. Metherall P, Three dimensional electrical impedance tomography of the human thorax. Ph.D. Thesis, University of Sheffield, 1998
 35. Broyden, C.G.: A class of methods for solving nonlinear simultaneous equations. *Math. Comput.* **19**, 577 (1965)
 36. Broyden, C.G.: Quasi-Newton methods and their application to function minimization. *Math. Comput.* **21**, 368 (1967)
 37. Biswas, S.K.: Experimental and theoretical studies towards the development of a direct 3-d diffuse optical tomographic Imaging system. Ph.D. Thesis, Indian Institute of Science, Bangalore, 2012
 38. Branes, J.: An algorithm for solving nonlinear equations based on the secant method. *Comput. J.* **8**, 66 (1965)
 39. Dennis Jr, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs (1983)
 40. Gockenbach, M.S.: Lecture Series of MA5630-spring, 2003, Mathematical Science, Michigan Technological University, <http://www.math.mtu.edu/~msgocken/ma5630spring2003/lectures/>
 41. Bera, T.K., Nagaraju, J.: Studies and Evaluation of EIT Image Reconstruction in EIDORS with Simulated Boundary Data. In: International Conference on Soft Computing for Problem Solving (SocProS 2012), Institute of Engineering and Technology, JK Lakshmiapat University, Jaipur 28–30 Dec 2012

Predicting Total Number of Failures in a Software Using NHPP Software Reliability Growth Models

Poonam Panwar and A. K. Lal

Abstract For a software development project, management often faces the dilemma of when to stop testing the software and release it for operation. Estimating the remaining defects (or failures) in software can help test management to make release decisions. Several methods exist to estimate the defect content in software; among them are also a variety of software reliability growth models (SRGMs). SRGMs have underlying assumptions that are often violated in practice, but empirical evidence has shown that a number of models are quite robust despite these assumption violations. However it is often difficult to know which model to apply in practice. In the present study a method for selecting SRGMs to predict total number of defects in a software is proposed. The method is applied to a case study containing 3 datasets of defect reports from system testing of three releases of a large medical record system to see how well it predicts the expected total number of failures in a software.

Keywords Reliability testing · Software reliability growth models · Goodness of fit · Least squared estimation · Release time

Notation

$m(t)$ mean value function

$a(t)$ error content function

$b(t)$ error detection rate per error at time t

$N(t)$ random variable representing the cumulative number of software errors predicted by time t

P. Panwar (✉) · A. K. Lal
Thapar University, Patiala 147004, India
e-mail: rana.poonam1@gmail.com

A. K. Lal
e-mail: aklal@thapur.edu

1 Introduction

Several Software Reliability Growth Models (SRGMs) have been proposed in literature to estimate the reliability of a software system. Software reliability is one of the most important attributes of software quality and is closely related to being developed defects. It is expected to grow as defects are corrected and removed from the software. To estimate the remaining number of defects in a software system under test several approaches have been proposed in literature. SRGMs are being applied to guide test management in their release decisions. A number of estimation models exist to estimate the expected number of total defects (or failures) or the expected number of remaining defects (or failures) to make the release decisions. Static defect estimation models include capture–recapture models [1–6], curve-fitting methods, such as the Detection Profile Method and the Cumulative Method [5, 7], and experience-based methods [8, 9]. Software reliability growth models (SRGMs) [10–15] have also been used to estimate remaining failures. In these models the problem to decide release date is that they have underlying assumptions that are often violated in practice. Some of the assumption violations are:

- Performing functional testing rather than testing an operational profile.
- Varying test effort due to holidays and vacations.
- Imperfect repair and introduction of new errors.
- Calendar time is used instead of execution time.
- Using defect reports instead of failure reports.
- Partial or complete scrubbing of duplicate defect reports.
- Failure intervals those are not independent of each other, etc.

Despite the fact that empirical evidence has shown that many of the models are quite robust in practice, it is often difficult to decide which model to apply in light of these assumption violations. The “best model” can vary across systems, and even releases. One cannot expect to select one model and apply it successfully in later releases or other systems. The underlying problem is the complexity of the factors that interact while influencing software reliability.

This paper proposes a selection method that is able to determine the most appropriate best model(s) for estimating the total number of failures in a software. From this estimate, the expected number of remaining failures may be estimated and this estimate can help in making release decisions. For its validation the proposed method has been applied to a case study consisting of three releases of a large medical record system. The rest of the paper is arranged as under.

[Section 2](#) describes some well-known SRGMs. [Section 3](#) describes the method for model selection proposed in this study. The proposed method is next applied in [Sect. 4](#) on a case study. Conclusions based on present study are finally drawn in section.

2 Software Reliability Growth Models (SRGMs)

A Software Reliability Growth Model is one of the fundamental techniques to assess software reliability quantitatively. The Software Reliability Growth Model required having a good performance in terms of goodness-of-fit, predictability, and so forth. In order to estimate as well as to predict the reliability of software systems, failure data need to be properly measured by various means during software development and operational phases. Any software required to operate reliably must still undergo extensive testing and debugging. This can be a costly and time consuming process, and managers require accurate information about how software reliability grows as a result of this process in order to effectively manage their budgets and projects [16]. The effects of this process, by which it is hoped software is made more reliable, can be modeled through the use of Software Reliability Growth Models, hereafter referred to as SRGMs. Research efforts in software reliability engineering have been conducted over the past three decades and many software reliability growth models (SRGMs) have been proposed [17–19]. SRGMs can estimate the number of initial faults, the software reliability, the failure intensity, the mean time-interval between failures, etc. Ideally, these models provide a means of characterizing the development process and enable software reliability practitioners to make predictions about the expected future reliability of software under development. Such techniques allow managers to accurately allocate time, money, and human resources to a project, and assess when a piece of software has reached a point where it can be released with some level of confidence in its reliability.

There are two main types of software reliability growth models: Failure rate models and the Non Homogeneous Poisson Process Models (NHPP). The failure rate group of models is used to study the program failure rate (rate at which failures occur in a certain time interval $[t_1, t_2]$) per fault at the failure intervals. These models studies how failure rate changes at the failure time during the failure intervals. As the number of remaining faults changes, the failure rate of the program changes accordingly. Since the number of faults in the program is a discrete function, the failure rate of the program is also a discrete function with discontinuities at the failure times. The models included in this group are: Jelinski and Moranda Model, Schick and Wolverton, Jelinski Moranda Geometric Model, Moranda Geometric Poisson, Negative-binomial Poisson, Modified Schick and Wolverton and Goel and Okumoto imperfect debugging. The another class of models i.e. Non-Homogeneous Poisson Process (NHPP) group of models provides an analytical framework for describing the software failure phenomenon during testing. The NHPP model represents that the number of failures experienced up to time t is a non-homogeneous Poisson process $\{N(t), t \geq 0\}$. The main issue in the NHPP model is to estimate the mean value function ($m(t)$) of the cumulative number of failures experienced up to a certain point of time [20]. With different assumptions, the model will end up with a different mean value function. The NHPP model is based on the following assumptions:

- The failure process has an independent increment, i.e. the number of failures during the time interval $(t, t + s)$ depends on the current time t and the length of time interval s , and does not depend on the past history of the process.
- The failure rate of the process is given by

$$\begin{aligned} P\{\text{exactly one failure in}(t, t + s)\} &= P\{N(t + \Delta t) - N(t) = 1\} \\ &= \lambda(t)\Delta t + o(\Delta t), \end{aligned} \quad (1)$$

where $\lambda(t)$ is the intensity function.

- During a small interval Δt , the probability of more than one failure is negligible, that is,

$$P\{\text{two or more failures in}(t, t + \Delta t)\} = o(\Delta t). \quad (2)$$

- The initial condition is $N(0) = 0$.

On the basis of these assumptions, the probability that exactly n failures occurring during the time interval $(0, t)$ for the NHPP is given by:

$$\Pr\{N(t) = n\} = \frac{[m(t)]^n}{n!} e^{-m(t)} \quad n = 0, 1, 2, \dots, \quad (3)$$

where

$$m(t) = E[N(t)] = \int_0^t \lambda(s) ds \quad (4)$$

where $\lambda(t)$ is intensity function and $m(t)$ is non decreasing function.

A number of methods exist in literature for selection of optimal software reliability model to predict total number of failures in a software. Goel discussed the applicability and limitations of SRGMs during the software development life cycle in [10]. He proposed a step-by-step procedure for fitting a model and applied the procedure to a real-time command and control software system. His procedure selects an appropriate model based on an analysis of the testing process and a model's assumptions. Khoshgoftaar and Woodcock [21] proposed a method to select a reliability model among various alternatives using the log-likelihood function. They apply the method to the failure logs of a project. The method selected an S-shaped model as the most appropriate one. Lyu and Nikora in 1992 implemented GOF model selection criteria in their computer aided software reliability tool [22]. In 1996, Wood applied eight reliability models to a subset of software products with four releases to determine which model performed the best in predicting the number of residual defects. This study shows that SRGMs based on cumulative defects predict the number of remaining defects that are close to the number of defects reported after release. While Wood used correlation as the measure of GOF, there are other criteria for comparing software reliability models such as evaluating and comparing predictive validity [23]. Wood in 1997 proposed three approaches to accommodate model assumption violations with advantages and disadvantages discussed for each. The easiest approach is to ignore the

violations [24]. This keeps the model simple, but causes some loss of accuracy. Parameter estimation, however, may compensate. The second solution proposed involves modifying the data. This approach is easy to implement, because the standard models can be used with the modified data. Data modification, however, needs to be based on a mathematical model. A third approach is to derive new models to fit the test environment. These models are more complex. While they are more accurate, they are more difficult to derive and to apply. In Gaudoin et al. [25] the power of several of these statistical tests for GOF has been evaluated for a variety of reliability models including those based on an NHPP, on a Weibull distribution, and the Miranda model. In 2002 Stringfellow and Andrews [26] proposed an empirical method to predict total number of failures in a software that is later criticized by Andersson [27] in 2006. Sharma et al. [18] proposed a distance based approach in 2010 for ranking of various selected SRGMs. But it is found that no software reliability growth model is optimal for all contributing criteria. Due to all these reasons an approach is proposed here to determine total number of failures in software at early stage. The proposed approach is very interesting and has useful interpretation that may be used to predict the release of software.

3 Proposed Approach

A large number of software reliability models have been proposed in literature to select an optimal SRGM to predict the reliability, total number of failures and failure intensity of a software. Some of the points which must be kept in mind while developing a software reliability model selection techniques are:

- Since no model is expected to match company's development and test process exactly, the selected model should yield a curve fit that shows Goodness-Of-Fit value as high as possible.
- Data is usually grouped by weeks resulting in a smaller amount of data. A smaller amount of data means it will take longer for the predicted total number of failures to stabilize.
- Testing effort may vary from week to week. This is generally a problem with a small amount of data that is when data is only for a few test weeks.
- Models are usually based on execution time whereas most of the information available is based on calendar time. In view of these, none of the above techniques can be regarded fully satisfactory.
- When a software has been reasonably developed and is to be shipped we should choose a model that gives as accurate estimate of known past defects as possible.
- Should be able to predict its future defects in a reasonably accurate manner so that necessary precautions can be taken.
- It should provide guideline for next release of the software.

If the new software is released earlier, it may contain bugs which may start appearing in a very short span of time and thus, lower acceptance of product by the users. On the contrary, if one takes too much time to make release decision, it may result in wastage of time, money as well as resources.

3.1 Unknown Parameter Estimation Using Genetic Algorithm

The main problem in using SRGMs is to estimate their unknown parameters. There exists a number of parameter estimation techniques in literature like maximum likelihood estimation (MLE) and least square estimation (LSE) etc. We have estimated the parameters in proposed approach using Genetic Algorithm (GA). Genetic Algorithms (GA) are direct, parallel, stochastic method for global search and optimization, which imitates the evolution of the living beings, described by Charles Darwin. GA are part of the group of Evolutionary Algorithms (EA). The evolutionary algorithms use the three main principles of the natural evolution: reproduction, natural selection and diversity of the species, maintained by the differences of each generation with the previous. Genetic Algorithms works with a set of individuals, representing possible solutions of the task. The selection principle is applied by using a criterion, giving an evaluation for the individual with respect to the desired solution. The best-suited individuals create the next generation. The large variety of problems in the engineering sphere, as well as in other fields, requires the usage of algorithms from different type, with different characteristics and settings.

The mathematical formulation of parameter estimation problem is depicted in Eq. 5. Using the observed failure data (t_i, y_i) for $i = 1, 2, \dots, n$, one can use the mean value function $m(t_i)$ for any model to determine expected numbers of errors detected by time t_i for $i = n + 1, n + 2$, etc.

$$\max f = \sum_{i=1}^n (y_i - y_{i-1}) \cdot \log[m(t_i) - m(t_{i-1})] - m(t_n) \quad (5)$$

where

y_i = cumulative number of detected errors in a given time interval $(0, t_i)$

$i = 1, 2, \dots, n$ is the failure index

t_i = failure time index

$m(t_i)$ = total number of failures observed at time t_i according to the actual data.

The parameters are estimated using GA Tool of Matlab has been implemented on a laptop having Intel(R) Core(TM) 2 Duo 1.67 GHz. Following parameters have been used in our study:

- Population Type: Double Vector
- Population Size: 20
- Scaling Function: Rank
- Selection Function: Stochastic Function
- Reproduction: Elite Count
- Crossover Fraction: 0.8
- Crossover Function: Scattered
- Migration:
 1. Direction: Forward
 2. Fraction: 0.2
 3. Interval: 20
- Evaluate: Fitness and constraint function in serial
- Total number of trials = 10 for each problem

3.2 Predicting Total Number of Failures in Software Using Proposed Approach

Keeping above facts in view we propose alternative method for choosing a software reliability model which tries to take care of the issues listed above in an appropriate way. The proposed method works as under:

1. Record the cumulative number of failures found in system test at the end of each week.
2. Decide possible Models that are likely to fit the data.
3. Fit selected models to the available failure data to its expected specified interval (that about say 60 % of test plan has been executed) and use a software tool such as Matlab to determine unknown parameters for each models using Genetic Algorithm or use MLE or LSE parameter estimation technique, values of R-Square (square of the correlation between the response values and the predicted response values), RMSE (Root mean squared error) and expected number of cumulative failures based on each of these models.
4. Set threshold value for R-square and RMSE according to the observed values.
5. Select the models that have RMSE less than the threshold value of RMSE and R-square value greater than the threshold value of R-square.
6. Out of the models which qualifies step 5, select those models that have estimate greater than the actual value of failures and the model is stable (as per Wood’s [20] recommendation a model is stable if the prediction in week i is within 10 % of the prediction in week $i-1$).
7. Select those models for which the value of $a(t)$ (error content function) satisfies following condition:

$$(m(t) * 2) + 2 > a(t) > m(t) + (0.1 * m(t)) \tag{6}$$

8. If more than one qualifies the criteria than take the average of the value of the error content function for all the selected models to predict total number of failures in the system.

4 Case Study

4.1 Data

The failure data come from three releases of a large medical record system, consisting of 188 software components. Each component contains a number of files. Initially, the software consisted of 173 software components. All three releases added functionality to the product. Over the three releases, 15 components were added. Between three and seven new components were added in each release. Many other components were modified in all three releases as a side effect of the added functionality. The tracking database records the following attributes for defect reports:

- Defect report number
- Release identifier
- Phase in which defect occurred (development, test, post-release)
- Test site reporting defect
- Defective entity [code component(s), type of document by subsystem]
- Whether the component was new for a release
- The report date
- The “drop” in which the defect occurred.

Software is released to testers in stages or “drops.” Developers work on drop “i p 1” when testers test drop “i.” each successive drop includes more of the functionality expected for a given release. Release 1 has three drops. Release 2 has two drops. Release 3 had one drop. In system test and after release, a defect report is actually a failure. This study uses failure data from the last drop of system test and after release. Table 1 shows cumulative number of failures by week in the last drop of system test for all three releases. The cumulative number of failures after release are also shown in the last row.

4.2 Application of Proposed Approach

The proposed method is applied to 60 % of the test plan data. Tables 2, 3, 4, 5, 6, 7, shows the data from three releases using all the sixteen SRGMs to predict the total number of failures.

Table 1 Failure data for all three releases

Test week	Cumulative number of failures		
	Release 1	Release 2	Release 3
1	28	90	9
2	29	107	14
3	29	126	21
4	29	145	28
5	29	171	53
6	37	188	56
7	63	189	58
8	92	190	63
9	116	190	70
10	125	190	75
11	139	192	76
12	152	192	76
13	164	192	77
14	164	192	
15	165	203	
16	168	203	
17	170	204	
18	176		
Post release	231	245	83

Table 2 Predicted total number of failures for release 1 using 60 % of data for all the sixteen SRGMS

Model name	R-square	RMSE	Estimated total failures/error content function (a(t))	Estimated Failures when only 60 % test plan is executed
Zeng Teng Pham	0.96	13.10	154.39	139.29
Yamda Rayleigh	0.90	16.89	239.30	139.16
Yamda imperfect debugging model II	0.92	14.29	38,469.95	143.35
Yamda imperfect debugging model I	0.93	13.92	180.39	149.54
Yamda exponential	0.84	21.21	16,120.00	122.79
PZ model	0.90	18.38	612.85	136.35
PNZ model	0.92	13.92	348.10	144.66
Pham Zhang IFD	0.90	14.66	830.70	143.65
Musa Okumoto	0.84	18.68	23,610.00	122.86
Modified Duane	0.91	15.05	183,700.00	142.67
Logistic growth	0.95	11.44	291.20	145.32
Inflection S shaped	0.92	13.92	347.80	144.64
Gompert	0.94	11.93	15,070.00	147.81
Goel Okumoto	0.84	18.68	24,020.00	122.89
Generalized Goel	0.91	15.27	10,310.00	138.69
Delayed S shaped	0.90	14.66	830.70	143.65

Table 3 Final estimates by the SRGM model not rejected in release 1

Model	Estimate compare to 231	R-square value	Error
Yamda Rayleigh	239	0.90	+8

Table 4 Predicted total number of failures for release 2 using 60 % of data for all the sixteen SRGMS

Model name	R-square	RMSE	Estimated total failures/ error content function (a(t))	Estimated failures when only 60 % test plan is executed
Zeng Teng Pham	0.97	9.29	199.1796	195.5688
Yamda Rayleigh	0.71	24.43	198.6	183.7163
Yamda imperfect debugging model II	0.93	11.22	199.9233	195.2671
Yamda imperfect debugging model I	0.93	11.22	199.62	195.1034
Yamda exponential	0.94	11.34	195.0077	195.0077
PZ model	0.94	10.82	198.8839	196.7964
PNZ model	0.93	11.22	199.9193	195.2639
Pham Zhang IFD	0.80	18.07	186.1412	186.1412
Musa Okumoto	0.94	10.03	53.76	202.6516
Modified Duane	0.95	9.77	251.9	199.9116
Logistic growth	0.98	6.08	196.7	195.0576
Inflection S shaped	0.93	10.65	195	192.9295
Gompert	0.97	7.10	199.6	196.1563
Goel Okumoto	0.93	10.65	195.4	193.2791
Generalized Goel	0.95	9.42	212.5	197.881
Delayed S shaped	0.80	18.07	186.2	186.1414

Table 5 Final estimates by the SRGM model not rejected in release 2 taking RMSE threshold ≤ 11 and R-Square ≥ 0.93

Model	Estimate compare to 245	R-square value	Error
Modified Duane	251	0.95	+6

Table 2 shows that according to our process framework, where threshold for R-square is 0.89 and for RMSE is 17 than only one model Yamda Rayleigh will qualify all the criteria specified in proposed approach. Yamda Exponential, Musa Okumoto, Goel Okumoto, PZ Model, are rejected because they have not qualified the R-square and RMSE criteria. Then only ZT Pham and Yamda Rayleigh are the only stable models. ZT Pham is rejected because its error content function not satisfying the criteria given in Eq. 6. The predicted number of failures is 239 by

Table 6 Predicted total number of failures for release 3 using 60 % of data for all the sixteen SRGMS

Model name	R-square	RMSE	Estimated total failures/ error content function (a(t))	Estimated failures when only 60 % test plan is executed
Delayed S shaped	0.95	5.33	82.81	64.99
Generalized Goel	0.95	5.53	67.21	63.53
Goel Okumoto	0.93	6.14	761.80	66.96
Gompert	0.96	6.33	74.23	64.48
Inflection S shaped	0.96	5.00	66.28	63.20
Logistic growth	0.97	4.53	65.55	63.08
Modified Duane	0.94	6.11	158.70	65.83
Musa Okumoto	0.93	6.14	808.30	66.98
PNZ model	0.96	5.00	66.28	63.20
PZ model	0.95	7.27	160.79	64.75
Pham Zhang IFD	0.95	5.33	82.81	64.98
Yamda exponential	0.93	7.53	14210	61.12
Yamda imperfect debugging model I	0.93	6.72	739.80	66.91
Yamda imperfect debugging model II	0.93	6.71	74.04	68.21
Yamda Rayleigh	0.95	6.39	64.15	61.37
Zeng Teng Pham	0.93	10.78	35480	73.28

Table 7 Final estimates by the SRGM model not rejected in release 2 taking RMSE threshold ≤ 6 and R-Square ≥ 0.95

Model	Estimate compare to 83	R-square value	Error
Delayed S shaped	82	0.95	-1
Pham Zhang	82	0.95	-1

Yamda Rayleigh model, is compared to the actual number of failures, 231 and found that it is a close estimate to the actual failures.

Results show that the proposed method works well in choosing a SRGM that predicts total number of failures in a software.

5 Conclusion

This paper addresses the issue of optimal selection of software reliability growth models for prediction of total number of failures in a software. The decision has unrestricted choices in exploring the influences of various different sets of model selection criteria to the final decision. The proposed method is suitable for ranking

SRGMs based on a number of conflicting criteria taken all together. The selection method is robust in the sense that it is able to adjust to the differences in the data. This enables it to differentiate between the models: Different models were selected in the releases. At least one model of those investigated is acceptable by the time testing is complete enough to consider stopping and releasing the software. In the first and third release, the S-shaped models performed well in predicting the total number of failures. These two releases had defect data that exhibited an S-shape. The data in Release 2 was concave, rather than S-shaped. It is no surprise that the S-shaped models did not perform well on this data. The Modified Duane model, however, performed very well on the data from Release 2. (Other concave models under predict the total number of failures.) SRGMs may provide good predictions of the total number of failures or the number of remaining failures. Wood's empirical study (Wood 1996) has shown that predictions from simple models of cumulative defects based on execution time correlate well with field data. In given study, predictions from simple models based on calendar time correlate well with data from our environment. The selection method described in this paper helped in choosing an appropriate model. We would like to caution, though, that this does not guarantee universal success. It is always useful to employ complementary techniques for assessment.

References

1. Briand, L., El Emam, K., Freimut, B., Laitenberger, B., Laitenberger, O.: Quantitative evaluation of capture–recapture models to control software inspections. In: Proceedings of the Eighth International Conference on Software Reliability Engineering, Albuquerque, NM, pp. 234–244 (1997)
2. Eick, S., Loader, C., Long, M., Votta, L., VanderWeil, S.: Estimating software fault content before coding. In: Proceedings of the International Conference on Software Engineering, Melbourne, Australia, pp. 59–65 (1992)
3. Runeson, P., Wohlin, C.: An experimental evaluation of an experience-based capture–recapture method in software code inspections. *Empirical Softw. Eng. Int. J.* **3**(4), 381–406 (1992)
4. Vander Wiel, S., Votta, L.: Assessing software designs using capture–recapture methods. *IEEE Trans. Softw. Eng.* **19**(11), 1045–1054 (1993)
5. Wohlin, C., Runeson, P.: Defect content estimations from review data. In: Proceedings of the International Conference on Software Engineering, Kyoto, Japan, pp. 400–409 (1998)
6. Yang, M., Chao, A.: Reliability-estimation and stopping-rules for software testing based on repeated appearances of bugs. *IEEE Trans. Reliab.* **44**(2), 315–321 (1995)
7. Briand, L., El Emam, K., Freimut, B.: A comparison and integration of capture–recapture models and the detection profile method. In: Proceedings of the Ninth International Conference on Software Reliability Engineering, Paderborn, Germany, pp. 32–41 (1998)
8. Biyani, S., Santhanam, P.: Exploring defect data from development and customer usage on software modules over multiple releases. In: Proceedings of the Ninth International Conference on Software Reliability Engineering, Paderborn, Germany, pp. 316–320 (1998)
9. Yu, T., Shen, V., Dunsmore, H.: An analysis of several software defect models. *IEEE Trans. Softw. Eng.* **14**(9), 1261–1270 (1988)

10. Goel, A.L.: Software reliability models: assumptions, limitations, and applicability. *IEEE Trans. Reliab.* **11**(12), 1411–1421 (1985)
11. Goel, A.L., Okumoto, K.: A time dependent error detection model for software reliability and other performance measures. *IEEE Trans. Reliab.* **28**(3), 206–211 (1979)
12. Kececioglu, D.: *Reliability Engineering Handbook*, vol. 2. Prentice-Hall, Englewood Cliffs, NJ (1991)
13. Musa, J., Ackerman, A.: Quantifying software validation: when to stop testing. *IEEE Softw.* **6**, 19–27 (1989)
14. Musa, J., Iannino, A., Okumoto, K.: *Software Reliability: Measurement, Prediction, Application*. McGraw-Hill, New York (1987)
15. Yamada, S., Ohba, M., Osaki, S.: S-shaped reliability growth modeling for software error detection. *IEEE Trans. Reliab.* **32**(5), 475–478 (1983)
16. Zhang, X., Teng, X., Pham, H.: Considering fault removal efficiency in software reliability assessment. *IEEE Trans. Syst. Man. Cybern. Part A: Syst. Hum.* **33**(1) (2003)
17. Sharma, K., Garg, R., Nagpal, C.K., Garg, R.K.: Selection of optimal software reliability growth model using distance based approach. *IEEE Trans.* **59**(2), 266–276 (2012)
18. Garg, R., Sharma, K., Kumar, R., Garg, R.K.: Performance analysis of software reliability models using matrix method. *World Acad. Sci. Eng. Technol.* **71**, 31–38 (2010)
19. Duygulu, H.B., Tosun, O.: An algorithm for software reliability growth model selection. *IADIS International Conference Informatics* (2008)
20. Wood, A.: Predicting software reliability. *IEEE Comput.* **29**(11), 69–78 (1996)
21. Khoshgoftaar, T., Woodcock, T.: Software reliability model selection: a case study. In: *Proceedings of the Second International Symposium on Software Reliability Engineering*, pp. 183–191. IEEE Computer Society Press, Austin, TX (1991)
22. Lyu, M., Nikora, A.: CASREA: a computer-aided software reliability estimation tool. In: *Proceedings of the Fifth International Workshop on Computer-Aided Software Engineering*, Montreal, CA, pp. 264–275 (1992)
23. Iannino, A., Musa, J., Okumoto, K., Littlewood, B.: Criteria for software model comparisons. *IEEE Trans. Softw. Eng.* **10**(6), 687–691 (1984)
24. Wood, A.: Software reliability growth models: assumptions vs. reality. In: *Proceedings of the International Symposium on Software Reliability Engineering* **23**(11), 136–141 (1997)
25. Gaudoin, O., Xie, M., Yang, B.: A simple goodness-of-fit test for the power-law process, based on the Duane plot. *IEEE Trans. Reliab.* (2002)
26. Stringfellow, C., Andrews, A.A.: An empirical method for selecting software reliability growth models. *Empirical Softw. Eng.* **7**(4), 319–343 (2002)
27. Andersson, C.: An empirical method for selecting software reliability growth models. *Empirical Softw. Eng.* **7**, 161–182 (2007)
28. Huang, C.Y., Lyu, M.R., Kuo, S.Y.: A unified scheme of some non-homogenous Poisson process models for software reliability estimation. *IEEE Trans. Softw. Eng.* **29**(3), 261–269 (2003)
29. Musa, J.D., Okumoto, K.: A logarithmic Poisson execution time model for software reliability measurement. In: *Conference Proceedings of the 7th International Conference on Software Engineering*, pp. 230–237 (1983)
30. Yamada, S., Tokuno, K., Osaki, S.: Imperfect debugging models with fault introduction rate for software reliability assessment. *Int. J. Syst. Sci.* **23**, 2241–2252 (1992)
31. Pham, H.: Software reliability and cost models: perspectives, comparison and practice. *Eur. J. Oper. Res.* **149**, 475–489 (2003)
32. Pham, H., Nordmann, L., Zhang, X.: A general imperfect software debugging model with s-shaped fault detection rate. *IEEE Trans. Reliab.* **48**, 169–175 (1999)
33. Pham, H., Zhang, X.: An NHPP software reliability models and its comparison. *Int. J. Reliab. Qual. Saf. Eng.* **14**(3), 269–282 (1997)
34. Pham, H.: *System Software Reliability*. Springer, London (2006)

Design Optimization of Shell and Tube Heat Exchanger Using Differential Evolution Algorithm

Pawan Singh and Millie Pant

Abstract Shell and tube heat exchangers (STHE) are the most common type of heat exchangers widely used in various kinds of industrial applications. Cost minimization of these heat exchangers is of prime concern for designers as well as for users. Heat exchanger design involves processes such as selection of geometric and operating parameters. Generally, different exchangers geometries are rated to identify those that satisfy a given heat duty and a set of geometric and operational constraints. In the present study we have considered minimization of total annual cost as an objective function. The different variables used include shell internal diameter, outer tube diameter and baffle spacing for which two tube layout viz. triangle and square are considered. The optimization tool used is differential evolution (DE) algorithm, a nontraditional stochastic optimization technique. Numerical results indicate that, DE can be used effectively for dealing with such types of problems.

Keywords Shell and tube heat exchanger • Optimization • Differential evolution

1 Introduction

Heat exchangers are devices that facilitate heat transfer between two fluids at different temperatures. They are used in industrial process to recover heat between two process fluids. The shell-and-tube heat exchangers (STHE) are probably the

P. Singh (✉)

Department of Mechanical Engineering, Global Institute of Technology, Jaipur, India
e-mail: pawan.mechengg@gmail.com

M. Pant

Department of Applied Science and Engineering, Indian Institute of Technology,
Roorkee, India
e-mail: millifpt@iitr.ernet.in

most common type of heat exchangers applicable for a wide range of operating temperatures and pressures.

The design of STHes, including thermodynamic and fluid dynamic design, cost estimation and optimization, represents a complex process containing an integrated whole of design rules and empirical knowledge of various fields. There are many previous studies on the optimization of heat exchanger. Investigators have used different optimization techniques considering different objective functions like minimum entropy generation and minimum cost of STHes to optimize heat exchanger design.

Some studies have focused on a single geometric parameter like optimal baffle spacing while some have considered optimizing a variety of geometrical and operational parameter of STHes.

Strategies applied for solving such problems vary from traditional mathematical methods to sophisticated non-traditional optimization methods such as genetic algorithms (GA), differential evolution (DE), particle swarm optimization (PSO) etc. Relevant literature can be found in [1–7].

In the present study we have considered the cost optimization of STHe and have employed DE for solving the optimization model.

The remaining of the paper is divided into three sections. In Sect. 2, we discuss the DE optimization technique, in Sect. 3, mathematical model considered is discussed in brief. In Sect. 4, we give numerical results and finally in Sect. 5, the conclusions based on the present study are given.

2 Differential Evolution (DE)

DE, an evolutionary algorithm, was proposed by Storn and Price [8, 9]. The main operators of DE are *mutation, crossover and selection* to guide the search process. The algorithm uses mutation operation as a search mechanism; crossover operation is applied to induce diversity and selection operation is to direct the search toward the potential regions in the search space.

DE starts with a set of solutions, which is randomly generated when no preliminary knowledge about the solution space is available. This set of solution is called population.

Let $P^G = \{X_i^G, i = 1, 2, \dots, NP\}$ be the population at any generation G which contain NP individuals where an individual can be defined as a D dimensional vector such as $X_i^G = (x_{1,i}^G, x_{2,i}^G, \dots, x_{D,i}^G)$.

For basic DE (DE/rand/1/bin) mutation, crossover and selection operations are defined as below:

- i. **Mutation:** For each target vector X_i^G , mutant vector V_i^{G+1} is defined by:

$$V_i^G = X_{r_1}^G + F(X_{r_2}^G - X_{r_3}^G) \tag{1}$$

Where $r_1, r_2, r_3, r_4, r_5 \in 1, 2, \dots, NP$ are randomly chosen integers, distinct from each other and also different from the running index i . F is a real and constant factor having value between $[0, 2]$ and controls the amplification of differential variation $(X_{r_2}^G - X_{r_3}^G)$.

ii. **Crossover:** Crossover is introduced to increase the diversity of perturbed parameter vectors $V_i^G = (v_{1,i}^G, v_{2,i}^G \dots v_{D,i}^G)$.

Let $U_i^G = (u_{1,i}^G, u_{2,i}^G \dots, u_{D,i}^G)$ as the trial vector then U_i^G is defined as:

$$u_{j,i}^G = \begin{cases} v_{j,i}^G, & \text{if } Cr < rand(0, 1) \forall j = jrand \\ x_{j,i}^G & \text{otherwise} \end{cases} \tag{2}$$

$rand(0, 1)$ is uniform random number between 0 and 1; Cr is the crossover constant takes values in the range $[0, 1]$ and $jrand \in 1, 2, \dots, D$; is the randomly chosen index

iii. **Selection:** It decides which vector (X_i^G or U_i^G) should be a member of next generation $G + 1$. During the selection operation we generate a new population $P^{G+1} = \{X_i^{G+1}, i = 1, 2, \dots, NP\}$ for next generation $G + 1$ by choosing the best vector between trial vector and target vector.

$$X_i^{G+1} = \begin{cases} U_i^G, & \text{if } f(U_i^G) < f(X_i^G) \\ X_i^G & \text{otherwise} \end{cases} \tag{3}$$

Thus the points obtained after selection are either better or at par with the points of the previous generation.

3 Mathematical Model

A. Objective Function

Total cost C_t is taken as the objective function, which includes capital investment (C_i), energy cost (C_e), annual operating cost (C_o) and total discounted operating cost (C_{tod}) [10]. This optimization function is subjected to design variables d_0, D_s and B .

$$\text{Minimize } C_t = C_i + C_{tod}$$

- (a) Capital investment includes cost of purchase, instalment and piping for heat exchanger and shell side and tube side pumps multiplied by annualization factor.

$$C_i = A_f(C_{ps} + C_{pt} + C_{he})$$

Capital cost is calculated by using Goff's correlation and it depends on exchanger surface area (A).

$$C_i = a_1 + a_2 A^{a_3}$$

Where $a_1 = 8000$, $a_2 = 259.2$ and $a_3 = 0.93$ for stainless steel is used as material of construction for both shell and tubes.

- (b) Total operating cost is cost of energy required to operate pumps. This is given by

$$C_o = PC_e H$$

And total discounted total cost is given by

$$C_{od} = \sum_{x=1}^{ny} \frac{C_o}{(1+i)^x}$$

B. Heat Transfer equations

Heat transfer between fluids in shell and tube is due to convection currents in liquids and conduction in wall of tube. But as steel is highly conducting material and thickness of tube walls is assumed to be very small, resistance offered to conduction heat transfer in tube walls is ignored. Depending on flow pattern, tube side heat transfer coefficient is calculated by following correlation.

$$h_t = \frac{k_t}{d_i} \left[3.657 + \frac{0.0677(Re_t Pr_t \left(\frac{d_i}{L}\right))^{1.33^{1/3}}}{1 + 0.1 Pr_t (Re_t \left(\frac{d_i}{L}\right))^{0.3}} \right]$$

(if $Re_t < 2,300$ [10])

$$h_t = \frac{k_t}{d_i} \left[\frac{\left(\frac{f_t}{8}\right) (Re_t - 1000) Pr_t}{1 + 12.7 \left(\frac{f_t}{8}\right)^{1/2} (Pr_t^{2/3} - 1)} \left(1 + \frac{d_i}{L}\right)^{0.67} \right]$$

(if $2300 < Re_t < 10,000$ [10])

$$h_t = 0.027 \frac{k_t}{d_0} Re_t^{0.8} Pr_t^{1/3} \left(\frac{\mu_t}{\mu_{wt}} \right)^{0.14}$$

(If $Re_t > 10,000$ [10])

Where f_t is Darcy friction factor given by

$$f_t = (1.82 \log 10^{Re_t} - 1.64)^{-2}$$

Re_t is Reynolds number which is given by

$$Re_t = \frac{\rho_t v_t d_i}{\mu_t}$$

Flow velocity at tube side is given by

$$v_t = \frac{m_t}{(\pi/4) d_i^2 \rho_t} \left(\frac{n}{N_t} \right)$$

N_t is the number of tubes and n is the number of tube passes. Total number of tubes (N_t) is depended on shell diameter and tube diameter. These can be approximated by the equation

$$N_t = C \left(\frac{D_s}{d_0} \right)^{n_1}$$

C and n_1 are coefficients that take values according to flow arrangement and number of passes. These coefficients are given Table 1 for different flow arrangements.

Tube side prandtl number (Pr_t) is given by

$$Pr_t = \frac{\mu_t C_{pt}}{k_t}$$

Also $d_i = 0.8 d_o$

Kerns formulation [11] is used to find shell side heat transfer coefficient of segmental baffle shell and tube heat exchanger

$$h_s = 0.36 \frac{k_t}{d_e} Re_s^{0.55} Pr_s^{1/3} \left(\frac{\mu_s}{\mu_{wts}} \right)^{0.14}$$

Where d_e is shell hydraulic diameter and computed as

$$d_e = \frac{4(S_t^2 - (\pi d_0^2/4))}{\pi d_0}$$

(for square pitch)

Table 1 Value of C and n_1 coefficients

No. of passes	Triangle tube pitch $S_t = 1.25 d_0$		Square tube pitch $S_t = 1.25 d_0$	
	C	(n_1)	C	(n_1)
1	0.319	2.142	0.215	2.207
2	0.249	2.207	0.156	2.291
4	0.175	2.285	0.158	2.263
6	0.0743	2.499	0.0402	2.617
8	0.0365	2.675	0.0331	2.643

$$d_e = \frac{4(0.43S_t^2 - (0.5\pi d_0^2/4))}{0.5\pi d_0}$$

(for triangular pitch)

Cross sectional area normal to flow direction is determined by

$$A_s = D_s B \left(1 - \frac{d_0}{S_t}\right)$$

$$v_s = \frac{m_s}{\rho_s A_s}$$

Reynolds number for shell side follows,

$$Re_s = \frac{m_s d_e}{A_s \mu_s}$$

Prandtl number for shell side follows,

$$Pr_s = \frac{\mu_s C_{ps}}{k_s}$$

The overall heat transfer coefficient (U) depends on both the tube side and shell side heat transfer coefficients and fouling resistances are given by

$$U = \frac{1}{(1/h_s) + R_{fs} + (d_0/d_i)(R_{ft} + (1/h_t))}$$

Considering the cross flow between adjacent baffle, the logarithmic mean temperature difference (LMTD) is determined by,

$$LMTD = \frac{(T_{hi} - T_{co}) - (T_{ho} - T_{ci})}{\log((T_{hi} - T_{co})/(T_{ho} - T_{ci}))}$$

The correction factor F for the flow configuration involved is found as a function of dimensionless temperature ratio for mostflow configuration of interest, [12, 13]

$$F = \frac{\sqrt{R^2 + 1}}{R - 1} * \frac{\ln((1 - P)/(1 - PR))}{\ln((2 - PR - P + P\sqrt{R^2 + 1})/(2 - PR - P - P\sqrt{R^2 + 1}))}$$

Where R is a coefficient given by,

$$R = \frac{(T_{hi} - T_{ho})}{(T_{co} - T_{ci})}$$

And P is the efficiency given by,

$$P = \frac{(T_{co} - T_{ci})}{(T_{hi} - T_{ci})}$$

Considering overall heat transfer coefficient, the heat exchanger surface area (A) is computed by,

$$A = \frac{Q}{U * F * LMTD}$$

For sensible heat transfer

$$Q = m_h C_{ph}(T_{hi} - T_{ho}) = m_c C_{pc}(T_{co} - T_{ci})$$

Based on total heat exchanger surface area (A) the necessary tube length (L) is,

$$L = \frac{A}{\pi d_0 t}$$

C. Pumping Power

The power required to pump the fluid into the shell and tube heat exchanger is a function of the pressure drop allowance which is actually the static fluid pressure which may be expended to drive the fluid through the exchanger. There is a very close physical and economic affinity between pressure drop and heat transfer for every type of heat exchanger. Increasing the flow velocity will result in rise of heat transfer coefficient (for a constant heat capacity in a heat exchanger). This results in a compact exchanger design and reduced investment cost. But we cannot neglect the fact that increase in flow velocity will cause more pressure drop which will result in additional running cost. For this reason pressure drop is considered with heat transfer in order to find best solution for the system.

Tube side pressure drop includes distributed pressure drop along the tube length and concentrated pressure losses in elbows and in the inlet and outlet nozzle [11]

$$\Delta P_t = \Delta P_{tubelength} + \Delta P_{tubelbow}$$

$$\Delta P_t = \frac{\rho_t v_t^2}{2} \left(\frac{L}{d_i} f_t + p \right) n$$

Different values of constant p are considered by different authors. Kern [13] assumed $p = 4$, while Sinnott et al. [14] assume $dp = 2.5$.

The shell side pressure drop is,

$$\Delta P_s = f_s \left(\frac{\rho_s v_s^2}{2} \right) \left(\frac{L}{B} \right) \left(\frac{D_s}{D_e} \right),$$

Where,

$$f_s = 2b_0 Re_s^{-0.15}$$

And $b_0 = 0.72$ [15] valid for $Re_s < 40,000$

Considering pumping efficiency (η), pumping power computed by,

$$P = \frac{1}{\eta} \left(\frac{m_t}{\sigma_t} \Delta P_t - \frac{m_s}{\sigma_s} \Delta P_s \right)$$

D. Design constraints for feasible design

Design of heat exchanger is constrained by various factors. Those can be classified into geometric and operating constraints.

Operating constraints

Maximum allowed pressure drop on both shell and tube side of heat exchanger. These pressure drops are directly proportional to maximum pump capacity available. Upper bounds of pressure drops are given by

$$\Delta P_t \leq \Delta P_{tmax}$$

$$\Delta P_s \leq \Delta P_{smax}$$

Velocity range allowed for both shell and tube sides. Velocity of fluid above certain value can cause erosion or flow induced tube vibrations and lower velocities can cause fouling. These bounds are given by

$$V_{tmin} \leq V_t \leq V_{tmax}$$

$$V_{smin} \leq V_s \leq V_{smax}$$

Recommended velocity by Sinnott on tube side are 1–2.5 m/s and 0.3–1 m/s on shell side.

Geometric constraints

Length and diameter of shell and tube heat exchanger can be restricted due to space constraints.

$$D_s \leq D_{s\max}$$

$$L_t \leq L_{t\max}$$

In case of baffle spacing, higher spacing can cause bypassing, reduced cross flow along with low heat transfer coefficient and lower spacing can lead to higher heat transfer coefficient in expense of higher pressure drops in shell side fluid

$$B_{\min} \leq B \leq B_{\max}$$

4 Numerical Results and Discussions

A. Parameter settings of DE

DE has certain control parameters which are to be set by the user. In the present study the following parameter setting is considered:

- Population size—50
- Scale factor F—0.5, 0.8
- crossover rate Cr—0.5, 0.9
- maximum number of iterations—100

B. Experimental settings

- We executed the algorithm 50 times and recorded the mean value.
- Programming language used is DEVC++.
- Random numbers are generated using *rand()* the inbuilt function of DEVC++.
- The effectiveness of the present approach using DE is assessed by analyzing the following case study: 1.44 (MW) duty, kerosene crude oil exchanger [16].

C. Results and comparison

In this section, numerical results are given in Tables 1 and 2. In Table 1 results are given on the basis of different parameter settings (denoted as 1,2,3 in Table 2) of DE. The values given by setting 2 (Cr = 0.9, F = 0.5) are optimum. Hence, we can see that parameter setting 2 (Cr = 0.9, F = 0.5) perform better in the comparison of other settings.

We observed that the numerical results obtained using DE are either better or at par with the results available in literature. It was observed that DE was able to achieve the optimal design variables successfully (Table 3).

Table 2 Results by DE with different parameter settings

	1 Cr = 0.5, F = 0.5	2 Cr = 0.9, F = 0.5	3 Cr = 0.9, F = 0.8
L (m)	1.43	1.35	1.65
d_0 (m)	0.0145	0.013	0.0137
B (m)	0.1032	0.1045	0.1082
D_s (m)	0.563	0.55	0.57
S_t (m)	0.172	0.0153	0.0149
cl (m)	0.0033	0.0031	0.0034
N_t	740	820	802
v_t (m/s)	0.96	0.98	0.975
Re_t	2870	2600	2785
Pr_t	55.2	55.2	55.2
h_t (W/m ² K)	1224	1240	1246
f_t	0.0457	0.046	0.0449
ΔP_t (Pa)	17,220	19,000	18,450
a_s (m ²)	0.0123	0.0121	0.0129
d_c (m)	0.0137	0.012	0.0150
v_s (m/s)	0.53	0.55	0.545
Re_s	13,744	12,247	12,720
Pr_s	7.5	7.5	7.5
h_s (W/m ² K)	1373	1402	1418
f_s	0.432	0.345	0.335
ΔP_s (Pa)	25,722	26,225	26,012
U (W/m ²)	425.4	438.7	439.5
A(m ²)	46.2	44.3	45.7

Table 3 Comparison of results with GA and PSO

	GA [10]	PSO [10]	DE
C_i	17599	16707	15808
C_o	440	523.3	528.5
C_{od}	2704	3215.6	3242.22
C_{tot}	20303	19922.6	18042.18

5 Conclusion

Heat exchangers are an integral component of all thermal systems. Their designs should be adapted well to the applications in which they are used; otherwise their performances will be deceiving and their costs excessive. Heat exchanger design can be a complex task, which requires a suitable optimization technique for a proper solution. The present work shows the effectiveness of differential evolution optimization algorithm. This technique can be easily modified to suit optimization of various thermal systems.

References

1. Ozkol, G., Komurgoz, I.: Determination of the optimum geometry of the heat exchanger body via a genetic algorithm. *Int. J. Heat Mass Transf.* **48**, 283–296 (2005)
2. Hilbert, R., Janiga, G., Baron, R., Thevenin, D.: Multi objective shape optimization of a heat exchanger using parallel genetic algorithm. *Int. J. Heat Mass Transf.* **49**, 2567–2577 (2006)
3. Xie, G.N., Sunden, B., Wang, Q.W.: Optimization of compact heat exchangers by a genetic algorithm. *Appl. Therm. Eng.* **28**, 895–906 (2008)
4. Ponce-Ortega, J.M., Serna-Gonzalez, M., Jimenez-Gutierrez, A.: Use of genetic algorithms for the optimal design of shell-and-tube heat exchangers. *Appl. Therm. Eng.* **29**, 203–209 (2009)
5. Ponce-Ortega, J.M., Serna-Gonzalez, M., Jimenez-Gutierrez, A.: Design and optimization of multipass heat exchangers. *Chem. Eng. Process.* **47**, 906e913 (2008)
6. Saffar-Avval, M., Damangir, E.: A general correlation for determining optimum baffle spacing for all types of shell-and-tube exchangers. *Int. J. Heat Mass Transf.* **38**, 2501–2506 (1995)
7. Soltan, B.K., Saffar-Avval, M., Damangir, E.: Minimizing capital and operating costs of shell-and-tube condensers using optimum baffle spacing. *Appl. Therm. Eng.* **24**, 2801–2810 (2004)
8. Storn, R., Price, K.: Differential evolution—A simple and efficient adaptive scheme for global optimization over continuous spaces. Technical report TR-95-012, International Computer Science Institute (1995)
9. Storn, R., Price, K.: DE: a simple evolution strategy for fast optimization. *Dr. Dobb's J.* **18–24**, **78**, (1997)
10. Caputo, A.C., Pelagagge, P.M., Salini, P.: Heat exchanger design based on economic optimization. *Appl. Therm. Eng.* **28**, 1151–1159 (2008)
11. Kern, D.Q.: *Process Heat Transfer*. McGraw-Hill, New York (1950)
12. Fraas, A.P.: *Heat Exchanger Design*, 2nd edn. Wiley, New York (1989)
13. Ohadi, M.M.: *The Engineering Handbook*. CRC, Florida (2000)
14. Sinnott, R.K., Coulson, J.M., Richardson, J.F.: *Chemical Engineering Design*, vol. 6. Butterworth-Heinemann, Boston (1996)
15. Peters, M.S., Timmerhaus, K.D.: *Plant Design and Economics for Chemical Engineers*. McGraw-Hill, New York (1991)
16. Patel, V.K., Rao, R.V.: Design optimization of shell-and-tube heat exchanger using particle swarm optimization technique. *Appl. Therm. Eng.* **30**, 1417–1425 (2010)
17. Taal, M., Bulatov, I., Klemes, J., Stehlik, P.: Cost estimation and energy price forecast for economic evaluation of retrofit projects. *Appl. Therm. Eng.* **23**, 1819–1835 (2003)

Compression of Printed English Characters Using Back Propagation Neural Network

Sunita, Vaibhav Gupta and Bhupendra Suman

Abstract In this paper, image data compression algorithm is presented using back propagation neural networks. Back propagation is used to compress printed English characters by training a net to function as an auto associative net (the training input vector and the target output vector are the same) with fewer hidden units than there are in the input or output units. The input and the output data files are formed in +1 and -1 form. The network parameters are adjusted using different learning rates and momentum factors. Mainly, the input pixels are used as target values so that assigned mean square error (MSE) can be obtained, and then the hidden layer output will be the compressed image. The proposed algorithm has been implemented in MATLAB to simulate the algorithm for the English characters A, B, C, D, E. The results obtained, such as compression ratio, mean square error, number of epoch for different learning rates and momentum factors are presented in this paper. Hebbian learning rule and Delta learning rules are used to train the network. Sigmoidal function, binary sigmoidal function and bipolar sigmoidal function are used in feed forward net and back propagation net respectively.

Keywords Compression · Back propagation neural network · Hebbian learning rule · Sigmoidal function and delta function

Sunita (✉) · V. Gupta
IRDE, DRDO, Bhavan, India
e-mail: sunnidma@yahoo.com

V. Gupta
e-mail: vaibhav.drdo@gmail.com

B. Suman
DEAL, DRDO, Bhavan, India

1 Introduction

The main objective of data compression is to decrease the redundancy of the data which helps in increasing the capacity of storage and efficient transmission. Data compression aids in decreasing the size in bytes of a digital data without degrading the quality of the data to a desirable level. The reduction in file size allows more data to be stored in a given amount of disk or memory space. Research activities on neural networks for image compression do exist in many types of networks such as—Multi Layer perceptron (MLP) [1–3], Hopfield [4], Self-Organizing Map (SOM). Artificial neural network models are specified by network topology and learning algorithms [5–7]. Network topology describes the way in which the neurons are interconnected and the way in which they receive input and output. Learning algorithms specify an initial set of weights and indicate how to adapt them during learning in order to improve network performance.

In this paper, network has an input layer, a hidden layer and an output layer. The input layer is connected to the hidden layer and the hidden layer is connected to the output layer through interconnection weights. The increase in the number of hidden layers results in computational complexity of the network. As a result, the time required for convergence and to minimize the error can be very high. In this paper, single hidden layer with fewer hidden units than there are in the input or output units are taken for neural network topology to reduce the complexity of the algorithm. The algorithm is simulated for the compression and decompression of the printed alphabetic characters (A–E) which are taken in form of 9×7 blocks. In these 9×7 blocks, if the pattern of character is recognized then it is +1, and if the pattern is not recognized, it is -1 for input and output files.

2 Proposed Back-Propagation Neural Network

Back-Propagation is a multi-layer forward network using extend gradient-descent based delta-learning rule. Back propagation provides a computationally efficient method for changing the weights in a feed forward network, with differentiable activation functions units, to learn a training set of input–output examples. Being a gradient descent method it minimizes the total squared error of the output computed by the net. The aim of this network is to train the net to achieve a balance between the ability to respond correctly to the input patterns that are used for training and the ability to provide good responses to the input that are similar.

2.1 Architecture

A multilayer feed forward back propagation network having one input layer, one hidden layer and one output layer is considered in this paper (Fig. 1). The input

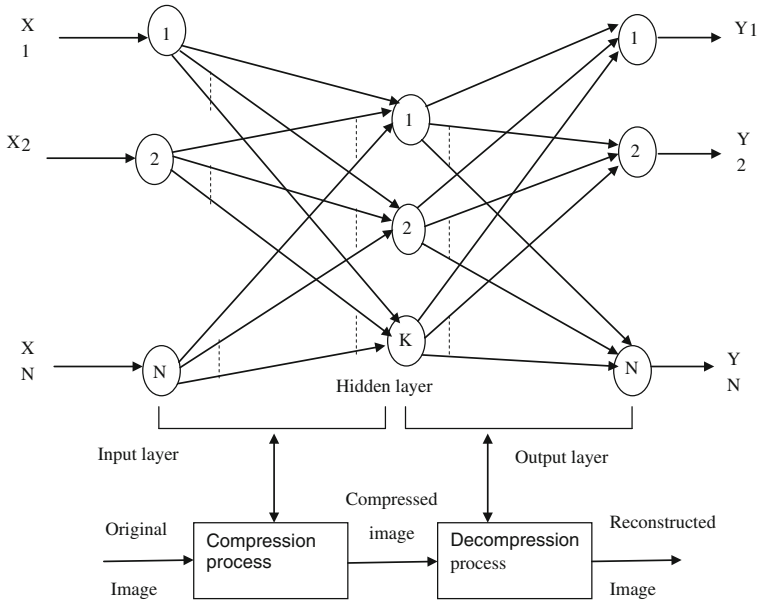


Fig. 1 BPNN image compression system

layer is connected to the hidden layer and the hidden layer is connected to the output layer by means of interconnection weights. The increase in the number of hidden layers results in the computational complexity of the network. As a result, the time taken for convergence and to minimize the error may be very high.

2.2 Training Algorithm

The various parameters used in the training algorithm are as follows:

- x: Input training vector
- $x = (x_1, \dots, x_i, \dots, x_n)$
- t: Output target vector
- $t = (t_1, \dots, t_k, \dots, t_m)$
- δ_k = error at output unit y_k
- δ_j = error at hidden unit z_j

- α = learning rate
- V_{oj} = bias on hidden unit j
- z_j = hidden unit j
- w_{ok} = bias on output unit k
- y_k = output unit k

The training algorithm used in the back propagation network involves following four stages:

2.2.1 Initialization of Weights

- Step 1:** Initialize weight to small random values
- Step 2:** While stopping condition is false, do step 3–10
- Step 3:** For each training pair do Steps 4–9

2.2.2 Feed Forward

- Step 4:** Each input receives the input signal x_i and transmits this signal to all units in the layer above i.e. hidden units
- Step 5:** Each hidden unit (z_j , $j = 1, \dots, p$) sums its weighted input signals using Hebbian learning rule and binary sigmoidal function

$$Z_{-inj} = v_{oj} + \sum_{i=1}^n x_i v_{ij}$$

applying identity function as activation function

$$Z_j = f(z_{-inj})$$

and sends this signal to all units in the layer above i.e. output units.

- Step 6:** Each output unit (y_k , $k = 1, \dots, m$) sums its weighted input signal using Hebbian learning rule and binary sigmoidal function

$$y_{-ink} = w_{ok} + \sum_{j=1}^p z_j w_{jk}$$

and applies its identity function as an activation function to calculate the output signals

$$y_k = f(y_{-ink})$$

2.2.3 Back Propagation of Errors

Step 7: Each output ($y_k, k = 1, \dots, m$) receives a target pattern corresponding to an input pattern, error information term is calculated as follows using delta learning rule and bipolar sigmoidal function

$$\delta_k = (t_k - y_k) f(y_{-ink})$$

Step 8: Each hidden unit ($z_j, j = 1, \dots, n$) sums its delta inputs from units in the layer above using Hebbian learning rule and bipolar sigmoidal function

$$\delta_{-inj} = \sum_{k=1}^m \delta_j W_{jk}$$

The error information term is calculated as

$$\delta_j = \delta_{-inj} f(z_{-inj})$$

2.2.4 Updation of Weight and Biases

Step 9: Each output unit ($y_k, k = 1, \dots, m$) updates its bias and weights ($j = 0, \dots, p$)

The weight correction term is given by

$$\Delta W_{jk} = \alpha \delta_k z_j$$

and the bias correction term is given by

$$\Delta W_{ok} = \alpha \delta_k$$

Therefore, $W_{jk}(\text{new}) = W_{jk}(\text{old}) + \Delta W_{jk}$, $W_{ok}(\text{new}) = W_{ok}(\text{old}) + \Delta W_{ok}$

Each hidden unit ($z_j, j = 1, \dots, p$) updates its bias and weights ($i = 0, \dots, n$).

The weight correction term

$$\Delta V_{ij} = \alpha \delta_j x_i$$

The bias correction term

$$\Delta V_{oj} = \alpha \delta_j$$

Therefore, $V_{ij}(\text{new}) = V_{ij}(\text{old}) + \Delta V_{ij}$, $V_{oj}(\text{new}) = V_{oj}(\text{old}) + \Delta V_{oj}$

Step 10: Test the stopping condition.

The stopping condition may be the minimization of the errors, numbers of epochs etc.

3 Selection of Parameters

For the efficient operation of the back propagation network, it is necessary for the appropriate selection of the parameters used for training. There are a number of different parameters that must be decided upon when designing a neural network. Among these parameters are the number of layers, the number of neurons per layer, the number of training iterations, etc. Some of the more important parameters in terms of training and network capacity are the number of hidden neurons, the learning rate and the momentum factor. The initial value assignments are discussed in this section. The details of these parameters are as follows:

3.1 Number of Neurons in the Hidden Layer

Hidden neurons are the neurons that are neither in the input layer nor the output layer. These neurons are essentially hidden from view, and their number and organization can typically be treated as a black box to people who are interfacing with the system. Using additional layers of hidden neurons enables greater processing power and system flexibility. This additional flexibility comes at the cost of additional complexity in the training algorithm. Having too many hidden neurons is analogous to a system of equations with more equations than there are free variables: the system is over specified, and is incapable of generalization. Having too few hidden neurons, conversely, can prevent the system from properly fitting the input data, and reduces the robustness of the system. In this BPNN algorithm, we are taking one hidden layer. For simulation result, different numbers of hidden neurons in one hidden layer are taken as 24, 14, and 10.

3.2 Learning Rate

Learning rate is the training parameter that controls the size of weight and bias changes during learning. In this algorithm, we have taken two different values of learning rate parameter as 0.3 and 0.5.

3.3 Momentum Factor

Momentum factor simply adds a fraction m of the previous weight update to the current one. The momentum parameter is used to prevent the system from converging to a local minimum or saddle point. A high momentum parameter can also help to increase the speed of convergence of the system. However, setting the

momentum parameter too high can create a risk of overshooting the minimum, which can cause the system to become unstable. A momentum coefficient that is too low cannot reliably avoid local minima, and can also slow down the training of the system. 0.2 and 0.4 are taken as momentum factor in the designed algorithm.

3.4 Training Type

0 = train by epoch

1 = train by minimum error

3.5 Epoch

Determines when training will stop once the number of iterations exceeds epochs. When training by minimum error, this represents the maximum number of iterations. We have taken 100 and 60 epoch for minimum error 0.008.

3.6 Minimum Error

Minimum mean square error of the epoch is called minimum error. It can be defined as the square root of the sum of squared differences between the network targets and actual outputs divided by number of patterns (only for training by minimum error). In this algorithm we have taken minimum squared error (MSE) as 0.008.

4 Methodology for Image Data Compression Using Proposed BPNN

The main idea of neural network applications for the problem of image compression is to construct such a network in which the input and output layer with the same number of neurons will be the connection to the middle (hidden) layer with smaller number of neurons, which is the precondition for compression realization. The relation input/hidden layer size is the compression rate. The goal of such compression is to reconstruct its input hence the output of the network that solves this class of problems is equal to its input. Network training starts with initialization of its weights and requires a set of examples with appropriate input–output pairs. Weights in the network are iteratively modified during the training time to minimize the performance function of the given network, which is taken as MSE (Mean Square Error between the calculated and expected network output). The procedure is repeating till deviation between calculated and expected output is 0.008.

4.1 Steps of BPNN Algorithm

A simulation program to implement back propagation was built which include the following steps:

1. Take 63 input neurons of block size 9×7 to form a character for training vector pair (input and the corresponding output) in training set files. For simulation 05 training vector pairs are taken.
2. Initialization of network weights to small random values, learning rate (alpha) and threshold error. Set iterations to zero.
3. Open training set file.
4. Total_error = zero; iterations \rightarrow iterations + 1.
5. Get one vector and feed it to input layer.
6. Get the target output of that vector.
7. Calculate the outputs of hidden layer units.
8. Calculate the outputs of output layer units.
9. Calculate error (desired output—actual output) and calculate total_error \rightarrow total_error + error.
10. Calculate delta sigma of output layer units and adjust weights between output and hidden layer.
11. Calculate delta sigma of hidden layer units and adjust weights between hidden and input layer.
12. While there are more vectors, go to Step 5.
13. If threshold error \geq total error then stop, otherwise go to Step 4.

For simulation, 5 alphabetical characters (A–E) are chosen. Neural network with 24, 14 and finally with 10 neurons in hidden layers are trained to achieve compression rates 3, 5, and 6 respectively.

5 Simulation Result

For simulation result, 5 alphabetical characters (A–E) are chosen. Block size of one character is 9×7 i.e. in input layer there are 63 neurons. Input patterns are shown in Table 1. Network has been trained for 24, 14, and 10 neurons in hidden layer. Learning rate (0.3, 0.5) and momentum factors (0.2, 0.4) are also adjusted to achieve the desired (0.008) minimum squared error. Most of the characters from “A” to “E” get their shape almost similar to the original shape after decompression.

Tables 2, 3 and 4 shows the minimum squared error (MSE) achieved by characters “A”–“E” in compression for different combinations of learning rates, momentum factor, epoch and compression ratio.

Table 2 Data compression results in terms of minimum squared error (MSE) for characters A–E. Compression ratio = 3

Character	Learning rate = 0.3	Learning rate = 0.5	Learning rate = 0.3	Learning rate = 0.5
	Momentum factor = 0.2 Epoch = 60 MSE	Momentum factor = 0.4 Epoch = 60 MSE	Momentum factor = 0.2 Epoch = 100 MSE	Momentum factor = 0.4 Epoch = 100 MSE
A	0.085339	0.006627	0.046953	0.005225
B	0.015665	0.076973	0.072312	0.007473
C	0.015560	0.081357	0.057334	0.008974
D	0.019492	0.103078	0.058548	0.006590
E	0.014536	0.083399	0.057228	0.006352

Table 3 Data compression result in terms of minimum squared error (MSE) for characters A–E. Compression ratio = 5

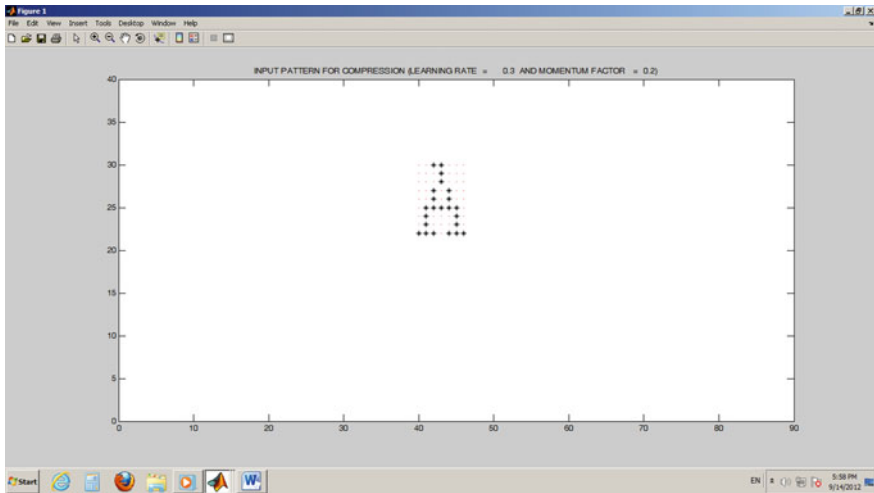
Character	Learning rate = 0.3	Learning rate = 0.5	Learning rate = 0.3	Learning rate = 0.5
	Momentum factor = 0.2 Epoch = 60 MSE	Momentum factor = 0.4 Epoch = 60 MSE	Momentum factor = 0.2 Epoch = 100 MSE	Momentum factor = 0.4 Epoch = 100 MSE
A	0.163688	0.034051	0.079771	0.022048
B	0.140148	0.057788	0.074416	0.020125
C	0.184303	0.029748	0.089696	0.028202
D	0.173084	0.030384	0.109465	0.026873
E	0.121639	0.047957	0.077747	0.015614

Table 4 Data compression result in terms of minimum squared error (MSE) for characters A–E. Compression ratio = 6

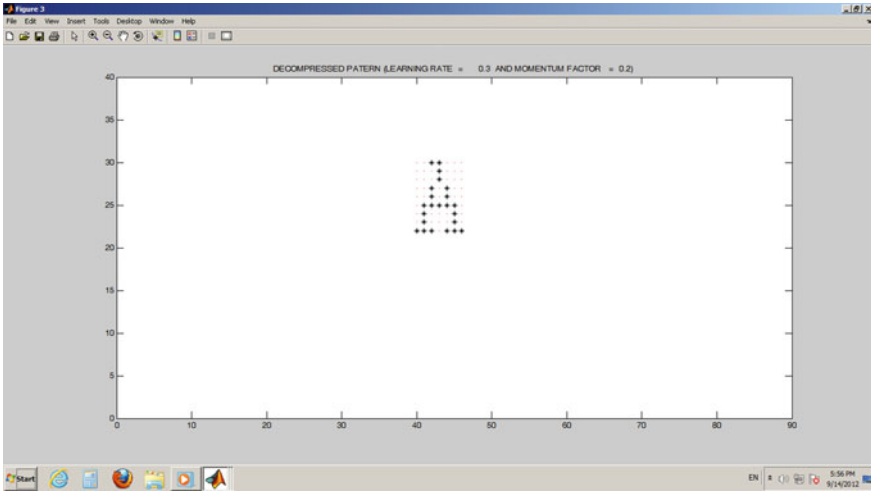
Character	Learning rate = 0.3	Learning rate = 0.5	Learning rate = 0.3	Learning rate = 0.5
	Momentum factor = 0.2 Epoch = 60 MSE	Momentum factor = 0.4 Epoch = 60 MSE	Momentum factor = 0.2 Epoch = 100 MSE	Momentum factor = 0.4 Epoch = 100 MSE
A	0.187018	0.065062	0.100105	0.039388
B	0.275004	0.042760	0.133900	0.031317
C	0.216361	0.077609	0.100406	0.042369
D	0.188603	0.047558	0.123496	0.071309
E	0.189438	0.057925	0.098847	0.031682

Graphical simulation result for compression ratio 3, 5 and 6 have been taken. [Section 5.1](#) shows the graphical simulation result for compression ratio 3 only to avoid repetition of the results. Algorithm is implemented in Matlab.

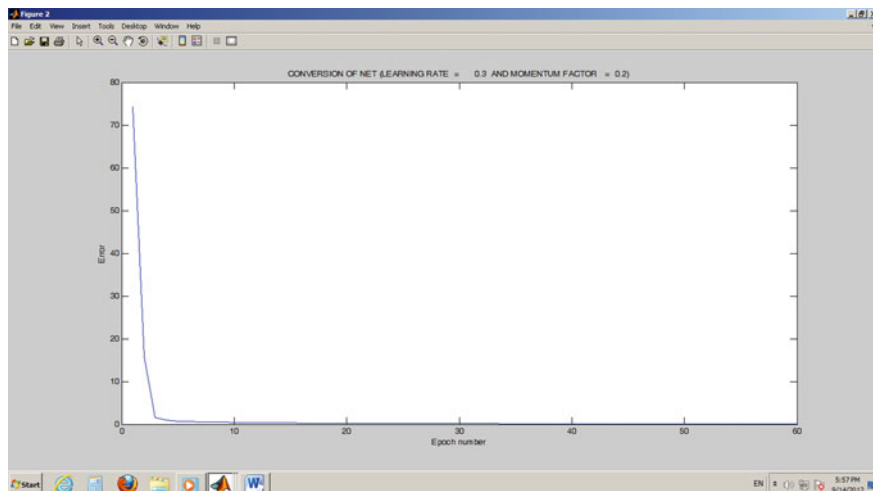
5.1 Graphical Simulation Result for Letter “A”–“E”. Compression Ratio = 3



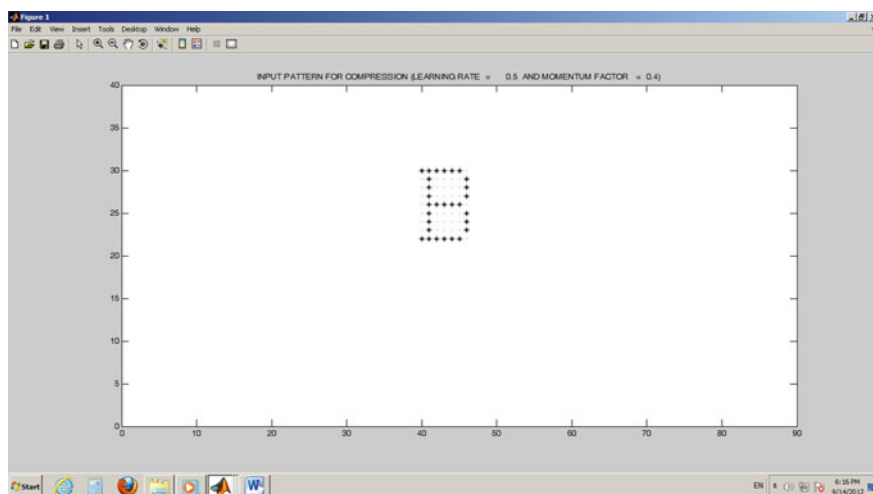
Input Pattern for Character ‘A’



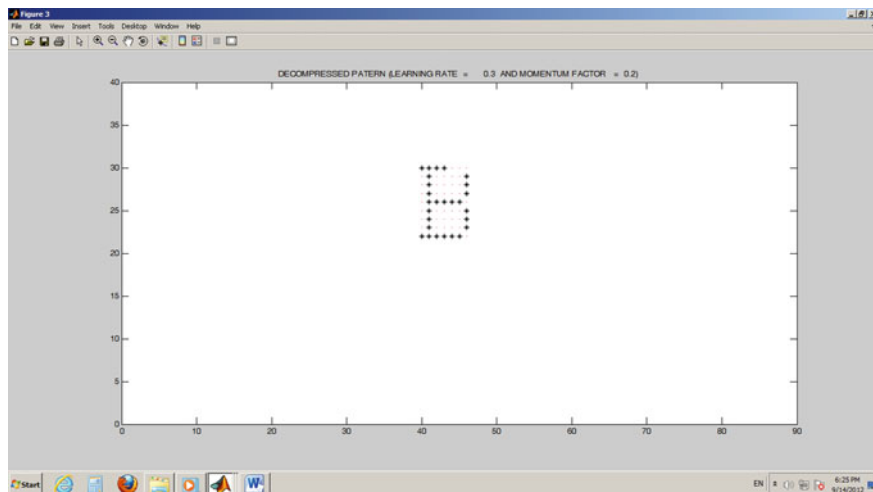
Decompressed Pattern (Learning Rate = 0.3 and Momentum Factor = 0.2)



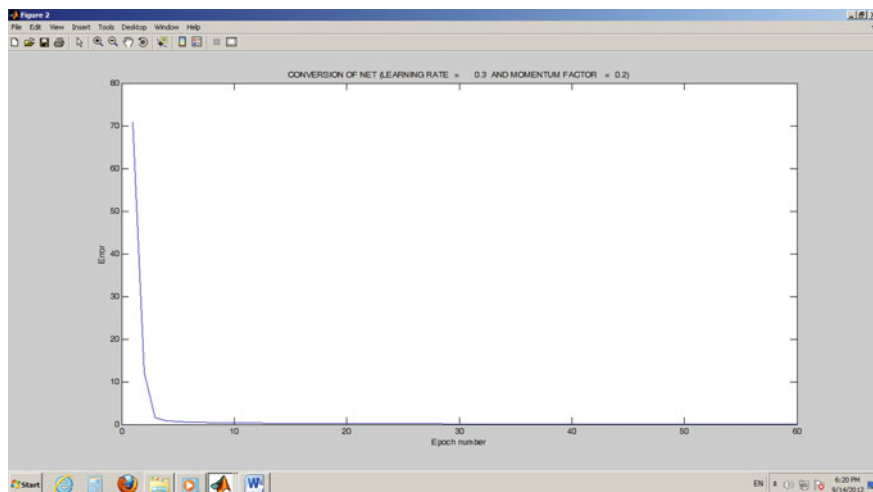
Conversion of NET (Learning Rate = 0.3 and Momentum Factor = 0.2)



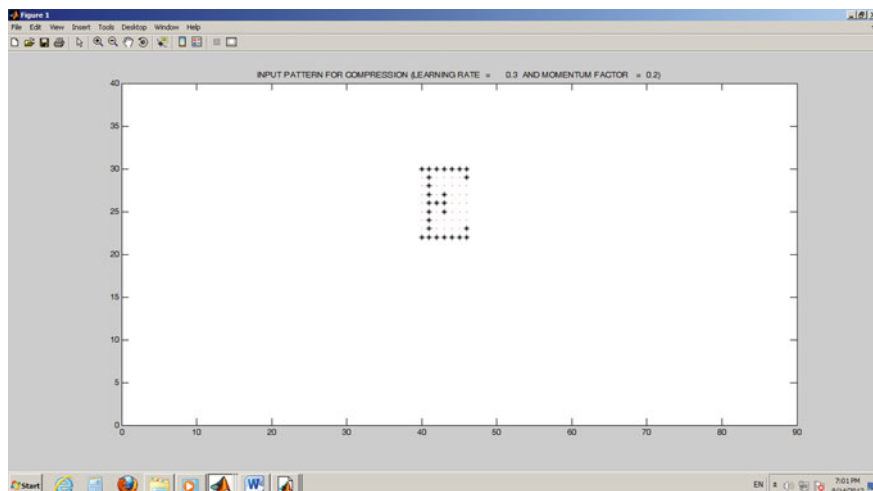
Input Pattern for Character 'B'



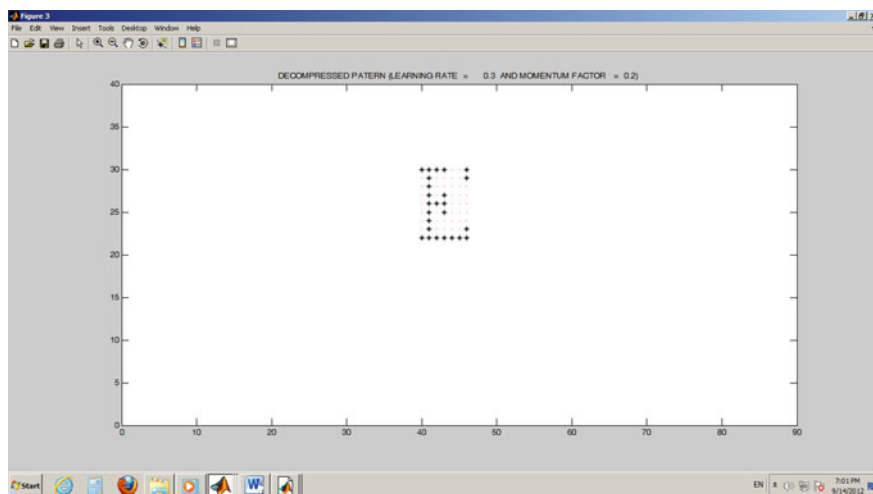
Decompressed Pattern (Learning Rate = 0.3 and Momentum Factor = 0.2)



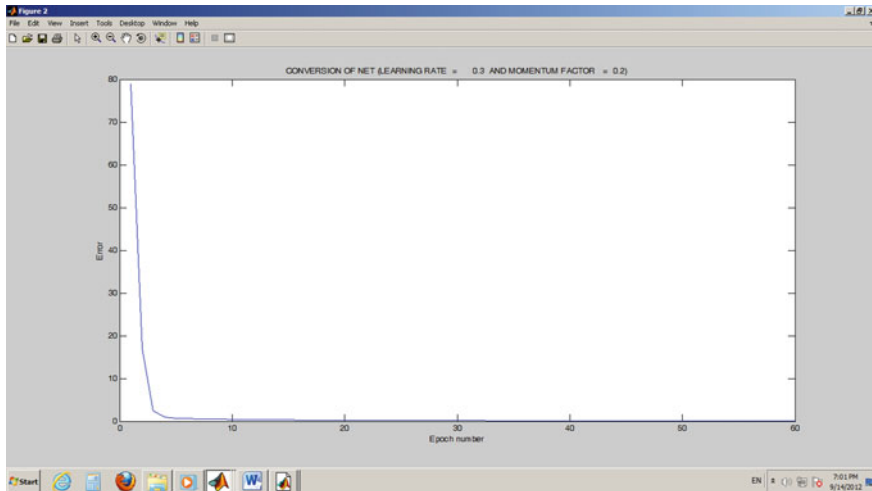
Conversion of NET (Learning Rate = 0.3 and Momentum Factor = 0.2)



Input Pattern for Character 'E'



Decompressed Pattern (Learning Rate = 0.3 and Momentum Factor = 0.2)



Conversion of NET (Learning Rate = 0.3 and Momentum Factor = 0.2)

6 Conclusion

The performance of the designed BPNN image data compression algorithm to the desired minimum squared error is achieved by modifying the network itself by using different values of learning rate parameters, momentum factors, number of neurons in hidden layer and epoch. Combination of Hebbian learning rule and Delta learning rule support the algorithm to converge to the MSE (0.008) with epoch (100).

References

1. Mohammed, A.A., Abdullah, M.T.: Compressed medical image transfer in frequency domain. *Int. J. Image Process. (IJIP)* **5**, 371–379 (2011)
2. Chen, C.H., Yao, Y., Page, D., Abidi, B., Koschan, A., Abid, M.: Comparison of image compression methods using objective measures towards machine recognition. *IJIAP Int. J. Inf. Anal. Process.* **1**, 63–74 (2008)
3. Panda, S.S., Prasad, M.S.R.S., Prasad, M.N.M., Naidu, C.S.K.V.R.: Image compression using back propagation neural network. *Int. J. Eng. Sci. Adv. Technol.* **2**, 74–78 (2012)
4. Patnaik, S., Pal, R.N.: Image compression using Auto associative neural network and embedded zero-tree coding. In: *Third IEEE Processing Workshop Advances in Wireless Communication*, Taiwan (2001)
5. Mittal, M., Lamba, R.: Image compression using vector quantization algorithms. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**(6) (2013)

Books

6. Haykin, S.: Neural Networks—A comprehensive Foundation. Pearson Prentice Hall, New Delhi (2004)
7. Sivanandam, S.N., Sumathi, S., Deepa, S.N.: Introduction to Neural Network using Matlab 6.0. Tata McGraw-Hill Publishing Company Limited, New Delhi (2006)

Conceptual Design of EPICS Based Implementation for ICRH DAC System

Ramesh Joshi, Manoj Singh, S. V. Kulkarni and Kiran Trivedi

Abstract The VME based Ion Cyclotron Resonance Heating (ICRH) Data Acquisition Control system (DAC) is commissioned for remote operation of heating experiment on SST-1 Tokamak. ICRH-DAC is physically distributed into two sections at RF Lab and SST-1 hall. RF Generation section at RF Lab and Transmission line, interface and antenna section at the SST-1 hall having its own independent VME based DAC system. VME system of both section is running on master/slave configuration when synchronization mode operation is needed. This synchronization of both DAC could be possible with EPICS (Experimental Physics and Industrial Control System) process variables, which broadcast and describe itself in Ethernet network. The existing system uses the TCP/IP Ethernet network for the same. The proposed program is used for real-time state parameters transmission and storage, dynamic graphical display, modification of the interactive system and synchronization. This paper will describe the conceptual design of EPICS based ICRH DAC software to achieve desired goal of experiment for generic development platform oriented toward complex data acquisition is proposed.

Keywords VME-versa module eurocard · ICRH-ion cyclotron resonance heating · DAC-data acquisition and control system · EPICS-experimental physics and industrial control system

R. Joshi (✉) · M. Singh · S. V. Kulkarni
Institute for Plasma Research, Bhat Village, Gandhinagar 382428, India
e-mail: arjoshi07@gmail.com

M. Singh
e-mail: parihar@ipr.res.in

S. V. Kulkarni
e-mail: kulkarni@ipr.res.in

R. Joshi · K. Trivedi
Shantilal Shah Engineering College, Bhavnagar 364060, India
e-mail: krtrivedi@gmail.com

1 EPICS Introduction

The EPICS software [1] was originally designed to be tool based approach to process control and this continues to be its primary application. An infrastructure that encourages proper design of distributed software systems is also important. For example, in multi-threaded distributed systems, toolkit needs communication software interfaces designed to avoid application programmer introduced mutual exclusion deadlocks. Interfaces must also be properly structured to encourage robust response to loss of communication or other hardware resources. Portability among workstations and embedded systems is an important requirement for certain advanced applications. These capabilities are required by process control components.

The EPICS based software tools includes application software that offers a satisfying solution for measuring, processing tasks and secondary development software which is a powerful technology that allows software integrators and application users to customize and automate the application software. Several extensions provided with EPICS could also be used for alarm handler, interlocking and alarming mechanism implementation. User interface design with MEDM which provides motif based user interface. Channel Archiver (CA) for the storage of data during shot and analysis with Strip tool or data browser. We could also able to use the CSS Best Opi Yet (BOY) for the user interface development which provides the XML based markup language integration with bundled widget options and integration based on Eclipse platform. Certain aspects of the existing EPICS communication software interfaces appear to be important facilitators for advanced toolkits. Close integration with process control systems requires efficient publish-and-subscribe communication strategies. Message-batching capabilities also improve communication efficiency. Software interfacing with systems capable of independent actions needs interfaces that can generate an asynchronous response synchronized with external events.

2 ICRH System

ICRH is a promising heating method for a fusion device due to its localized power deposition profile, a direct ion heating at high density, and established technology for high power handling at low cost. For the same reason 1.5-MW ICRH system is developed for Steady State Superconducting Tokamak (SST-1). SST-1 is a large aspect ratio Tokamak with wide range of elongation (1.7–1.9) and triangularity (0.4–0.7). The maximum plasma current is expected to be 220 kA. The maximum toroidal magnetic field at the center is 3.0 T. The major and minor radii are 1.1 and 0.2 m. 1.5 MW of RF power is to be delivered to the plasma for pulse lengths of up to 1,000 s [2]. The block diagram of the integrated ICRH system has shown in Fig. 1. The first block named RF Source will generate RF Power as per the

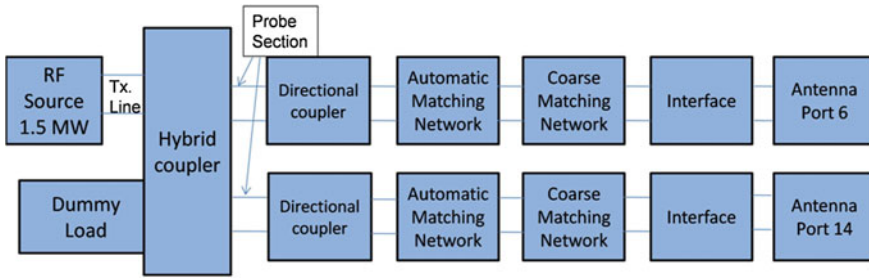


Fig. 1 ICRH block diagram

requirement with different frequencies for the experiment. This operation remotely controlled by Master DAC system and the other part means generated power transmission with matching network and antenna diagnostics would be controlled and monitored by slave DAC system. Both DAC systems have been described in next sub sections.

2.1 ICRH Master DAC System

ICRH has two separate DAC systems, which have been used for two different requirements. The first DAC system been installed at RF Lab to control and monitor the RF generator system, which are provides RF power at 22–25, 45.6 and 91.2 MHz frequencies. Master DAC would remotely operate the different stages of RF Generator like 2, 20, 200 kW and 1.5 MW. 2 and 20 kW stages require two different plate power supplies. 200 kW and 1.5 MW systems would have four different power supplies each which are named as Plate, Control grid, Screen grid and Filament power supply. These all power supplies have different voltage and current setting with raise/lower functionalities. These all functionalities would be remotely controlled and monitored by Master DAC system with required interlocks. The block diagram has been shown in Fig. 2 with different stages. All the signals coming from different stages have been connected through front end electronics and signal conditioning which ends at VME terminal end. The application allows control of a variety of hardware, ranging from straightforward, continuous data streaming of a few channels to multi-rack based data acquisition instruments.

2.2 ICRH Slave DAC System

Slave DAC system is responsible to control and monitor RF transmission and diagnostics using two transmission lines with offline and online matching system

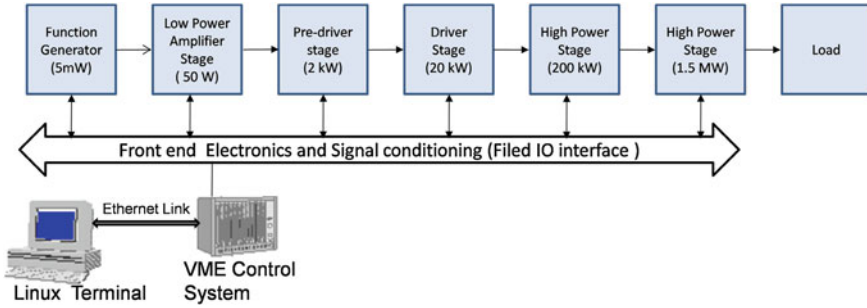


Fig. 2 Block diagram of ICRH generator DAC

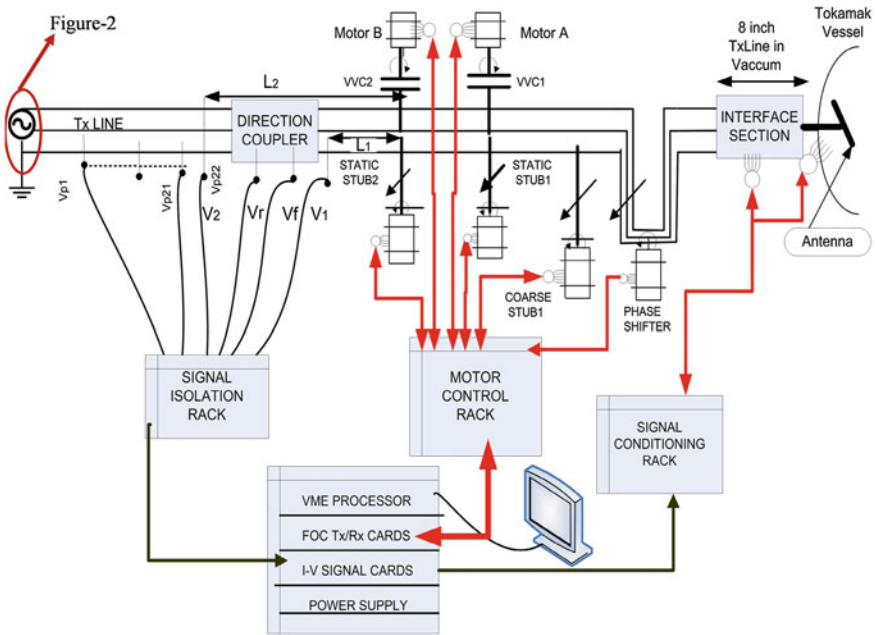


Fig. 3 Block diagram of ICRH transmission line with matching network

followed by Interface sections and antenna diagnostics. The systematic block diagram of the all connected systems has been shown in Fig. 3 for single transmission line. The red oval indicates the master DAC system which have been used as RF generator system as explained via Fig. 2. The DAC server would only get triggered with the trigger pulse provided by the Master DAC either by external hardware or software. The transmission line consists of (a) a pressurized 9 in. 50 Ω coaxial line, (b) matching systems at two different time scales and (c) vacuum transmission line called interface, linking the transmission line to the fast wave

antennae. One single line transmits the power from the RF generator to the two antenna boxes placed at diametrically opposite radial ports. Each transmission line arm has 24 voltage probes, motorized automatic matching network (in ms) and course tuner (in s). Automatic matching network consists of two stubs coupled with stepper motor and two numbers of vacuum variable capacitors coupled with high-speed servomotor. Interface section has vacuum system which has been monitored and acquired with the desired time scale with Linux terminal using serial port. The systems would have connected with some diagnostics which requires faster scale acquisition of the signals like density calculations and some probe signals coming from Antennae interface connected with SST1 machine.

Two another probe signals among the 24 probe signals have been used for local voltage 1 and local voltage 2 for the Automatic matching network error signal calculation along with forward and reflected power. Coarse matching network have been used for the offline matching before the experiment starts and the automatic matching network have been used during shot as it works as fast matching network to match the input impedance with the plasma impedance. This algorithm is based on feedback control loop at every 5 ms which uses the four signals from the transmission lines, calculate the reflection coefficient and that will be used for the error signal calculation. There error signals are being used for the deciding parameter for the cylindrical capacitor direction either clockwise or anticlockwise. The faster 1 kHz sampling rate has been used for all 32×2 signals and around 24 diagnostic signals. There is hardware as well as software interlock have been implemented and tested before experiment and maintained periodically [3, 4].

2.3 Integrated DAC Approach

Integrated DAC system block diagram has been shown in Fig. 4. Both DAC systems have been connected with Ethernet communication and hardware link for fast controller trigger line. The acquisition of the data during experiment would be done at very fast scale like 1 kHz. This process will follow the cycle like the VME processor board will get data from the analog or digital board register memory buffer. At the time of shot the data would be available at the digitizer cards. After shot the data would be demanded by the Linux user interface the socket communication is been established and the packet structure wise data have been available on Ethernet. The processor board will store the data in buffer memory from the digitizer board using messages queue implemented by program. Linux terminal will take the data from that buffer using Ethernet network. The data will be acquired (stored) on user interface computer by the GNU socket library program.

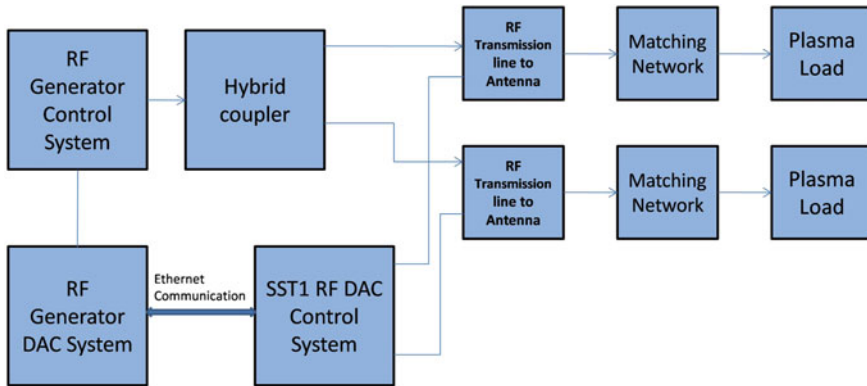


Fig. 4 Block diagram of integrated DAC systems

3 EPICS Implementation

There are about 32 analog input channels, 12 analog output channels and 64 digital I/O channels consist at master DAC same way 96 analog input channels, 12 analog output and 128 digital I/O channels are connected with slave DAC system. The requirement of different channel IOs have been catered with the EPICS base module installation at Linux terminal. To make user interface the state notation language (XML) has been used with CSS IDE and assign required field widgets with respective process variables. The signal naming has been specified at the ICRH: <signal_name>. Using soft IOC module we have broadcasted the process variables (PVs). XY graph has been chosen for monitoring the voltage and current signals. Python script has been used for the periodic assignment of the channels process variables using caput command for apply periodic new value to the respective process variable. Separate python script is running periodically using execute command function provided on action button click event. In this script we have used pyepics [5] package and import epics as python module and will able to process capget and caput command as per requirements.

We have also used cython package for the load shared object module as dynamic library which was made using C program for socket communication with fast controller [6]. Using this module we can run python script on action button event for socket communication and reading and writing data as per need. In separate thread we can run this module which will not affect the main monitoring and control process.

The fast fiber optic trigger network will give trigger pulse to fast controller at Master DAC. This fast controller get triggered that will trigger the fast controller at RF transmission DAC fast controller. As fast controller triggered the digitizer card buffer memory will get filled with the given on-time reference time. This data will be acquired by the Linux terminal user interface program with acquire button by socket command using Ethernet. The integrated architecture of both DAC systems

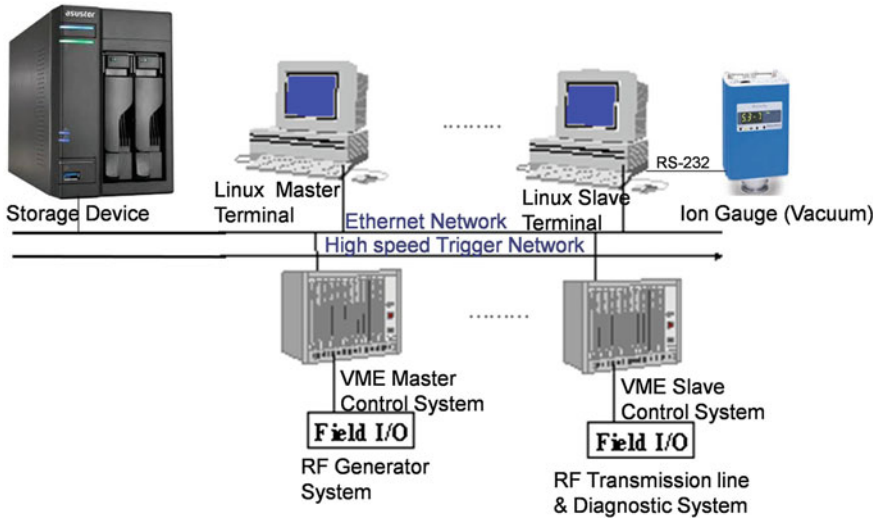


Fig. 5 Integrated architecture of both DAC systems

has been shown in Fig. 5. Master DAC will be communicated by Central Control System (CCS) communication with details of shot number and experimental parameters by Ethernet based communication program and time synchronization by Network Time Protocol (NTP) from Master GPS timer. Master DAC will communicate with slave DAC system with EPICS process variables. Data acquired at master DAC have been sent to the slave DAC and that will acquire data accordingly.

3.1 Advantages Over Existing System

1. Synchronize in required timescale for both DAC in terms of data monitoring, interlock and acquisition.
2. Smooth operation of real time feedback control loop with interlocking using lightweight process variables.
3. Data monitoring and acquire of vacuum data using serial communication on Linux terminal using micro ion gauge with synchronization of acquire trigger pulse.
4. Automatic acquisition of data at slave DAC for transmission lines and diagnostics in synchronous with master digital pulse from master DAC.
5. Interlocking the master DAC system from slave DAC parameters with real time action like reflection coefficient and vacuum degradation which was the limitation of existing system.

6. Process variables available in real time requires by other sub system for synchronous operation which was another limitation of existing system.
7. Experimental Parameters exchange in real time with Central Control System.
8. Accelerate development of next-generation, broad-reach applications.
9. Using web technology revolution one can get the platform independency.
10. Save time with the ultimate developer cockpit.

4 Database

MDSplus (Modal data system) [7] is a data management system used in several Nuclear Fusion experiments to handle experimental and configuration data [8, 9]. Many application programming interface (API) for local and remote data access are available with most languages, namely C, C++, Fortran, Java, Python, MATLAB and IDL, and a set of visualization and analysis tools are available for data browsing and display. In this way, it is possible to take advantage of the availability of the local and remote data access layers of MDSplus, widely used in the fusion community to handle large sets of data. The MDSplus system has been successfully adopted in many fusion experiments for data acquisition, storage and access [9]. MDSplus provide the hints to achieve continuous data acquisition during the experiments required to use acquired data in the active control of the experiment. We could use Channel Archiver tool for binary data storage of process variable to MDSplus database. The Channel Archiver [10] is a generic periodic sampling toolset for EPICS. Using the EPICS Channel Access (CA) network protocol [11], it can collect real-time data from any CA server on the network. The Channel Archiver acts as a Channel Access Client and stores recorded data, acquired via periodic scan or monitored, into indexed binary files [12].

The acquire data at the slave DAC will be done with the shot number at slave DAC embedded master DAC shot number for synchronization. Shot detail form master DAC would be sent in separate child process because it should not affect the main running process of monitoring and control and master DAC. Same way another thread is waiting at slave DAC will watch for the coming information from master DAC and acquires data accordingly.

5 Conclusion

We can handle experimental requirement with broadcasting mechanism using channel access protocol with EPICS. Channels as process variables have been broadcasted every required time period so when the shot is applied to master DAC, the information has been automatically available to slave DAC on same network. There is no need to run separate thread for this purpose and lightweight process of

broadcasting process variable will be accomplished easily. By using this proposed system we could also able to synchronize the ICRH DAC system with Central control system. We haven't considered this section in this paper. The final implementation with experimental results will be published after deployment of conceptual implementation in sub sequent paper.

References

1. EPICS home page. [Online] Available <http://www.aps.anl.gov/epics>
2. Engineering design report (EDR) of ICRH-SST1; SST-2.03-050199-RF Group
3. Durodie, F., Veriver, M.: Eur. Phys. Soc. (EPS), p. 81 (1992)
4. Bora, D., et al.: Nucl. Fusion **46**(3) (2006)
5. PyEpics Python Package, <http://cars.uchicago.edu/software/python/pyepics3/>
6. Ctypes Python Package, <http://python.net/crwe/theller/ctypes/>
7. MDSPlus online reference <http://www.mdsplus.org>
8. Stillerman, J., Fredian, T.: The MDSplus data acquisition system, current status and future directions. Fusion Eng. Des. **43**, 301–308 (1999)
9. Fredian, T., Stillerman, J., Manduchi, G.: MDSplus extensions for long pulse experiments. Fusion Eng. Des. **83**, 317–320 (2008)
10. Dalesio, L.R., et al.: Data Archiving in Experimental Physics, ICALEPCS'97, Beijing (1997)
11. Hill, J.O.: Channel Access: A Software Bus for the LAACS, ICALEPCS'89, Vancouver (1989)
12. Manduchi, G., et al.: New EPICS channel archiver based on MDSplus data system. IEEE Trans. Nucl. Sci. **58**, 3158 (2011)

An Efficient Network Management and Power Saving Wake-On-LAN

Pranjal Daga, D. P. Acharjya, J. Senthil and Pranam Daga

Abstract In distributed systems a computer generally process information of distributed application or provide service in distributed system. Therefore, computers connected in distributed system need to keep on all time. It leads to the concept of Wake-on-LAN (Local Area Network). However, keeping on during the idle period is the wastage of power. Therefore, it is essential to save the power while being efficient in network management. In this paper we propose an improved Wake-on-LAN device that incorporates the usage of basic Wake-on-LAN technology into a network management and power saving product.

Keywords Wake-on-LAN · Distributed system · Magic packet · Traffic routing · Network management · Header checksum

1 Introduction

In the earlier decades, computer systems were highly centralized within a single room. But, the merging of computers and communication technology has brought a profound influence on the way computer systems are organized. As a result single computer serving system is replaced by a large number of separate but interconnected computers. This leads to computer network. Therefore, a collection of autonomous computers are interconnected by a single technology to exchange

P. Daga (✉)

School of Computing Science and Engineering, VIT University, Vellore, India
e-mail: pranjal_daga@gmail.com

D. P. Acharjya · J. Senthil

Faculty, School of Computing Science and Engineering, VIT University, Vellore, India

P. Daga

Purdue University, West Lafayette, IN, USA

their information [6]. This helps common people in different applications such as business expansion, access to remote information, person-to-person communication etc. With these applications people are now interested in computer network and distributed systems. However, there is a considerable overlap between computer network and distributed system. A distributed system consists of multiple computers that communicate through a computer network [4]. The main objective of a distributed system is to solve large computational problems. The basic purpose is to coordinate the use of shared resources or provide communication services to the users. Therefore, a computer needs to keep on while being in the distributed system. But, it is observed that a computer in a distributed system may not be busy always and remain idle sometimes. Therefore, power consumption is to be minimized while it is idle. At the same time, a computer should have efficient network management while being busy. Hence, power consumption and efficient network management are two major issues in distributed systems. The purpose was to resolve a major issue inherent to all computer networks: operating cost. Modern computers require a substantial amount of energy to operate. On a per-system basis, especially in a home environment, computers are economical. However in larger operations such as business networks, in which anywhere from tens to hundreds to even thousands of computers are deployed, operating costs are significant. To combat excessive power consumption, idle computers can be allowed to fall asleep. But under typical network configurations cannot be allowed as it would compromise network performance. Under sleep mode, a computer's network presence can be lost and data connections cannot be made. Furthermore, allowing computers to fall asleep is impractical because a user must manually wake the system if its services are required again. Wake-on-LAN [1] is an attempt to minimize the operating cost. Chandrakasan and Brodersen [3] have discussed minimizing power consumption in digital CMOS circuits. Stemm and Katz [5] have discussed measuring and reducing energy consumption of network interfaces in handheld devices. Ayers et al. [2] developed a lightweight, cost-effective device known as Whack-on-LAN capable of forcing a hardware reset over Ethernet. Emulab [7, 8] is a time- and space-shared network test bed with hundreds of publicly accessible nodes.

In this paper, we propose an improved Wake-on-LAN which minimizes the cost while being efficient in network management. The design is an intermediate network device that is placed between client computers and the network at large. Under normal usage, when its client is actively communicating across the network, the intermediate device allows traffic to flow through as required. However when a client is inactive and falls asleep, the device maintains the client's network presence and allows connections to continue to be made. In this sleep-mode operation, the device simultaneously caches data for its client while waking it using the standard Wake-on-LAN implementation. After its client wakes, the cached data is sent to the client, ensuring network operation.

The rest of the paper is organized as follows. Section 2 discusses the hardware and software specifications required for improved Wake-on-LAN. Detail design of the proposed device is presented in Sect. 3 followed by a conclusion in Sect. 4.

Fig. 1 Abstract view of the proposed device



2 Specifications

In this section, we propose the abstract view of the proposed device as shown in Fig. 1. The proposed device is an intermediate network device placed between client and network. During active schedule, the proposed device allows traffic to flow through as required by the client whereas in sleep mode it caches data for its client while waking it using the standard Wake-on-LAN implementation.

2.1 Software Specifications

The device is an intermediate network device that needs to be able to process network data, receive network packets, and forward network packets. To accomplish this, the software needs to be able to direct traffic to proper addresses. Furthermore, it also needs to be able to detect functional statuses (awake or asleep) of its client as well as to be able to broadcast a magic packet to wake a sleeping computer. Finally, it needs to be able to 2 simultaneously monitor traffic on both client facing and network facing ends. Therefore, operation, the software component of the device needs to handle the following:

- Packet capture, analysis, and injection
- Packet caching
- Traffic routing (via an IP table of addresses)
- Status detection
- Wake-on-LAN via Magic Packet broadcast
- Parallel operation.

2.2 Hardware Specifications

One of the motivations for the design of the device is to save power. Therefore, its power consumption must be considerably lower than that of its clients. To accomplish this task, the main hardware chosen was the Beagleboard-XM, a single-board computer. A brief overview of the relevant specifications of the Beagleboard-XM is listed below:

- Advanced RISC Machine (ARM) Cortex A8 CPU operating at 1 GHz
- 512 MB of DDR Random Access Memory (RAM)
- 4 Universal Serial Bus (USB) ports

- Ability to run Linux
- Programmable General Purpose Input Output (GPIO)
- Peak power consumption of 2 W.

Because the ultimate goal of the design is to be able to operate in a “headless” plug and play manner, the device incorporates a series of 8 light emitting diode (LED) status indicators driven by the Beagleboard-XM s programmable GPIO.

3 Design Details

This section discusses the software design details such as state diagram of software when it receives traffic and state diagram of the software when it sends traffic of the proposed device. The design of the proposed intermediate device meets the basic objective of Wake-on-LAN by maintaining client’s network presence and allowing connections to continue in sleep mode. In this mode, the device simultaneously caches data for its client while waking it using the standard Wake-on-LAN implementation. Clear idea on these design issues is discussed in the following section.

3.1 Software Design

The software implementation needs to first capture incoming traffic. Once it captures a packet it needs to populate an IP table where it will keep the addresses which data can be routed to. Then it determines if the computer that the data is ultimately destined for is available. If the destination is not active, it will cache the traffic, wake the computer, and then forward the data over. If the destination is active, the software will immediately forward the traffic to the destination. This method is valid for both incoming and outgoing traffic. The state diagram of the device when it receives traffic is presented in the following Fig. 2 whereas in Fig. 3 we represent the state diagram of the device when it sends traffic.

- Packet Capture, Analysis, and Injection

Packet capture and packet injection is essential to the device’s traffic forwarding function. When a client sends data to the network, it sends data to the intermediate device’s client facing interface. The intermediate device must capture this packet and then inject it out its network-facing interface. Data sent to the client computer is handled in the same manner: the packets are captured from the network-facing interface and then injected to the client from the client-side interface. This capture and inject process is implemented using the packet capture (PCAP) library. We represent the general structure of a TCP packet in Fig. 4.

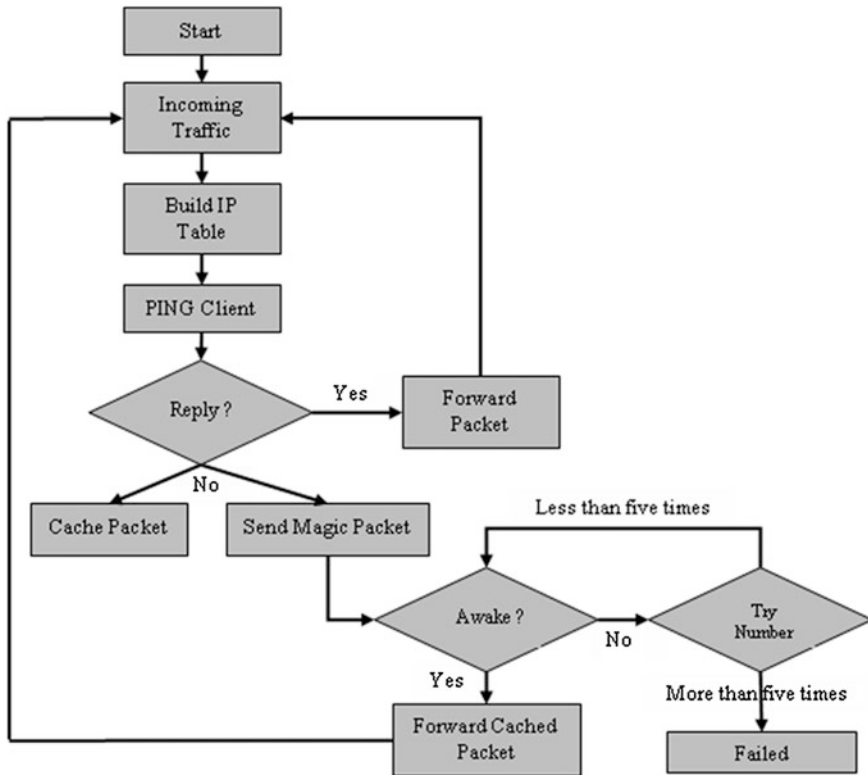


Fig. 2 State diagram of the device when it receives traffic

When a packet is received, it is put into the network stack, in a receiving process. As the packet moves through the receiving process, it progresses up the network stack. At each stage of the process, a header is read and removed before going to the next step. This occurs starting at the Ethernet header, where the source and destination hardware MAC addresses are contained. After the Ethernet header is processed, the IP header, where source and destination IP addresses as well as the IP header checksum, is processed. Then, the transmission control protocol (TCP) header, where the packet’s flags and port information is contained, is processed. More header information is processed if they exist. Eventually, only the payload data remains. This payload data is then forwarded to the application that requested it.

In the case of the intermediate device, none of the payload data is relevant to any of its applications. All network traffic it receives is directed at either the client or the rest of the network. Thus, to the Ubuntu Linux operating system upon which the software implementation is built upon, all of the data moving through the network stack is garbage and will be discarded unless specifically preserved. This is where the PCAP library becomes useful.

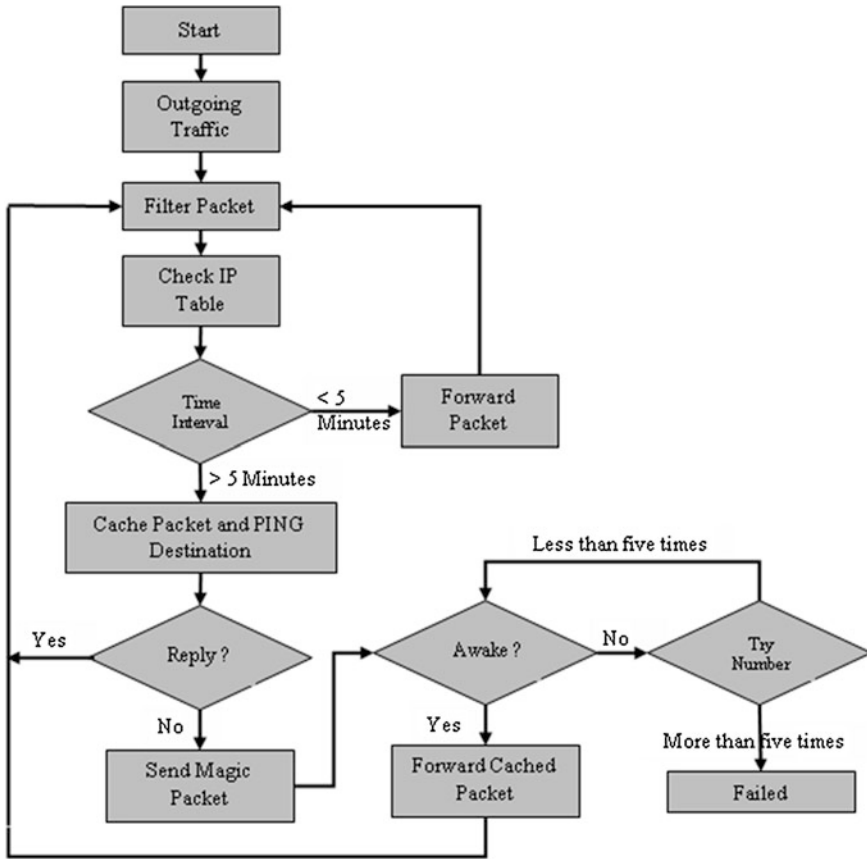


Fig. 3 State diagram of the device when it sends traffic

Fig. 4 The general structure of a TCP packet



The function `pcap_loop()` captures packets continuously from a specified Ethernet interface using a preset filter. In the case of the packet capture, this filter must be set to only accept data with destination to IP address of either one of the device's Ethernet interfaces. Upon capture of each packet, `pcap_loop()` calls a callback function from which analysis and processing can be run on the packet and the other functions, to control the LED status light, to cache packets, to build the IP table, to inject packets, and to broadcast the Magic Packet, are run. After a packet is captured, it is temporarily stored in serial form as a string of characters. Since a character is a single byte, data can be easily parsed through a method called typecasting. Typecasting allows pointers to structures types in C to be used

frames. To locate specific data in the packet, such as the Ethernet header or the IP header, all one needs to do is create a structure and frame it over the desired region. To frame the Ethernet header, one would just need to create a pointer to a structure defining the contents of the Ethernet header and frame it over the first 14 bytes of the packet. To frame the IP header, one would simply do the same except with a pointer to an IP header structure and place it starting at the fifteenth byte of the packet. In pseudo-code it would look like the following:

```
string Packet;  
struct Ethernet_header * eth_hdr;  
eth_hdr = (struct * Ethernet_header *) Packet;  
struct IP_header * ip_hdr;  
ip_hdr = (struct * IP_header *) (Packet + 14);
```

The IP header needs to be modified in two ways, first the destination and source IP addresses need to be properly changed and secondly, the header checksum field must be filled with a recalculated checksum value. In the case of a packet sent from network to client, the destination address field originally contains the IP address of the Beagleboard-X's network-facing interface. This is because computers on the network only see the Beagleboard-XM and not the client computer that is shielded behind it. The Beagleboard-XM effectively acts as a secondary router to its client. Therefore, the destination IP address of the packet needs to be modified to reflect the IP address that the Beagleboard-XM's Linux operating system assigned to the client.

The IP addresses of the packets originating from the client computer do not need to be modified as they proper assignment is taken care of internally by Linux operating system. However, these packets must still be captured in the case that their destination computer is asleep and thus need to be cached by The Engine. Upon modification of any component of the IP header, a new IP header checksum needs to be calculated and appended into the proper field. The IP header checksum is simply the ones' complement of the one's complement sum of the bytes in the IP header. This is an important step in IP header modification because if the checksum is detected to be erroneous by the destination computer, the entire packet is designated as corrupt and promptly discarded.

The checksums of possible headers following the IP header do not need to be modified as they pertain to the payload data as well as their internal values. These contents are never tampered with so no action needs to be done. For example, in the case of a TCP packet, the checksum is computed over the TCP header and its data. All of this higher-level data needs to be preserved; therefore, the checksum does not need to be recomputed.

- Packet Caching

If and when The Engine detects that the destination computer to which a packet is sent is asleep, the traffic is temporarily cached. To cache this data, the strings of characters in which the `pcap_loop()` function stores captured packets is stored into a global array. Packet data is cached into the global array until The Engine

determines the destination computer is awake again. After confirmation of wake, each element of the array is modified and injected to the destination computer.

- Traffic Routing via IP Table

The IP Table is a dynamically allocated global memory structure that essentially routes outgoing traffic from the client. The IP Table is preset with a default connection timeout value. Whenever the client sends a packet to a machine, the IP Table is referenced to determine the time since last connection. If this time interval was greater than the preset timeout value then the program will cache the packet and send out a Magic Packet to the connection destination. However, if the time interval was less than the timeout value then the device assumes the destination is awake, and will forward the packet directly to the destination address. If the IP address had not been previously recorded, a new IP Table entry will be added and a timestamp will be recorded for future reference.

- Status Detection

Status detection of wake or sleep states of computers is implemented using a PING function. The device is set to ping the client every upon program execution but can also be set to ping at specific time interval. The PING function added a select function to include a default timeout to determine a lack of destination response. The default system call `recvfrom()` function does not support a timeout setting.

```
/* PING FUNCTION */
FD_SET(sockfd, &rfds);
tv.tv_sec = 1, tv.tv_usec = 0;
retval = select(sockfd + 1, &rfds, NULL, NULL &tv);
```

With the code above, the PING is able to have a timeout value set by `tv.tv` structure. As an example, it is set as one second. Based on the response of the client from the PING function the device will decide to either forward the packet or to cache and send a magic packet.

- Wake-on-LAN via Magic Packet

The device sends out a magic packet via a user datagram protocol (UDP) broadcast containing the destination machine's MAC address. The magic packet is simply a packet with six `0xFF` values followed by sixteen repetitions of the target's MAC address. The MAC address can be set either via user input from The Engine directly.

- Parallel Operation

Parallel operation is required in The Engine in order to wake a sleeping computer through the broadcast of a magic packet while simultaneously continuously caching incoming packets through a second instance of `pcap_loop()`. To run operations, also known as threads, in parallel, two functions of the PTHREAD library are required.

- `pthread_create()` - used to create a thread.
- `pthread_join()` - used to join threads.

Specifically, in the code, two threads of type `pthread_t` are created. One thread is to run the function to broadcast a magic packet. The second thread is to call a second instance of `pcap_loop()` to store packets into the global array. After the completion of either thread, they are joined together using the `pthread_join()` function and The Engine continues onto its functions. The code implemented in the Engine code is replicated below for clarity.

```
pthread_t thread1;
pthread_t thread2;
/* starts thread 1 to call Magic Packet function */
pthread_create(&thread1, NULL, &magic, NULL);
/* starts thread 2 to call pcap_loop() function to */
pthread_create(&thread2, NULL, &cachePackets, NULL);
/* join the two threads */
pthread_join(thread1, NULL);
pthread_join(thread2, NULL);
```

3.2 *Hardware Specifications*

The hardware design of this project created an LED status indicator for users. It displays the device's operating state in real-time, providing users with visual confirmation of device function or malfunction. Therefore, the user can manage and troubleshoot a network based on the status indicators. This is especially useful in management of large networks, where a status report of the sub-networks under the intermediate device is important. The LEDs also fulfill a debugging role as it indicates all of the statuses: capture, process, sleep status, and Wake-on-LAN. The hardware design is an add-on to the Beagleboard-XM. It consists of software and hardware elements. The software design deals with the C code that implements Beagleboard-XM's general purpose input/output (GPIO) drivers. The hardware design pertains to the physical board design the LED indicators.

- Hardware for GPIO Drivers

The Beagleboard-XM comes with GPIO pins, to which the additional hardware board consisting of 8 LEDs is attached. This made the initial hardware design simple as all of the work is done internally on the Beagleboard-XM. No voltage regulator or jumpers were required as the device is already programmed to maintain a sufficient amount of voltage to drive any LEDs under 5 V. Therefore, the final hardware add-on design for this device consisted of a circuit of 8 LEDs and 8 1 k Ω in series with the GPIO connectors. These resistors ensured that the forward bias current that goes through the LEDs will not burn the LEDs. The most difficult aspect of the design was finding a suitable connector for the Beagleboard-

XM's GPIO pin array. Once all the circuits are completed on the process control block (PCB) board, the software will take care of everything else.

4 Conclusion and Future Extension

The goal of the device is to ensure that packets are forwarded to the correct location on the network and the machines can be waken when required. All of the testing verification passed as stated in the design review. Throughout the research and development process, many publicly available tutorials were consulted to achieve a grasp of network programming. The performance of the device is tested not only in a closed environment, but also under real traffic scenarios and found satisfactory. Despite the overall success, there were several areas that were left unaddressed. Specifically, The Engine and underlying software contains memory leaks because it was a failure to allocate all of the memory that was assigned throughout the runtime. Furthermore, the end product was not a completely automated device. The intention of the project was to allow the device to work without any user configuration. Finally, the product packaging was very rudimentary and could have been created for a more durable product.

In future work, the software memory issues will be addressed. Though they did not affect performance in demonstration, these bugs need to be fixed before this product can be reliably used in real-world applications. Furthermore, it is planned to completely automate the device so it can be used as a plug-and-play product. Finally a case design could be created so the device is protected. In addition, improvements can still be made in code efficiency and product packaging.

The software libraries employed, namely PCAP, PTHREAD, and GPIO as well as their respective documentation are available in the public domain. The reliance on publicly available information creates little ethical concern as credit is given to the source parties.

References

1. AMD Corp. Magic Packet Technology. White Paper 20213, Rev A, AMD, Nov. (1995). <http://www.amd.com/us-en/assets/contenttype/whitepapersandtechdocs/20213.pdf>
2. Ayers, G., Webb, K., Hilber, G.A.K.W.M., Lepreau, J.: Whack-on-LAN: Inexpensive Remote PC Reset. (2005)
3. Chandrakasan, A.P., Brodersen, R.: Minimizing power consumption in digital CMOS circuits. *Proc. IEEE* **83**(4), 498–523 (1995)
4. Sahani, G.J., Suthar, K., Rao, U.P.: A proposed power saving scheme for distributed systems by powering on/off computers remotely. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* **3**(2), 207–210 (2013)
5. Stemm, M., Katz, R.H.: Measuring and reducing energy consumption of network interfaces in handheld devices. In: *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Science* (1997)

6. Tanenbaum, A.S: Computer Networks. Prentice_Hall, USA (1989)
7. White, B., Lepreau, J., Guruprasad, S.: Lowering the barrier to wireless and mobile experimentation. In Proceedings of HotNets-I, Princeton, New Jersey (2002)
8. White, B., Lepreau, J., Stoller, L., Ricci, R., Guruprasad, S., Newbold, M., Hibler, M., Barb, C., Joglekar, A.: An integrated experimental environment for distributed systems and networks. In: Proceedings of the Fifth Symposium on Operating Systems Design and Implementation, pp. 255–270. Boston, MA (2002)

Authentication in Cloud Computing Environment Using Two Factor Authentication

Brijesh Kumar Chaurasia, Awanish Shahi and Shekhar Verma

Abstract The security is one of the indispensable concerns in the cloud computing. Different authentication and access mechanisms are implemented for different services in the cloud. For uniform strong authentication and obviate the need for password registration, two factor authentication (T-FA) has been proposed. The data owner provides one of the credentials in this is a two tier mechanism. T-FA needs two identities based on what the user knows and what he possesses which is implemented through Software as a Service (SaaS) in the cloud computing environment. However, key-compromise impersonation makes the software only mechanism vulnerable if a server is compromised. This study examines this vulnerability and proposes an identity based signature technique to make T-FA resilient to impersonation attack. Analysis shows that the proposed technique is able to make T-FA mechanism robust to key-compromise impersonation attacks.

Keywords Cloud · Impersonation · Security · Two factor authentication (T-FA)

1 Introduction

Cloud computing is a computing for enabling convenient, dynamic scalable, on demand network access to a shared of pool of configurable computing resources over the internet [1]. The cloud computing provides many advantages such as on-demand self service, ubiquitous network access, storage utility, software utilization, rapid elasticity, platform and infrastructure utilization, managed distributed computing power etc. [2]. In cloud computing, cloud basically is a collection of

B. K. Chaurasia (✉)
ITM University, Gwalior, India
e-mail: bkchaurasia.itm@gmail.com

A. Shahi · S. Verma
Indian Institute of Information Technology, Allahabad, India
e-mail: sverma@iiita.ac.in

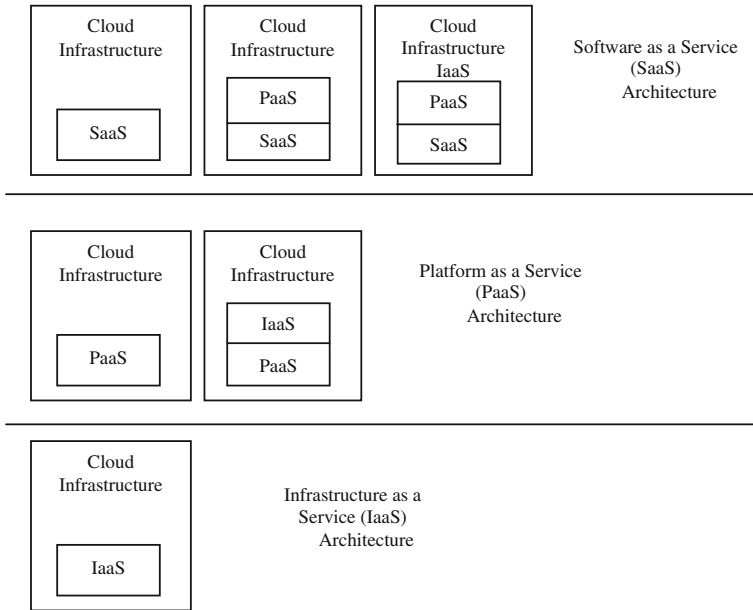


Fig. 1 Cloud standard services

storage systems, servers and devices to amalgamate data, software and computing power distributed across the network at various points. Cloud computing proposes a model the gives an on-demand, economy of scale and pay for use IT services over internet. The infra structure setup by the cloud computing architecture is known as cloud [3]. The cloud computing can be categorized according to services available on the cloud namely IaaS (Infrastructure as a Service), SaaS (Software as a Service) and PaaS (Platform as a Service) [4]. SaaS software as a service model provides applications, software on demand or on subscription basis where user is free from overhead of installation or upgrade. To make the functionality of application or software available over web or internet it is hosted on a centralized servers. SaaS has become most demanding service because of the great services, functionality, flexibility, less maintenance and enhanced scalability. Google docs, CRM applications are the few instances of SaaS. It lowers downs the cost of business [5]. PaaS provides platforms and services to deploy, develop, test, maintain and host application in clouds integrated environment. PaaS acts as a support for application creation and usage. PaaS too is based on subscription or metering model i.e. pay and use. It provides web application platform raw compute and business application platform. IaaS provides hardware services, data storage, bandwidth and networking services. Basically IaaS can be perceived as machine in the sky which provides facility to choose among various operating systems, hardware sizes and storage amount. IaaS provides ability to scale up, upfront cost to minimum by moving your infrastructure to cloud (Fig. 1).

2 Authentication in Cloud Environment

In general, authentication is the mechanism of verifying a user's identity and ensures that user who attempts to access a network is a legitimate user. Another aspect of authentication is to deny, the accessibility of the network to an unauthorized user. The authentication mechanism can be categorized into two types: message authentication and entity authentication. In cloud computing, security services are as follows. SaaS in which a user relies on cloud provider to perform all security, the PaaS model accepts user to assume almost all security functions except availability and IaaS which depends on providers to maintain data integrity as well as availability but loads users with confidentiality and privacy control. As the IT industry and its whole business is moving its sensitive information to cloud, so SaaS providers have to provide strong support for this data in the cloud by taking strong security measures. Authentication in cloud is completely different from that of an organization and the verification of the user is different from that in an enterprise network. Enterprise can rely on multi layer of authentication but in cloud user are not necessarily connected via enterprise LAN. The cloud computing is subjected to various risks as attacks targeting shared-tenancy environment, virtual machine based malware etc. Conventional security methodology do not work well for cloud and hence cloud is subjected to various attacks. So seeing the cloud vulnerability it needs well structured and well defined security procedures. Among various proposed protocols two factor authentication $T-FA$ is popular and is supposed to be used and adopted by many cloud service providers in future. Two factor-or multi factor authentication is two or multi proof based authentication i.e. a proof which user knows (password etc.) and other proof which user has as security code, mobile phone etc. This $T-FA$ can be categorized a hardware based technique or software only technique.

3 Two Factor Authentication $T-FA$

$T-FA$ is a mechanism that provide two way authentication to the user. Authentication depends on at least one of three kinds of information—something that a user knows (password), something that a user possesses (security token or USB or mobile phone) or something that a user is (biometric data like fingerprints).

Two-factor authentication is required if any information that is personally identifiable is handled outside the firewall or in a large communication mechanism. Hardware based token such as USB, smart card etc. deployment is unfavorable in large scale communication scenario and have flaws. $T-FA$ as SaaS allows service level agreements which ensure quality of service especially availability. It also offers administrative and other facilities to the user. But, the system is opened to the Internet making it vulnerable as $T-FA$ has to be deployed for all SaaS applications. A weak authentication mechanism is a problem for the SaaS

providers and which makes the data in the cloud critical in which customers email id and password are under scanner of attacker [6]. *T-FA* in software only technique is also vulnerable as masquerading becomes critical issue in *T-FA*. The notations are used in algorithms is illustrated in Table 1.

4 Vulnerability of Software Only Technique in *T-FA*

The digital signature based solution proposed in [7] requires prime numbers p , q and a base g for creation of a private key (x) and a public key (y) pair. However, when a client needs to authenticate itself by its signature, its public key certificate is sent to the cloud for signature verification. The on-demand transmission and verification of the public key certificate is a communication and computation overhead and causes authentication latency. The client tries to login by sending a request to cloud servers and the servers generates a random challenge m for the user with k as security parameter suppose a number which is generated by key generating function. Under password deriving keys $f(pw)$ user decrypts x where pw is password [8]. Now, user randomly selects c in $[1, q - 1]$ and calculates r and s using α . Thus user sends the result to server where server decrypts w and b and gets (r, s) and (y, g, p, q) [7]. Server verifies the certificate and computes u_1 and u_2 and in turn gets v . Server compares v and r if they match login is granted. Otherwise server suspects the user (Fig. 2).

T-FA minimizes the risk of system compromise as the probability of loss both passwords concurrently is minimal. However, *T-FA* is subject to impersonation attack. Impersonation attack can take place when even without compromising any of the authentication factors. The situations are one, the public key of user was encrypted with symmetric key which is shared by server, then the same domain impersonation can occur [7] or two, the public key of user was encrypted with public keys of the server and private keys remain unchanged. Then, the impersonation will be done in a different domain. So, when a key in long term use is compromised, then the attacker will be able to masquerade and situation grows worse when attacker is able to masquerade as another entity. To prevent this vulnerability, the possibility of an attacker masquerading as another entity must be excluded.

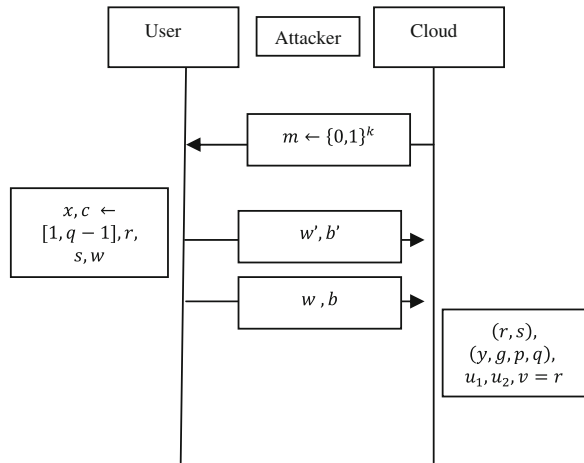
5 Proposed Solution

The same domain and different domain impersonation attacks are possible in software only *T-FA* mechanism. However, this possibility can be excluded by using the following two approaches. One, if the public–private key pair of a client is tied to its identity and this is generated by a trusted third party who is specific to a particular domain. Second, if the client generates key pairs in large number targeting on many servers. Under both the conditions, the impersonation will not

Table 1 Notation used for algorithm

Notation	Description
A	User
$(serv, serv_1)$	Cloud server
p, q	Large prime numbers
$m \leftarrow \{0, 1\}^k$	Random number
g	Base
(x)	Private exponent
$f()$	Key driving function
$f(pw)$	Password driving key
pw	Password
$a = E(f(pw), x)$	Variable
$D()$	Inverse of a
$b = E(\alpha(y, g, p, q))$	Variable
α, μ	Server key pair

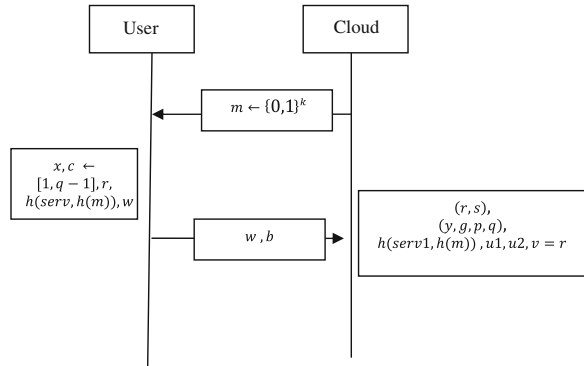
Fig. 2 The impersonation attack in software only based T-FA



happen as a signature is accepted only when the client is directing a particular server or when a key pair is specific to a domain. When the targeted server is unable to identify the actual user receiving the signature and impersonation attack takes place and two, when user loosely signs the digital signature as plain text, impersonation attack takes place. Tying the identifier of the client to the key pair in signature for recipient can be used in successful acceptance of signature in the aforesaid conditions.

However, communication and computation efficiency is imperative when large number of key pairs is generated. Intensive computation will decrease the client’s interest in the cloud service, while communication load increase the cost and add to the service’s risk of failure especially in a wireless mobile environment. This makes digital signature scheme based on public key cryptography less attractive in

Fig. 3 Proposed methodology with one way hash function



this scenario. In the proposed mechanism, the second approach of multiple key pair generation is combined with identity based signature (IBS) mechanism. In the proposed Identity based cryptography [9], a client’s identifier is its public key, and the corresponding private key is generated by a trusted private key generator (PKG). IBS achieves less overhead by reducing the size of the keys and the need for transmission and verification of public key certificates. The client generates an authenticator using its private key which is verified by the server.

A. Generation of Domain Specific Key Pair

There are three participants in the signature generation process: client, the server of a domain and an offline PKG specific to the domain. The PKG is a trusted third party that issues identity based private keys to legitimate clients. Such PKGs are quite common in the real world. The system setup is implemented by the PKG for enrolling the client. In this step, given the security parameter, the PKG determines its public/private key pair and makes the system parameters public. The client enrolls itself using its identity. The PKG generates the corresponding public key and passes the pair to the client over a secret channel. The PKG may also generate a time dependent key pair using the time stamp. The key pair becomes invalid after a time period and requires reissue of the keys. This prevents long term usage of the same signature which is a source of key compromise.

B. Resilience to Authentication Attacks

If an attacker poses as a server say $serv_1$ computes v on the other hand the user calculates s . So v and r are different for attacker as it he uses $serv_1$ to compute all parameters (Fig. 3).

This can be achieved through one way has function $h()$. As $serv_1$ calculates $h(serv_1, h(n))$ for getting v while user computes $h(serv_1, h(n))$ for s as result, $v \neq r$ for the attacker. So attacker cannot malign $serv_1$ as intended server identity is sent with the signature. Moreover in IBS, the identity of the client is tied to the signature and the signature is domain specific, impersonation attacks are clearly excluded.

6 Conclusion

Two factor authentication mechanism when implemented as software as a service is vulnerable to same and different domain impersonation attacks. The elimination of the security weakness requires usage of digital signature and multiple key pairs which has a large overhead. To overcome the vulnerability to impersonation, identity based signature with domain specific key generation for multiple key pairs has been proposed. The mechanism is not only able to prevent impersonation attacks but is also efficient. However, the key pair lifetime must be limited to control spread of the keys from compromised servers.

References

1. Takabi, H., Joshi James, B.D., Ahn, G.J.: Security and privacy challenges in cloud computing environments. *IEEE Secur. Soc.* pp. 24–31 (2010)
2. Choudhury, A.J., Kumar, P., Sain, M., Lim, H., Hoon, J.L.: A strong user authentication framework for cloud computing. *IEEE Asia-Pacific Services Computing Conference*, pp. 110–115 (2011)
3. Robert, L.G.: The case of cloud computing. *IT Prof.* **11**(2), 23–27 (2009)
4. Ramgovind, S., Eloff, M.M., Smith, E.: The management of security in cloud computing. *IEEE Info. Secur. South Africa*, pp. 1–7 (2010)
5. Kim, W.: Cloud computing today and tomorrow. *J. Object Technol* **8**(1), 65–72 (2009)
6. Rash, W.: Is cloud computing secure? Prove it, In tech in depth, eWeek, pp. 8–10 (2009)
7. FIPS.: Digital Signature Standard. Federal Information Processing Standards Publication **186**, 186–2 (1994)
8. Taekyoung, K.: Impersonation attacks on software-only two-factor authentication schemes. *IEEE Commun. Lett.* **6**(8), 358–360 (2002)
9. Cao, Xuefei, Zeng, Xingwen, Kou, Weidong, Hu, Liangbing: Identity-based anonymous remote authentication for value-added services in mobile networks. *IEEE Trans. Veh. Technol.* **58**(7), 3508–3517 (2009)

Expedited Artificial Bee Colony Algorithm

Shimpi Singh Jadon, Jagdish Chand Bansal, Ritu Tiwari and Harish Sharma

Abstract Artificial Bee Colony (ABC) is one of the latest and emerging swarm intelligence algorithms. Though, there are some areas where ABC works better than other optimization techniques but, the drawbacks like sticking at local optima and preferring exploration at the cost of exploitation, are also associated with it. This paper uses position update equation in ABC as in Gbest-guided ABC (GABC) and attempts to improve ABC algorithm by balancing its exploration and exploitation capabilities. The proposed algorithm is named as Expedited Artificial Bee Colony (EABC). We altered the onlooker bee phase of ABC by forcing the individual bee to take positive direction towards the random bee if this selected random bee has better fitness than the current bee and if it is not the case then the current bee will move in reverse direction. In this way, ABC colony members will not follow only global best bee but also a random bee which has better fitness than the current bee which is going to be modified. So the mentioned drawbacks of the ABC may be resolved. To analyze the performance of the proposed modification, 14 unbiased benchmark optimization functions have been considered and experimental results reflect its superiority over the Basic ABC and GABC.

Keywords Artificial bee colony · Swarm intelligence · Optimization · Gbest artificial bee colony

S. S. Jadon (✉) · R. Tiwari

ABV-Indian Institute of Information Technology and Management, Gwalior, India
e-mail: shimpisingh2k6@gmail.com

R. Tiwari

e-mail: tiwariritu2@gmail.com

J. C. Bansal

South Asian University, New Delhi, India
e-mail: jcbansal@gmail.com

H. Sharma

Government Engineering College, Jhalawar, India
e-mail: harish.sharma0107@gmail.com

1 Introduction

Swarm Intelligence is an emerging category in nature inspired algorithms group in current decade. Collaborative trail and error method is the main engine behind the Swarm Intelligence which enables the algorithmic procedure to find the solution. Researchers analyzed such collaboration between the social insects while searching food for them and created the structures known as Swarm Intelligence based algorithms like ant colony optimization (ACO) [6], particle swarm optimization (PSO) [14], bacterial foraging optimization (BFO) [15]. The work presented in the articles [6, 14, 16, 17] under the mentioned category proved their efficiency and potential to deal with non linear, non convex and discrete optimization problems etc. Karaboga [10] contributed the recent addition to this category known as Artificial bee colony (ABC) optimization algorithm. The ABC algorithm mimics the foraging behavior of honey bees while searching food for them. ABC is a simple and population based optimization algorithm. Here the population consists of possible solutions in terms of food sources for honey bees whose fitness is regulated in terms of nectar amount which the food source contains. The swarm updating in ABC is due to two processes namely the variation process and selection process which are responsible for exploration and exploitation respectively. However the ABC achieves a good solution at a significantly faster rate but, like the other optimization algorithms, it is also weak in refining the already explored search space, mainly due to less diversity in later search. On the other part, it is also required to tune the ABC control parameters based on problem. Also literature says that basic ABC itself has some drawbacks like stop proceeding toward the global optimum even though the population has not converged to a local optimum [11] and lacking to balance between exploration and exploitation [19]. Therefore these drawbacks require an enhanced and more efficient ABC to deal with. Here we propose a modification in ABC called as Expedited ABC (EABC) to overcome the mentioned drawbacks. Researchers are also working to enhance the capabilities of ABC. To enhance the exploitation, Gao et al. [8] improved position update equation of ABC such that the bee searches only in neighborhood of the previous iteration's best solution. Banharnsakun et al. [2] proposed the best-so-far selection in ABC algorithm and incorporated three major changes: The best-so-far method, an adjustable search radius, and an objective-value-based comparison in ABC. To solve constrained optimization problems, Karaboga and Akay [12] used Deb's rules consisting of three simple heuristic rules and a probabilistic selection scheme in ABC algorithm. Karaboga [10] examined and suggested that the *limit* should be taken as $SN \times D$, where, SN is the population size and D is the dimension of the problem and coefficient ϕ_{ij} in position update equation should be adopted in the range of $[-1, 1]$. Further, Kang et al. [9] introduced exploitation phase in ABC using Rosenbrock's rotational direction method and named modified ABC as Rosenbrock ABC (RABC).

Rest of the paper is organized as follows: ABC is explained in Sect. 2. In Sect. 3, proposed modified ABC (EABC) is explained. In Sect. 4, performance of the proposed strategy is analyzed. Finally, in Sect. 5, paper is concluded.

2 Artificial Bee Colony (ABC) Algorithm

Artificial Bee Colony is made of three groups of bees: employed bees, onlooker bees and scout bees. The number of employed and onlooker bees is equal. The employed bees searches the food source in the environment and store the information like the quality and the distance of the food source from the hive. Onlooker bees wait in the hive for employed bees and after collecting information from them, they start searching in neighborhood of food sources with better nectar. If any food source is abandoned then scout bee finds new food source randomly in search space. ABC is a population-based iterative search procedure. In ABC, first initialization of the solutions is done as:

2.1 Initialization of the Swarm

If D is the number of variables in the optimization problem then each food source $x_i(i = 1, 2, \dots, SN)$ is a D -dimensional vector among the SN food sources and is generated using a uniform distribution as:

$$x_{ij} = x_{minj} + rand[0, 1](x_{maxj} - x_{minj}) \tag{1}$$

here x_i represents the i th food source in the swarm, x_{minj} and x_{maxj} are bounds of x_i in j th direction and $rand[0, 1]$ is a uniformly distributed random number in the range $[0, 1]$. After initialization phase ABC requires the cycle of the three phases namely employed bee phase, onlooker bee phase and scout bee phase to be executed.

2.2 Employed Bee Phase

In this phase, i th candidate's position is updated using following equation:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) \tag{2}$$

here $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, D\}$ are randomly chosen indices and $k \neq i$. ϕ_{ij} is a random number between $[-1, 1]$. After generating new position, the position with better fitness between the newly generated and old one is selected.

2.3 Onlooker Bees Phase

In this phase, employed bees share the information associated with its food source like quality (nectar) and position of the food source with the onlooker bees in the hive. Onlooker bees evaluate the available information about the food source and based on its fitness it selects a solution with a probability $prob_i$. Here $prob_i$ can be calculated as function of fitness (there may be some other):

$$prob_i(G) = \frac{0.9 \times fitness_i}{maxfit} + 0.1, \quad (3)$$

here $fitness_i$ is the fitness value of the i th solution and $maxfit$ is the maximum fitness amongst all the solutions. Again by applying greedy selection, if the fitness is higher than the previous one, the onlooker bee stores the new position in its memory and forgets the old one.

2.4 Scout Bees Phase

If for a predetermined number of cycles, any bee's position is not getting updated then that food source is taken to be abandoned and this bee becomes scout bee. In this phase, the abandoned food source is replaced by a randomly chosen food source within the search space. In ABC, the number of cycles after which a particular food source becomes abandoned is known as *limit* and is a crucial control parameter. In this phase the abandoned food source x_i is replaced by a randomly chosen food source within the search space same as in initialization phase 1.1.

2.5 Main Steps of the ABC Algorithm

The pseudo-code of the ABC is shown in Algorithm 1 [11].

Algorithm 1 Artificial Bee Colony Algorithm:?

Initialize the parameters;

while Termination criteria is not satisfied **do**

 Step 1: Employed bee phase for generating new food sources;

 Step 2: Onlooker bees phase for updating the food sources depending on their nectar amounts;

 Step 3: Scout bee phase for discovering the new food sources in place of abandoned food sources;

 Step 4: Memorize the best food source found so far;

end while

Output the best solution found so far.

In 2010, Zhu and Kwong [19] proposed Gbest-guided ABC (GABC) algorithm, an improvement in ABC algorithm by introducing global best (gbest) solution of the current population into the position update equation to improve the exploitation. This proposed version GABC is inspired by PSO [14], which uses the global best (gbest) solution to guide the search. The modified the position update equation of ABC as follows:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}) + \psi_{ij}(y_j - x_{ij}) \quad (4)$$

The only difference in the modified Eq. (4) is the third term on the right-hand side, which is called gbest term. Here y_j is the j th dimension of the global best solution, ψ_{ij} is a uniform random number in $[0, C]$, where C is a non negative constant. According to Eq. (4), the gbest term tries to drive the search towards the global best solution, therefore, responsible for the exploitation. Note that setting the parameter C in (4) is a crucial task as to balance the exploration and exploitation of the solution search. In this paper ABC uses Eq. (4) of GABC as its position update equation and which is modified further to enhance its search capability.

3 Expedited Artificial Bee Colony

Karaboga and Akay [11] shows inefficient balance between exploration and exploitation of the search space in ABC while experimenting its different variants. Recently Zhu and kwong [19] proposed a modified version of ABC namely Gbest-guided Artificial Bee Colony algorithm (GABC) which incorporated information of the bee having highest fitness in the swarm within the position update equation. As a result of which, ABC may suffer from premature convergence as in early iterations the whole population gets influence to best solution (which is not necessarily a optimum) in the current swarm. At the same time, we can not lose this best solution information also, which tries to guide the population towards the better one. So in order to make balance between losing and adopting the best solution information, we modify the difference term in position update equation of ABC, which adopts the directional difference between the current bee and a random bee in the swarm, through taking positive scaled value of this term in that bee's direction which is better between current and random bee. In this way, we adopt information of both the swarm best and best between current and a random bee and modified the position update equation inspired from PSO [14] which does not give more weightage to global best, but a equal weightage to personal best also. After analyzing these concerns, we propose following modification in **onlooker bee phase** of ABC. The proposed modification can be mathematically seen as follows:

if ($prob_{current(i)} > prob_{random(k)}$)

$$v_{ij} = x_{ij} + rand[0, 0.5](x_{ij} - x_{kj}) + \psi_{ij}(y_j - x_{ij}) \quad (5)$$

else

$$v_{ij} = x_{ij} + rand[0, 0.5](x_{kj} - x_{ij}) + \psi_{ij}(y_j - x_{ij}) \quad (6)$$

here $prob_{current(i)}$ and $prob_{random(k)}$ are the probabilities calculated in Eq. (3) of the current bee i th and random bee k th respectively. $rand[0, 0.5]$ is a random number between 0 and 0.5. Basically in ABC, we follow positive direction towards best bee in current swarm and random direction towards randomly selected k th bee, but in proposed EABC we follow positive direction towards gbest and also towards random bee if it has better fitness than the current bee otherwise we follow negative direction towards random bee. Here, in both Eqs. (5) and (6) we take random number between $[0, 0.5]$ instead of $[0, 1]$ so that this term may not takeover the gbest influence over the search. Therefore instead of completely following gbest bee in colony, we add some better random direction to current bee also to ensure convergence not prematurely.

4 Experiments and Results

4.1 Test Problems Under Consideration

To validate the performance, we compared the proposed strategy with basic ABC and GABC over 14 unbiased benchmark optimization functions which are given in Table 1. For a successful run, the minimum error criteria is fixed as in last colon of Table 1 i.e. an algorithm is considered successful if it finds the error less than acceptable error in a specified maximum function evaluations.

4.2 Experimental Setting

To prove the efficiency of proposed EABC, it is compared with ABC and GABC. To test EABC, ABC and GABC over considered problems, following experimental setting is adopted:

Parameter setting for ABC:

- Colony size $NP = 50$ [5, 7],
- $\phi_{ij} = rand[-1, 1]$ and $\psi_{ij} = rand[0, 1.5]$ [19],
- Number of food sources $SN = NP/2$,
- $limit = D \times SN$ [1, 13],
- The number of simulations/run = 100.

Table 1 Test problems

Test problem	Objective function	SS	OV	D	AE
Beale function	$f_1(x) = [1.5 - x_1(1 - x_2)]^2 + [2.25 - x_1(1 - x_2^2)]^2 + [2.625 - x_1(1 - x_3^2)]^2$	$[-4.5, 4.5]$	$f(3, 0.5) = 0$	2	$1.0E-05$
Colville function	$f_2(x) = 100[x_2 - x_1^2]^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$	$[-10, 10]$	$f(\mathbf{1}) = 0$	4	$1.0E-05$
Brannin's function	$f_3(x) = a(x_2 - bx_1^2 + cx_1 - d)^2 + e(1 - f) \cos x_1 + e$	$-5 \leq x_1 \leq 10, 0 \leq x_2 \leq 15$	$f(-\pi, 12.275) = 0.3979$	2	$1.0E-05$
Kowalik function	$f_4(x) = \sum_{i=1}^{11} \left[a_i - \frac{x_i(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right]^2$	$[-5, 5]$	$f(0.1928, 0.1908, 0.1231, 0.1357) = 3.07E-04$	4	$1.0E-05$
Shifted Rosenbrock	$f_5(x) = \sum_{i=1}^{D-1} (100(z_i^2 - z_{i+1})^2 + (z_i - 1)^2) + f_{bias}, z = x - o + 1, x = [x_1, x_2, \dots, x_D], o = [o_1, o_2, \dots, o_D]$	$[-100, 100]$	$f(o) = f_{bias} = 390$	10	$1.0E-01$
Shifted sphere	$f_6(x) = \sum_{i=1}^D z_i^2 + f_{bias}, z = x - o, x = [x_1, x_2, \dots, x_D], o = [o_1, o_2, \dots, o_D]$	$[-100, 100]$	$f(o) = f_{bias} = -450$	10	$1.0E-05$
Shifted rastrigin	$f_7(x) = \sum_{i=1}^D (z_i^2 - 10 \cos(2\pi z_i)) + 10 + f_{bias}, z = (x_1, x_2, \dots, x_D), o = (o_1, o_2, \dots, o_D)$	$[-5, 5]$	$f(o) = f_{bias} = -330$	10	$1.0E-02$
Shifted schwefel	$f_8(x) = \sum_{i=1}^D \left(\sum_{j=1}^i z_j \right)^2 + f_{bias}, z = x - o, x = [x_1, x_2, \dots, x_D], o = [o_1, o_2, \dots, o_D]$	$[-100, 100]$	$f(o) = f_{bias} = -450$	10	$1.0E-05$
Shifted griewank	$f_9(x) = \sum_{i=1}^D \frac{z_i^2}{4000} - \prod_{i=1}^D \cos\left(\frac{z_i}{\sqrt{i}}\right) + 1 + f_{bias}, z = (x - o), x = [x_1, x_2, \dots, x_D], o = [o_1, o_2, \dots, o_D]$	$[-600, 600]$	$f(o) = f_{bias} = -180$	10	$1.0E-05$

(continued)

Table 1 (continued)

Test problem	Objective function	SS	OV	D	AE
Ackley	$f_0(x) = -20 + e + \exp\left(-\frac{0.2}{D} \sqrt{\sum_{i=1}^D x_i^2}\right)$	$[-1, 1]$	$f(0) = 0$	30	$1.0E-05$
Goldstein-price	$f_{11}(x) = (1 + (x_1 + x_2 + 1)^2 \cdot (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)) \cdot (30 + (2x_1 - 3x_2)^2)$	$[-2, 2]$	$f(0, -1) = 3$	2	$1.0E-14$
Easom's function	$f_{12}(x) = -\cos x_1 \cos x_2 e^{(-(x_1 - \pi)^2 - (x_2 - \pi)^2)}$	$[-10, 10]$	$f(\pi, \pi) = -1$	2	$1.0E-13$
Meyer and Roth	$f_{13}(x) = \sum_{i=1}^5 \left(\frac{x_i x_i}{1 + x_i + x_i^2} - y_i \right)^2$	$[-10, 10]$	$f(3.13, 15.16, 0.78) = 0.4 E-04$	3	$1.0E-03$
Shubert	$f_{14}(x) = -\sum_{i=1}^5 i \cos((i+1)x_1 + 1) \sum_{j=1}^5 i \cos((i+1)x_2 + 1)$	$[-10, 10]$	$f(7.0835, 4.8580) = -186.7309$	2	$1.0E-05$

SS Search space; OV Optimum value; D Dimension; AE Acceptable error

4.3 Results Comparison

With experimental setting of Sect. 4.2, Table 2 reports the numerical results in terms of standard deviation (*SD*), mean error (*ME*), average function evaluations (*AFE*), and success rate (*SR*). To minimize the effect of the algorithmic stochastic nature, the reported function evaluations is averaged over 100 runs. Results in Table 2 reflects that most of the time *EABC* outperforms in terms of reliability, efficiency and accuracy as compare to the *ABC* and *GABC*. For more intensive analyses of the results, the performance indices and boxplots have been carried out.

EABC, *ABC* and *GABC* are compared through *SR*, *AFE* and *ME* in Table 2. Here signs +/– indicate that for that particular function, *EABC* is better/worst than the considered algorithms, respectively. If *EABC* has more success rate *SR* than the other algorithm then it is said to be better and if *EABC* has equal success rate to other then comparison is done in order of preference of average function evaluation (*AFE*), mean error (*ME*). Total Outcome of comparison is mentioned in Table 3. The bottom line of Table 3 represents that of *EABC* is better than *ABC* in all 14 functions and better than *GABC* in 11 functions out of 14 functions and hence shows the superiority of *EABC*.

We also did boxplot analyses [18] to compare all the considered algorithms in terms of consolidated performance as it can efficiently represent the empirical distribution of data graphically. The boxplots for *EABC*, *ABC* and *GABC* are shown in Fig. 1. It is clear from this figure that *EABC* is better than the considered algorithms as interquartile range and median are comparatively low.

Table 3 shows the performance of *EABC* in the order of preference of *SR*, *AFE* and *ME*. Now giving weighted importance to these parameters, performance indices (*PI*) are calculated [4]. The values of *PI* for the *EABC*, *ABC* and *GABC* are calculated by using following equations:

$$PI = \frac{1}{N_p} \sum_{i=1}^{N_p} (k_1 \alpha_1^i + k_2 \alpha_2^i + k_3 \alpha_3^i)$$

where

$$\alpha_1^i = \frac{Sr^i Tr^i}{;} \alpha_2^i = \begin{cases} \frac{Mf^i}{Afi^i}, & \text{if } Sr^i > 0. \\ 0, & \text{if } Sr^i = 0. \end{cases} ; \text{ and } \alpha_3^i = \frac{Mo^i}{Ao^i}$$

$$i = 1, 2, \dots, N_p$$

- Sr^i = Successful simulations/runs of *i*th problem.
- Tr^i = Total simulations of *i*th problem.
- Mf^i = Minimum of average number of function evaluations used to obtain the required solution of *i*th problem.

Table 2 Comparison of the results of test problems

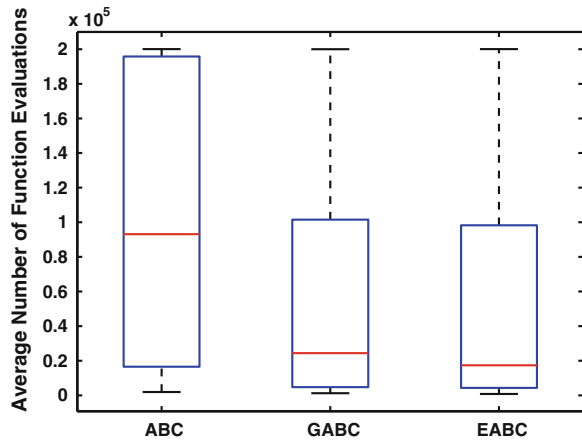
Test Function	Algorithm	SD	ME	AFEs	SR
f_1	ABC	1.66E-06	8.64E-06	16520.09	100
	GABC	3.05E-06	5.03E-06	9314.71	100
	EABC	2.77E-06	5.86E-06	5623.48	100
f_2	ABC	1.03E-01	1.67E-01	199254.48	1
	GABC	1.71E-02	1.95E-02	151300.35	46
	EABC	4.32E-03	8.79E-03	98278.88	85
f_3	ABC	6.83E-06	6.05E-06	1925.52	100
	GABC	6.54E-06	5.76E-06	1204.65	100
	EABC	6.35E-06	6.18E-06	816.13	100
f_4	ABC	7.33E-05	1.76E-04	180578.91	18
	GABC	2.15E-05	8.68E-05	90834.53	97
	EABC	1.49E-05	8.79E-05	63816.73	100
f_5	ABC	1.25E+00	8.32E-01	179015.68	18
	GABC	2.97E-02	8.76E-02	101517.9	95
	EABC	2.93E-02	8.84E-02	107569.72	92
f_6	ABC	2.62E-06	6.75E-06	9163.5	100
	GABC	2.45E-06	6.95E-06	5626	100
	EABC	2.09E-06	7.30E-06	5285	100
f_7	ABC	1.12E+01	8.76E+01	200000	0
	GABC	9.77E+00	8.63E+01	200000	0
	EABC	9.95E+00	8.76E+01	200000	0
f_8	ABC	3.40E+03	1.21E+04	200000	0
	GABC	3.31E+03	1.05E+04	200000	0
	EABC	3.00E+03	1.03E+04	200000	0
f_9	ABC	3.10E-03	1.41E-03	80078.81	82
	GABC	2.85E-06	5.25E-06	39423.63	100
	EABC	1.61E-03	3.75E-04	50717.76	95
f_{10}	ABC	1.75E-06	7.73E-06	16790.5	100
	GABC	1.64E-06	8.04E-06	9349.5	100
	EABC	1.32E-06	8.44E-06	8784	100
f_{11}	ABC	2.18E-06	4.97E-07	106142.69	65
	GABC	4.38E-15	4.80E-15	3965.99	100
	EABC	4.39E-15	4.67E-15	3093.09	100
f_{12}	ABC	6.19E-05	2.15E-05	195783.29	4
	GABC	2.98E-14	4.39E-14	48432.57	100
	EABC	2.80E-14	4.73E-14	25909.31	100
f_{13}	ABC	2.94E-06	1.95E-03	24185.13	100
	GABC	3.03E-06	1.95E-03	4770.87	100
	EABC	2.81E-06	1.95E-03	4311.85	100
f_{14}	ABC	5.47E-06	4.66E-06	4883.53	100
	GABC	5.88E-06	5.09E-06	2450.53	100
	EABC	5.46E-06	4.59E-06	2041.98	100

Table 3 Outcome of Table 2

TP	EABC Vs GABC	EABC Vs ABC
f_1	+	+
f_2	+	+
f_3	+	+
f_4	+	+
f_5	-	+
f_6	+	+
f_7	-	+
f_8	+	+
f_9	-	+
f_{10}	+	+
f_{11}	+	+
f_{12}	+	+
f_{13}	+	+
f_{14}	+	+
Total number of + sign	11	14

TP Test problems

Fig. 1 Boxplots graphs for average function evaluation



- Af^i = Average number of function evaluations used to obtain the required solution of i th problem.
- Mo^i = Minimum of standard deviation got for the i th problem.
- Ao^i = Standard deviation obtained by an algorithm for the i th problem.
- N_p = Total number of optimization problems evaluated.
- k_1, k_2, k_3 = weights assigned to the success rate, the average number of function evaluations and the standard deviation respectively such that $k_1 + k_2 + k_3 = 1$ and $0 \leq k_1, k_2, k_3 \leq 1$.

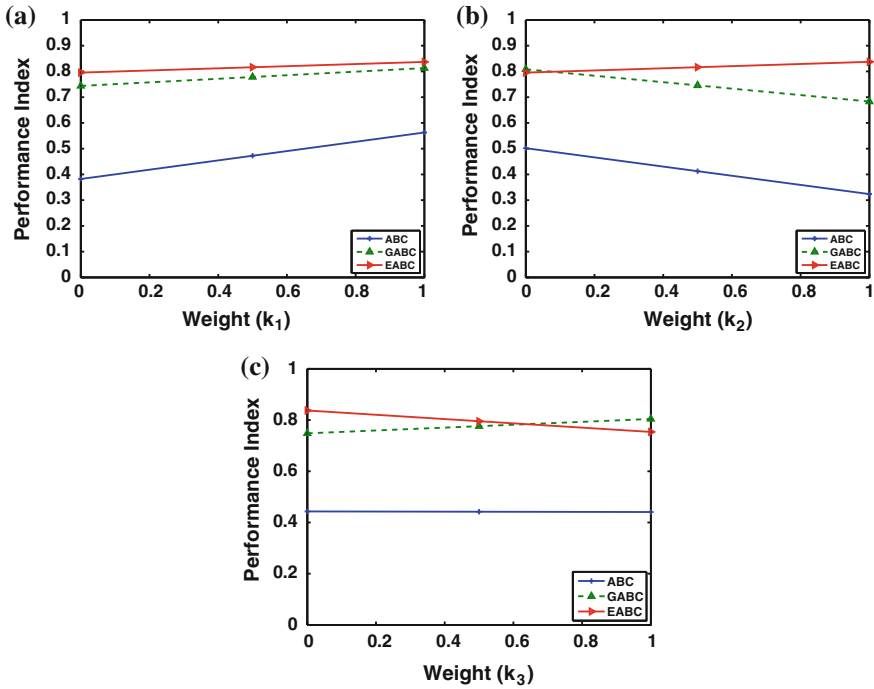


Fig. 2 Performance index for test problems; a for case (1), b for case (2) and c for case (3)

To calculate the *PIs*, equal weights are assigned to two variables while weight of the remaining variable vary from 0 to 1 as given in [3]. Following are the resultant cases:

1. $k_1 = W, k_2 = k_3 = \frac{1-W}{2}, 0 \leq W \leq 1;$
2. $k_2 = W, k_1 = k_3 = \frac{1-W}{2}, 0 \leq W \leq 1;$
3. $k_3 = W, k_1 = k_2 = \frac{1-W}{2}, 0 \leq W \leq 1$

The graphs corresponding to each of the cases (1), (2) and (3) for *EABC*, *ABC* and *GABC* are shown in Fig. 2a–c respectively. In these figures, horizontal axis represents the weights k_1, k_2 and k_3 and while vertical axis represents the *PI*.

For case (1), *PIs* of the considered algorithms are superimposed in Fig. 2a by giving equal weights to average number of function evaluations and the standard deviation. It is observed that *PI* of *EABC* are higher than the considered algorithms. Similarly for case (2), equal weights are assigned to the success rate and standard deviation and for case (3), equal weights are assigned to the success rate and average function evaluations. It is clear from Fig. 2b, c that in these cases also, the *EABC* algorithm has better performance than the others.

5 Conclusion

In this paper, ABC algorithm is modified through alteration in position update equation in onlooker bee phase to balance the exploration and exploitation capabilities of ABC. Instead focusing only on the global best bee, we also emphasize on a random bee with better fitness in the colony and the so obtained modified ABC is named as Expedited ABC (*EABC*). To analyze the proposed algorithm, it is compared to *ABC* and a recent variant *GABC* and with the help of experiments over 14 unbiased benchmark optimization functions, it can be concluded that the *EABC* outperforms to the considered algorithms in terms of reliability, efficiency and accuracy.

References

1. Akay, B., Karaboga, D.: A modified artificial bee colony algorithm for real-parameter optimization. *Inf. Sci.* (2010). doi:[10.1016/j.ins.2010.07.015](https://doi.org/10.1016/j.ins.2010.07.015)
2. Banharnsakun, A., Achalakul, T., Sirinaovakul, B.: The best-so-far selection in artificial bee colony algorithm. *Appl. Soft Comput.* **11**(2), 2888–2901 (2011)
3. Bansal, J.C., Sharma, H., Arya, K.V., Nagar, A.: Memetic search in artificial bee colony algorithm. *Soft Comput.* **17**(10), 1–18 (2013)
4. Bansal, J.C., Sharma, H.: Cognitive learning in differential evolution and its application to model order reduction problem for single-input single-output systems. *Memetic Comput.* **4**, 1–21 (2012)
5. Diwold, K., Aderhold, A., Scheidler, A., Middendorf, M.: Performance evaluation of artificial bee colony optimization and new selection schemes. *Memetic Comput.* **3**, 1–14 (2011)
6. Dorigo, M., Di Caro, G.: Ant colony optimization: a new meta-heuristic. In: Proceedings of the 1999 Congress on Evolutionary Computation (CEC 99), vol. 2. IEEE (1999)
7. El-Abd, M.: Performance assessment of foraging algorithms vs. evolutionary algorithms. *Inf. Sci.* **182**(1), 243–263 (2011)
8. Gao, W., Liu, S.: A modified artificial bee colony algorithm. *Comput. Oper. Res.* **39**(3), 687–697 (2011)
9. Kang, F., Li, J., Ma, Z.: Rosenbrock artificial bee colony algorithm for accurate global optimization of numerical functions. *Inf. Sci.* **181**(16), 3508–3531 (2011)
10. Karaboga, D.: An idea based on honey bee swarm for numerical optimization. Technical report TR06, Erciyes University Press, Erciyes (2005)
11. Karaboga, D., Akay, B.: A comparative study of artificial bee colony algorithm. *Appl. Math. Comput.* **214**(1), 108–132 (2009)
12. Karaboga, D., Akay, B.: A modified artificial bee colony (abc) algorithm for constrained optimization problems. *Appl. Soft Comput.* **11**(3), 3021–3031 (2011)
13. Karaboga, D., Basturk, B.: Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. *Foundations of Fuzzy Logic and Soft Computing*, pp. 789–798. Springer, Berlin (2007)
14. Kennedy, J., Eberhart, R.: Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, vol. 4, pp. 1942–1948. IEEE (1995)
15. Passino, K.M.: Biomimicry of bacterial foraging for distributed optimization and control. *Control Syst. Mag. IEEE* **22**(3), 52–67 (2002)
16. Price, K.V., Storn, R.M., Lampinen, J.A.: Differential evolution: a practical approach to global optimization. Springer, Berlin (2005)

17. Vesterstrom, J., Thomsen, R.: A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In: Congress on Evolutionary Computation (CEC2004), vol. 2, pp. 1980–1987. IEEE (2004)
18. Williamson, D.F., Parker, R.A., Kendrick, J.S.: The box plot: a simple visual method to interpret data. *Ann. Intern. Med.* **110**(11), 916 (1989)
19. Zhu, G., Kwong, S.: Gbest-guided artificial bee colony algorithm for numerical function optimization. *Appl. Math. Comput.* **217**(7), 3166–3173 (2010)

Indian Ehealth Services: A Study

Shilpa Srivastava, Milli Pant and Namrata Agarwal

Abstract Advancement in Information and communication technology has made the healthcare information and services globally accessible. The paper throws light on the impact of e-health services in a developing nation like India, where densely populated communities spread over vast distances. The healthcare delivery systems are being overloaded in the developing and densely populated countries, so it is imperative to have an efficient and cost effective systems. The paper highlights the various social and technical issues in the successful deployment of e-health services in India. Further the inclusion of “Aadhaar” (A unique Identification number provided by UIDAI, an agency of Government of India) has been suggested to make the services more streamlined and secured.

Keywords Ehealth services · UIDAI · Aadhaar

1 Introduction

Information systems and web technologies have changed quality of services and people’s life style. Advances in communication systems have impacted all aspects of our daily life and the field of health care is no exception. The use of Information Communication Technology in the health sector is one of the most rapidly growing areas in health care today [1]. The advancement in computer and communication

S. Srivastava (✉)
RKGIT, Ghaziabad, India
e-mail: shri.shilpa03@rediffmail.com

M. Pant
IIT Roorkee, Saharanpur Campus, Roorkee, India

N. Agarwal
NIFM, Faridabad, India

technology has made the healthcare information and services globally accessible. The definition of e-Health as per WHO is: 'leveraging of the ICT to connect provider and patients and governments to educate and inform healthcare professionals, managers and consumers; to stimulate innovation in care delivery and health system management; and to improve our health care system'.

Millions of people across the world do not have opportunity to obtain the optimum medical services due to high cost and accessibility. IT revolution in healthcare has presented an opportunity for universal access to medical services and information at a very low cost. The medicines, medical education and medical research are progressively going high and so is their cost escalating. This cost can be lowered by making the medical diagnosis and treatment more precise. According to Dr. T. E Bell (IEEE spectrum 2006) the effective and efficient use of engineering can lower the costs provided it is focused on early detection of the disease. Ehealth is a joint effort of healthcare group and the industry that possess technology. The power of the Internet to advance telemedicine was first brought to light by a seminal event in April 1995. An SOS e-mail message was sent through the Internet requesting international help for a Chinese university student, who was suffering from an unknown severe disease. This led to the first recorded Internet diagnosis—of Guillian-Barré syndrome [2]. The term ehealth had been in used since 1960s but the majority of growth has started after 2000. The World Health Organization (WHO), has taken a number of key actions aimed at bringing the power of ICT on health challenges internationally, regional and global levels. Some of the main objectives of E-health can be listed as [3]:

1. Efficient healthcare to lower the cost of healthcare by avoiding unnecessary repetitions of diagnostic procedures, focused diagnosis and judicious use of medicines.
2. Evidence based healthcare emphasizing that clinical practice should be guided not by mere assumptions, intuitions and beliefs, but by rigorous scientific research. Thus reducing medical errors by applying evidence based acquired knowledge by the physicians.
3. Ethical practice that targets the health related privacy issues. Also, incorporating informed consent and security of patient's information during the online clinical practice.
4. Education of physicians through various e-learning modules [4]. Empowerment of patients by distributing the medical knowledge among the patients, to raise their level of awareness.

In this paper the various issues in the deployment of ehealth services shall be studied. Section 2 shall throw light on the status of ehealth in developing countries. Section 3 will bring out the status of ehealth in Indian society. Further the integration of "Aadhaar" (A Unique Identification Number) in the Indian ehealth system has been suggest to make the system more managed.

2 E-Health in Developing Countries

Health systems in low- and middle-income countries continue to face considerable challenges in providing high-quality, affordable and universally accessible care [5]. According to global survey by GOe (Global Observatory by E-Health) the strong growth of ehealth started since 2000 [6]. Developed countries higher and middle income groups are more advanced in their ehealth development. The developing nations still need special attention to ensure their proper development. Policy-makers, donors and program implementers are putting efforts for innovative approaches to eliminate the geographic and financial barriers to health. In this context many health program implementers are exploring the extent to which m-health (ehealth) can help in facing the challenges. It has been observed that there has been an unprecedented increase in the number of users of cell phone and internet technologies, as well as a decline in the price of devices and services, so this information can be well utilized in the implementation of ehealth [7]. For e.g.: with more than 900 million cell phone connections, India is the second largest mobile market in the world after China. A recent report by Boston Consulting Group stated that the total number of Internet users in India is expected to increase from 125 million in 2011 to 330 million by 2016. The number of internet users through mobile will be also increased in the coming future. Hence mobile phones can be considered as an efficient way to tap the online services like ehealth.

Despite the increased interest, in low- and middle-income countries the e-health field is still relatively nascent. To date, the literature on e-health in low- and middle-income countries has largely consisted of articles describing single uses of technology in health care delivery [5]. Developing countries are faced with many problems in the healthcare and medical services such as limited financial resources, lack of trained personnel, reach ability to remote locations, etc. The reason may include various interrelated factors such as poverty, malnutrition, poor hygiene, living environment, gender or caste based discrimination etc. According to [8] a model was designed considering the problems and challenges of e health in developing countries. It includes:

1. Technical & operational: loss of EHR, security issues, suitable medical equipment, maintenance & space support etc
2. Social & Cultural: Education of e-health services, disbelief of people in new services, resistance against change due to habits
3. Naive Environment: lack of coverage of various regions in using medical equipments
4. Legal: provision of legal aspects, balancing of laws and regulations
5. Policy: Lack of comprehensive and national strategies regarding e-Health
6. Financial: Need for investment and allocation of regular budget in e-Health field.

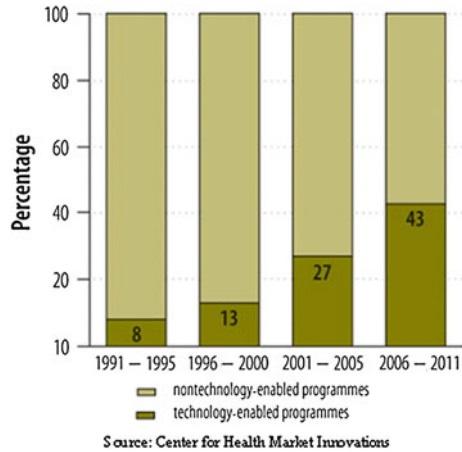
Developing countries should focus on e-health, in their strategic plan of nationwide health care, and subsequently the e-Health services. According to World Bank figures, expenditure on health in developed countries is 11 percent of their GDP as compared to just 6 % on an average for developing countries [9]. An analysis of Center of Health Market Innovations (CHMI) states that ICT may play a crucial role in improving the performance of health care sector in developing countries, as ICT enables access to health services delivery through proper network, coordinates research and knowledge management.

3 E-Health in Indian Context

India is a vast country consisting of 28 states and 7 union territories. Health is the primary responsibility of each state and there is paucity of infrastructure and dearth of doctors in rural areas. A survey by the Indian medical society has found that 75 % of qualified consulting doctors practice in urban centers and 23 % in semi urban areas and only 2 % from rural areas whereas majority of the patients come from rural areas [10]. Besides, there is no national health insurance scheme in the country. India has emerged as the leader in telemedicine with 380 plus telemedicine centers operating across the country for providing healthcare services to remote areas but unfortunately the percentage of active services are very less. Out of many challenges like poor reach of ICT services in rural areas, low literacy leading to low awareness, insufficient infrastructure, poverty, poor data management is also one of the major deterrents to large scale adoption of e-health. Efforts are directed towards setting up standards and IT enabled healthcare infrastructure in the country. Government, administrative bodies and the different players in the health service system are looking for innovative solutions to make health services most efficient and secure. In spite of many problems, it is heartening to note that India is among those developing nations, where the progress in ehealth has been encouraging. On this front India is much better placed than other developing countries. According to CHMI report the number of technology enabled programs in health care are emerging in all lower-income countries, as shown in Fig. 1. Southern Asia—India in particular—leads in terms of the absolute number of technology-enabled programs [5].

India pioneered first telemedicine connectivity in rural areas by connecting the village Aragonda through satellite connection in the year 2000. This was first of its kind in the world. The gap is not so much in terms of technical knowledge and actual infrastructure. India has potential and necessary expertise required to implement ehealth projects. Several state governments have earnestly taken e-health Monitoring and Reporting Systems for their public health programs. India is acquiring a sizable market in health care.

Fig. 1 Percentage of programmes using information and communications technology, by year launched. *Source* center for health market innovations



3.1 E-Health Initiatives in India

3.1.1 Telemedicine Initiatives

Telemedicine Program is a process of uniting the benefits of Satellite communication technology and information technology with Biomedical Engineering and Medical Sciences to deliver the health care services to the remote, distant and underserved regions of the country. The first telemedicine project was established by the Apollo group of hospitals in a village Aragonda in Andhra Pradesh in the year 2000. It started with simple web cameras and ISDN telephone lines and today, the village hospital has a video conferencing system and a VSAT (Very Small Aperture Terminal) satellite installed by ISRO (Indian Space Research Organization). In India the telemedicine program is mainly supported by:-

- Department of Information Technology (DIT)
- Indian Space Research Organization
- NEC Telemedicine program for North-Eastern states
- Apollo Hospitals
- Asia Heart Foundation
- State governments.

Telemedicine technology is also supported by some other private organizations. The telemedicine software system has been developed by the Centre for Development of Advanced Computing, C-DAC which supports Tele-Cardiology, Tele-Radiology and Tele-Pathology etc. It uses ISDN, VSAT, POTS and is used to connect the three premier Medical Institutes of the country (viz. All India Institute of Medical Sciences (AIIMS), New Delhi, Sanjay Gandhi Post Graduate Institute of Medical Sciences (SGPGIMS), Lucknow and Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh). Now it is being connected to include Medical centers in Rohtak, Shimla and Cuttack [11]. The main purpose

of telemedicine is to connect the remote District Hospitals/Health Centers with Super Specialty Hospitals in cities, through the INSAT Satellites for providing expert consultation to the needy and underserved population.

The major initiatives in the Telemedicine field at ISRO are:

- Providing Telemedicine Technology & connectivity between remote/rural hospital and Super Specialty Hospital for Tele-consultation, Treatment & Training of doctors & paramedics.
- Providing the Technology & connectivity for Continuing Medical Education (CME) between Medical Colleges & Post Graduate Medical Institutions/Hospitals.
- Providing Technology & connectivity for Mobile Telemedicine units for rural health camps especially in the areas of ophthalmology and community health.
- Providing technology and connectivity for Disaster Management Support and Relief.

The mobile vans are extensively used for Tele-ophthalmology, diabetic screening, mammography, child care and community health. The Mobile Tele-ophthalmology facilities provide services to the rural population in ophthalmology care including village level eye camps, vision screening for Cataract/Glaucoma/Diabetic Retinopathy. The telemedicine facilities are established at many remote rural district hospitals in many states and union territories of the country including Jammu & Kashmir, Andaman & Nicobar Islands, Lakshadweep Islands, North Eastern States etc. State level telemedicine networks are established in Karnataka, Kerala, Rajasthan, Maharashtra, Orissa and Chhattisgarh. Many interior districts in Orissa, Madhya Pradesh, Andhra Pradesh, Punjab, West Bengal and Gujarat have the telemedicine facility. About 1.5 lacs patients are getting the benefits of Telemedicine every year [12].

Apollo Telemedicine Networking Foundation (ATNF), a part of Apollo Hospitals Group has also emerged as India's single largest provider in the area of Telemedicine with over 125 peripheral centers including 10 overseas. ATNF works with multiple entities including the medical bodies, private and public sectors, Central and State Governments both at a domestic and international level to popularize the concept of Telemedicine. ATNF adheres to all legal and patient security laws guaranteeing patient confidentiality and security of data transmitted via the telemedicine network. Security is ensured by using secure login information, encrypted databases, secure websites and identification through hardware protocols including IRIS identification, when required. ATNF was the first to promote clinical Telemedicine in South Asia [13].

3.1.2 eHCF

eHealth-Care Foundation, (eHCF) a New Delhi based Not for Profit Organization. It is a web based patient care system that tracks a patient's health record. It offers them an opportunity to maintain their health record and authorize their physicians

to access it as needed. Physicians accessing the patient's data must have the capability to access the data. The Mission of eHealth-Care is to help patients, physicians, and community hospitals to make appropriate use of information and communication technologies (ICTs) in order to improve access and quality of healthcare delivery and reduce the cost of its management.

3.1.3 Recent Projects in Ehealth

- **Cloud Enabled E-health Center:-** India's first fully integrated cloud based e-health center was launched on 1st December 2012 at the Chausala village in Haryana Kaithal District. It was a joint effort of Council of Scientific Research and Industrial Research (CSIR) and Hewlett Packard. It was set up to provide affordable health services to remote areas. It has diagnostic equipment, a tele-medicine studio, laboratory and pharmacy and minimizes the need for onsite doctors in remote areas.
- **Virtual Medical Kiosk:-** E-health Access Pvt. Ltd, a healthcare based company launched Virtual Medical Kiosk in November 2012, which enables patient-doctor consultation in a secure environment. Patients and doctors can communicate through phone, web cams, video conferencing, messaging, or chat. The kiosk is embedded with diagnostic equipments, scanners and medical management software. It is first-of-its-kind. The main purpose of this medical kiosk is to act as a bridge to bring the medical practitioner closer to the people in need for medical consultation.
- **Alcohol Web India:-** An ehealth portal on alcohol use was developed by National Drug Dependence Treatment Centre (NDDTC) New Delhi as part of an initiative by the WHO, Geneva. The aim of this ehealth portal is to address the problem of alcohol use. This portal has sections devoted to alcohol users, family members of alcohol users, health professionals and policy makers.
- **RFID Individual Tracking and Records Management (RFID-ITRM) e-health project at Ahemdabad, Gujrat:-**IEEE launched this project successfully in Ahemdabad. RFID-ITRM technology is central to preventing medical errors, identifying victims of natural disasters, and tracking and monitoring diseases and outbreaks, as well as infants and vaccination history. An electronic medical record system is installed in a local community health care center. The system is managed by local NGO Manav Sadhna.
- **E-health Project at Punjab:-** An e-health clinic was established in Punjab in Malwa region. A Hyderabad based NGO Naandi Foundation played a major role in launching this project. This e-health clinic offers wide range of medical services for chronic disease like cancer apart from specialized health care services through telemedicine and broad band electronics methodology.

Besides this, the most important initiative being taken is standardization of exchange of health information between different entities within the healthcare sector. In this regard the ministry of health & family welfare and the ministry of communication and information technology are jointly creating a national health information infrastructure for easy capture and dissemination of health information and recently the government has introduced National Health Portal (NHP) which is scheduled to be launched on October 2013 [9, 14]. The NHP is a Union Ministry of Health and Family Welfare Project. The portal is intended to plug in the healthcare gaps through the effective use of IT. It will establish a national database for the medical records of all the citizens from birth to death. The three main objectives of NHP are:-

- To improve the health awareness among the masses India.
- Improvement in the access to health services across the country.
- Decrease the burden of disease by educating people on the preventive aspects of disease.

The NHP's main aim should be to reach and serve the 330 million Indians below poverty line through innovative ways. The challenge is that nearly 300 million of them are illiterate and less than one percent of them are reachable via internet.

4 Significance of Aadhaar in Ehealth

4.1 About Aadhaar

Aadhaar is a 12 digit unique identification number provided by UIDAI (Unique Identification Authority of India-A Government Body under Planning Commission Of India, established in 1999). It is based on the demographic and biometric information. This will ensure that the data collected is clean from the beginning of the program. The UIDIA will be the regulatory authority managing a Central Identification Repository (CIDR) which will issue Aadhaar, update resident information and authenticate the identity of the resident's whenever required [15]. There is a process to ensure that there will be no duplicate records. Residents have only one chance to be in the database. So the individuals are required to provide accurate data as many benefits and entitlement are going to be linked in future and "Aadhaar" will over time be recognized and accepted across the country and across all service providers. Efforts have already been started in introducing "Aadhaar" cards at school and college level. The University Grants Commission (UGC) has issued a letter to the State universities to enroll students for the Unique Identity Card [16]. Linking Aadhaar with students' bank accounts would help in disbursal of scholarships and fellowships. The education department is also taking initiative in enrolling the teaching and non-teaching staff for the Aadhaar card.

4.2 Security and Privacy in “Aadhaar”

- Online Authentication:- The UIDAI will provide a strong form of online authentication where the private and public agencies can compare demographic and biometric information of the resident with the record stored in the central database. It will answer all requests to authenticate identity only through a ‘Yes’ or ‘No’ response. For example banks can link the unique number to a bank account for every resident, and use the online authentication to allow its customer to access the account from anywhere in the country.
- Data Transparency:- All the aggregated data shall be available for public to access under RTI (Right to information act) but the Personal Identity Information (PII) will not be accessible by any entity [15].

The inclusion of Aadhaar will provide a strong authentication in e-health services. Several proposals have been given in integrating UID with Indian health services [17]. Besides this the Ministry of family and healthcare has also planned to integrate “Aadhaar” with NHP in the near future. Figure 2 describes the model for integrating Indian e-health system with “Aadhaar”. In this figure the Authentication in ehealth system is provided through the “Aadhaar” and the privileges shall be granted through the administrator. The Aadhaar number will be authenticated at CIDR (central identification repository (CIDR) maintained with UIDAI.

5 Challenges in Ehealth

The main aim behind promotion of e-Health is to try to provide a modern, up-to-date and quality healthcare service to far flung corners of India where the formal medical facilities are either totally absent or of poor quality. Other major benefits include continuous training of medical and Para-medical personnel, to bring awareness about the latest advances in medical field. The improvement in overall quality of healthcare personnel will benefit the entire local population. EHealth concept has been around for sometime but the developing countries are facing multiple issues like resource crunch, lack of finance, poor awareness etc. Some of the biggest issues which are hindering widespread deployment of eHealth are:-

- Lack of infrastructure: To support eHealth services, huge investments are required in basic infrastructure to ensure reliable power supply, proper road network and adequately fast Internet connectivity. Unfortunately in a developing country like India, there is insufficient availability of basic infrastructure. In India, tarred road is not available in nearly 66 thousand villages having a population of more than five hundred. Clean drinking water is not available in over fifty five thousand villages. The power supply is unreliable even in major cities while over a hundred thousand villages are not connected to power grid.

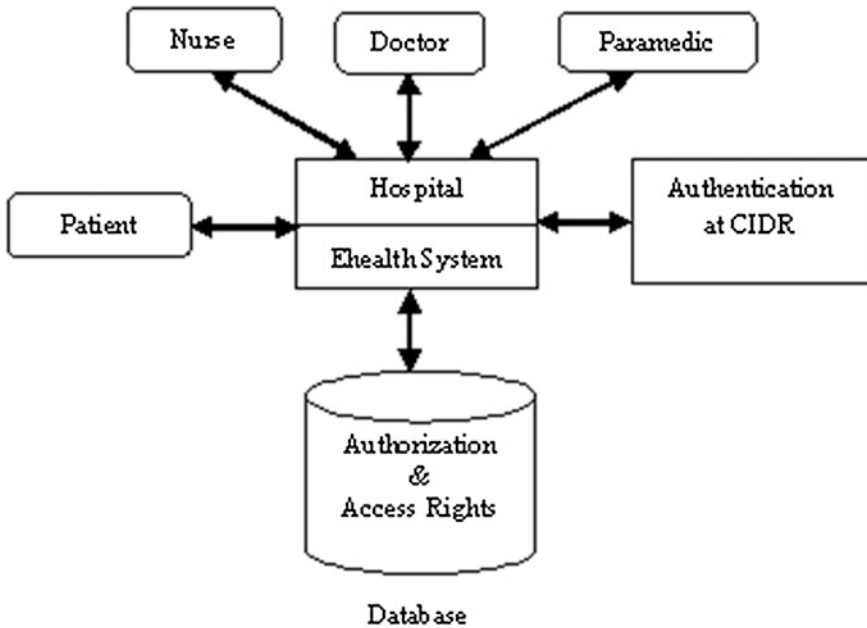


Fig. 2 Indian Ehealth system

- Lack of technically competency: A society which wants to use the ICT tools needs to have a general level literacy. In addition to medical training, the personnel involved in healthcare system from specialist doctors to low level staff, need to build competency in IT tools. In order to maximize the benefits of eHealth solutions, everybody, from top managers to workers in the field, should be familiar with the technical aspects. In a developing country, the basic level of literacy is generally poor. One of the aims for the planners of eHealth in a country like India will be to train the entire chain of health care workers.
- Financial support: As per World Bank analysis, the developed countries spend an average of 11 % of their GDP on Health related services while this figure is only 6 % for developing countries [9]. This clearly demonstrates the challenge of funding new initiative in Health care in developing countries. There is a need to allocate a regular fund in budget for improving the systems involved in eHealth. Ehealth combined with proper ICT tools has the potential to serve a larger proportion of population with latest medical knowledge with lower average cost per person.
- Awareness of ehealth: There is a general lack of confidence in the effectiveness of treatment via e-Medicine both amongst the health care professional as well patients. Even the top of the pyramid Doctors are not familiar with e-medicine tools and usage. A small survey was conducted in the NCR Delhi (National Capital Region). The aim of the survey was to know the acquaintance level of

Fig. 3 Ehealth awareness



educated persons living in the NCR region regarding ehealth. The results have been depicted in the Fig. 3.

- So the lack of awareness of people of developing countries towards ICT innovations and methods, is one of the biggest impediments towards wide spread roll out of eHealth. Even when health care information is disseminated via ICT initiatives, the general public may is still ignorant of its benefits.
- Security Issues: The ICT tools and systems used for eHealth are prone to the same security threats as other online applications. The problem is more pronounced for eHealth systems as utmost privacy, safety and anonymity of information regarding patients, doctors and line of treatment is required. The infrastructure and competencies required to protect data and its channels of transfer from one location to the other is severely limited in a developing country like India and is one of the road blocks for large scale rollout of eHealth. All the stakeholders involved in eHealth, specially the patients, must have utmost confidence that their privacy is fully secure and will not be compromised due to a security breach. There should be specific regulations and legislation by the government, clearly spelling out the protection of privacy and security of individual patients' data in eHealth tools and systems. Currently, the focus on protection of citizen's data and its confidentiality is limited in general and more or less absent for eHealth.
- Legal Issues: The absence of a coherent and clear law about eHealth is also one of the major problems. Systematic and coordinated development of ICT initiatives requires such a legal framework. So far we don't have dedicated e-health and telemedicine laws in India. The first Telemedicine project in India started in year 2000 without the establishment of legal protocols. We have an IT Act, but it does not specifically deals with eHealth. The laws are silent with respect to privacy and data protection issues. There is an urgent requirement of dedicated law which specifically deals with eHealth. It would be better for the larger interest of medical community and patients in India if such a law is enacted soon.

- **Multilingualism and cultural diversity:** If the information and services are provided in the local language, there is a greater chance that the people will respond to it. On the other hand, information in an unfamiliar language is likely to be cast aside, even if it has potential to benefit the target population. The creation of electronic eHealth content in local languages and modified according to the local cultural values is regarded as an essential element in widespread deployment and usage of eHealth. In India, the Centre for Development of Advanced Computing (CDAC) is developing and promoting computers that are capable of multilingual support. The Indian Institute of Technology Kanpur along with CDAC is running programs that are providing guidelines for prevention and control of diseases and other medical literature in the local languages to field workers of different regions.
- **Interoperability:** Rapid and secure access to information such as public health data records, patient information etc. requires seamless inter-working of diverse systems and services of ICT. This is a complex scenario as it depends on standardization of many inter dependent components like health care information systems, data records repository and patient identification systems. Interoperability between eHealth operations within an institution, a region, a country or internationally requires standardization. As the level of standardization increases, more and more developers will get involved and the users will have a greater choice of off-the-shelf components without worrying about interoperability. The entire economics of eHealth improves as the components and services progressively become cheaper to deploy and use. The transactional cost involved in data and information transfer between systems drops thereby allowing for bigger deployment of eHealth network.

6 Conclusion

The adoption of amalgamation of ICT and medicine should be taken care to see that the benefit of ICT revolution helps to transmit the medical facility and medical knowledge effectively to the underserved population of India, without compromising on their privacy. The success of ehealth is dependent upon the extent to which consumers can get online. The challenges in ehealth will not stop the stakeholders to move ahead. There is a strong belief that health care stakeholders and the Government will work together to overcome the current obstacles. Ministry of Health and family welfare, Government of India has been working hard to develop a National Policy on Telemedicine and e-health. Besides this the Government of India is also launching National Health Portal in October 2013. To make the services streamlined and secured the ministry of health and family welfare is also planning to integrate “Aadhaar” Unique Identification of India, with NHP in near future. Public private partnership has been successfully used to maximize the effectiveness of technical resources and personnel in ehealth sector. Funding and technical assistance has also been received from the external agencies

such as World Health Organization (WHO). Developing countries should accord health service a very important place in their strategic plan for improving the productivity and life expectancy of its citizens. The wide deployment of secured health is need of the hour, which shall definitely help us travel towards a healthy India.

References

1. Dong, M., Lee, B.W., Yoon, S.W., Park, C.O.: Actual condition of Korean e-Health: what do enterprisers want for developing e-Health industry? In: Proceedings of 9th International Conference on e-Health Networking, Application and Services, Taipei, pp. 304–307, (2007)
2. Subcommittee on telemedicine. Emergency medicine practice committee. Telemedicine in emergency medicine. American College of Emergency Physicians (1998)
3. Wickramasinghe, N.S., Fadlalla, A.M.A., Geisler, E., Schaffer, J.L.: A framework for accessing e-Health preparedness. *Int. J. Electron. Healthc.* **1**(3), 316–334 (2005)
4. Hinske, S., Ray, P.: Management of e-Health network for disease control: a global perspective. In: Proceedings of Healthcom. 9th International Conference on e-Health Networking, Application and Services, Taipei, pp. 52–57 (2007)
5. Bulletin of the World Health Organization. **90**:332–340 (2012). doi: [10.2471/BLT.11.099820](https://doi.org/10.2471/BLT.11.099820)
6. <http://www.who.int/GOe>
7. Center for health market innovations. Pesinet. Washington: CHMI; 2010. Available from: <http://healthmarketinnovations.org/program/pesinet>
8. Khalifensoltani, S.N., Germani, M.R.: EHealth challenges and opportunities and experiences of developing countries. In: Proceedings of International Conference on e-Education, e-Business, e-Management, and e-Learning, IC4E '10, pp. 264–268, 22–24 (2010)
9. Srivastava, A.: e-Health in developing countries: pitfalls, challenges and possibilities <http://www.e-healthmagazine.com>
10. Mishra, S.K., Gupta, D., Kaur, J.: Telemedicine in India: initiatives and vision. In: 9th International Conference on e-Health Networking Applications and services, Taipei, pp. 81–83 (2007)
11. Bedi, B.S.: Telemedicine in India: initiatives and perspective, e-Health 2003: addressing the digital divide (2003)
12. <http://www.isro.org/scripts/telemedicine.aspx>
13. <http://www.telemedicineindia.com/>
14. Ehealth Magazine, source (2013): <http://ehealth.eletsonline.com/2013/06/govt-to-launch-national-health-portal>
15. <http://uidai.gov.in/what-is-aadhaar.html>
16. http://negp.gov.in/index.php?option=com_newslatest&view=content
17. Gp Capt (Dr) Sanjeev Sood: Aadhaar opening up of new vistas in Healthcare. Ehealth Magazine, 50 (2012)

Cluster Based Term Weighting Model for Web Document Clustering

B. R. Prakash, M. Hanumanthappa and M. Mamatha

Abstract The term weight is based on the frequency with which the term appears in that document. The term weighting scheme measures the importance of a term with respect to a document and a collection. A term with higher weight is more important than a term with lower weight. A document ranking model uses these term weights to find the rank of a document in a collection. We propose a cluster-based term weighting models based on the TF-IDF model. This term weighting model update the inter-cluster and intra-cluster frequency components uses the generated clusters as a reference in improving the retrieved relevant documents. These inter cluster and intra-cluster frequency components are used for weighting the importance of a term in addition to the term and document frequency components.

Keywords Term weighting scheme · Document clustering · Information retrieval · Data mining

1 Introduction

A document clustering algorithm helps to find groups in documents that share a common pattern [1–5]. It is an unsupervised technique and is used to automatically find clusters in a collection without any user supervision. The main goal of the clustering is to find the meaningful groups so that the analysis of all the documents within clusters is much easier compared to viewing it as a whole collection.

The Vector Space Model (VSM) represents a document using a vector of T unique terms in a collection (T -dimension). Each term in a vector is associated

B. R. Prakash (✉) · M. Hanumanthappa
Department of Computer Science and Applications, Bangalore University, Bangalore, India
e-mail: India.brp.tmk@gmail.com

M. Mamatha
Department of Computer Science, Sri Siddaganga College for Women, Tumkur, India

with a weight (term weight) [6, 7]. The term weight is based on the frequency with which the term appears in that document. The term weighting scheme measures the importance of a term with respect to a document and a collection. A term with higher weight is more important than a term with lower weight. Each document can be located in T-dimensional space, where T is the number of unique terms in a collection (Euclidean space). With a document represented by a location in Euclidean space, we can compare any two documents by measuring the actual distance between them. In the same way, a user-supplied query can be represented as a vector and mapped in Euclidean space. In order to find a set of documents relevant to a query, we can find documents that are closer to this query in Euclidean space. A document ranking model finds the similarities between these documents and a query. If a document is more relevant to a query, it will get a higher ranking. VSM and term weighting schemes are widely used in many research areas such as document clustering, classification, information retrieval, document ranking, etc.

2 Cluster-Based Retrieval

Cluster-based retrieval uses the cluster hypothesis to retrieve a set of documents relevant to a query [8]. Cluster hypothesis Documents in the same cluster behave similarly with respect to relevance to information needs [9]. If a document in a cluster is relevant to a query, then the rest of the documents in that cluster are potentially relevant to the same query. There are two approaches in cluster-based retrieval. The first approach retrieves one or more clusters relevant to a query instead of retrieving documents relevant to a query. In other words, this approach retrieves and ranks the relevant clusters instead of the relevant documents. Based on the cluster hypothesis, the documents from the highly ranked clusters are more relevant to a query than the documents from the clusters with lower ranking. The main motive of this approach is to achieve higher efficiency and faster search.

The second approach uses the generated clusters as a reference in improving the retrieved relevant documents. In this approach, the given document collection is clustered (static clustering) beforehand. When a set of documents is retrieved for a query, the generated clusters (static clusters) of the collection are used as a reference to update the retrieved relevant document list. The main goal of this approach is to improve the precision-oriented searches.

3 Term Weighting Scheme

The Boolean retrieval model assigns 1 or 0 based on the presence or absence of a term in a document. This model performs undesirably in querying for a document. Later, VSM was introduced for ranked retrieval [10]. It is widely used in querying

documents, clustering, classification and other information retrieval applications because it is simple and easy to understand. It uses a bag of word approach. Each document d_i in the collection ζ is represented as a vector of terms with weight [11, 12].

One of the most commonly used term weighting schemes, TF-IDF, assigns weights to each term using the term frequency (tf) and inverse document frequency (idf). The term frequency of the term t_i is the number of times the given term t_i occurs in a document. The inverse document frequency is the total number of documents in a collection containing the term t_i with respect to the total number of documents (N) in a collection. Then, the document vector d_i , represented as:

$$d_i = \left\{ \text{term}_1, \text{tf}_{i1} \cdot \log \frac{N}{N_1}; \text{term}_2, \text{tf}_{i2} \cdot \log \frac{N}{N_2}; \dots \text{term}_T, \text{tf}_{iT} \cdot \log \frac{N}{N_T}; \right\}$$

The term weight w_{it} determines whether the term t_i will be included in the further steps. Only certain terms extracted from a document can be used for identifying and scoring a document within the collection. The term weighting schemes are used to assign weight to each term in a document. These term weights represent the importance of a term with respect to a collection. Document clustering uses these term weights to compare two documents for their similarity.

Table 1 shows representation of some of the term weighting schemes commonly used. Here, TF is the Term Frequency, IDF is the Inverse Document Frequency and ICF is the Inverse Cluster Frequency.

w_{ij} is the weight of the term t_i in the document d_j of the cluster C_j .

$\text{tf}_{if} = \text{fit}$ is the term frequency of the term t_i in the document d_i .

$\text{idf}_t = \log \frac{N}{N_t}$ is the inverse document frequency for the term t_i in the collection

where N is the total number of documents in the collection and d_t is the number of documents that contain the term t_i .

$\text{icf}_t = \log \frac{K}{k_t}$ is the inverse cluster frequency of the term t_i in the collection ζ , where K is the total number of clusters in the collection and K_t is the number of clusters that contains the term t_i .

We present a new term weighting method based on the traditional TF-IDF term weighting scheme. Our motivation is based on the idea that the terms of the documents in a same cluster have similar importance compared to the terms of the documents in a different cluster. We concentrated on the terms that are important within a cluster and considered the other terms as irrelevant and redundant. We presented this idea by giving more weight to the terms that are common within a cluster but uncommon in other clusters.

In our new term weighting scheme, we used unsupervised partitionial (K-means) clustering algorithms [2, 13, 14]. First, we ran the K-means algorithm with the four term weighting schemes, to show that the CBT term weighting scheme improves the quality of the clusters generated by a partitionial clustering algorithm.

Table 1 Term weighting schemes

$$\text{Norm - TF}w_{it} = \frac{f_{it}}{\sqrt{\sum_t f_{it}^2}}$$

$$\text{TF - IDE } w_{it} = f_{it} \log \frac{N}{N_t}$$

$$\text{TF - IDE - ICF } w_{ij} = f_{it} \log \frac{N}{N_t} \log \frac{K}{K_t}$$

Table 2 List of components in CBT term weighting scheme

tf_{it}	Term frequency component. High when term t occurs often in a document i
idf_t	Collection frequency component. High when term t occurs less of ten in the entire collection
df_{ij}	Intra-cluster frequency component. High when term t occurs more often in a cluster j
icf_t	Inter-cluster frequency component. High when term t occurs less often in clusters other than cluster j

From our experiment, we found that the new term weighting scheme based on the clusters gives better results than the other well-known term weight schemes traditionally used for both partitional and hierarchical clustering algorithms.

3.1 The Proposed Term Weighting Scheme

We introduce our new term weighting scheme. For the term $term_t$, document d_i and cluster C_j , CBT is given as:

$$w_{ij} = tf_{it} \cdot idf_t \cdot df_{ij} \cdot icf_t$$

$$= f_{it} \cdot \log \frac{N}{N_t} \cdot \frac{df_j}{|c_j|} \cdot \log \frac{K}{K_t}$$

Here, $df_{ij} = df_j$

$jC_{j,j}$ is the document frequency of the term $term_t$ within the cluster C_j , where df_j is the number of documents in the cluster C_j that contain the term $term_t$, and $jC_{j,j}$ is the total number of documents in the cluster C_j .

Our new term weighting scheme has four components. The first two components are based on the term weighting components discussed in [15]. The last two components are the cluster components as shown in Table 2. In other words, CBT assigns a weight to a term which is

- Highest when the term occurs more frequently in the documents of a cluster and uncommon in other clusters.
- Higher when the term occurs less frequently in the documents of a cluster and uncommon in other clusters.
- Lower when the term occurs often in a few clusters.

Lowest when the term occurs in most of the documents in a collection.

4 K-Means Algorithm with CBT

Initially, the K-means algorithm doesn't have any information about the cluster components, so we start the algorithm by setting df_{ij} and icf_t to 1 and update the inter- and intra-cluster components on each iteration. If a document has a set of terms that doesn't belong to a cluster, then its term weight will be reduced so that it will move to other clusters. It will be repeated until it finds a suitable cluster of its type.

```

Require: An integer  $K \geq 1$ , Document Collection  $\zeta$ 
1: if  $K = 1$  then
2:   return  $\zeta$ 
3: else
4:   Initialize  $l = 0$ 
5:    $\{C_1^{(0)}, \dots, C_k^{(0)}\} \leftarrow \text{RANDOM CLUSTERS}(\zeta, K)$ 
6:   repeat
7:     for all  $d_i \in \zeta, i : 1 \dots N$  do
8:        $m = \arg \min_j |c_j - d_i|$ 
9:        $C_m^{(l+1)} \leftarrow C_m^{(l)} \cup d_i$ 
10:    end for
11:     $l \leftarrow l + 1$ 
12:     $w_{ij} = df_{it} \cdot idf_t \cdot df_{ij} \cdot icf_t$ ;
for each term  $term_t$  in a document  $d_i$  for a cluster
 $C_j^{(l)}, t : 1 \dots T, i : 1 \dots N, j : 1 \dots K$ 
13:   for  $j = 1$  to  $K$  do
14:      $C_j \leftarrow \frac{1}{|C_j^{(l)}|} \sum_{d_i \in C_j^{(l)}} d_i$ 
15:   end for
16:   until No change in  $K$  centroids
17:   return  $\{C_1^{(l)}, \dots, C_k^{(l)}\}$ 
18: end if

```

4.1 Data Collections for CBT

We used the TREC [16], 20 Newsgroup and Reuters-21578 [17] data collections for our experiment. TR11, TR12, TR23, TR31, and TR45 collections are taken from TREC-5, TREC-6 and TREC-7. 20 NG S1–S5 are the five randomly chosen subsets of 20 Newsgroup documents [18]. RE S1 and RE S2 data sets are from Reuters-21578 collection. We got 4645 documents that have only one category. In addition to that, we used the Reuters transcribed subset (RE S2) [19]. For all the data sets shown in Table 3, we removed the stop words and stemmed using the Porter stemming algorithm [20].

Table 3 Data sets

Data set	Collection	No of documents	No of class
TR11	TREC	414	9
TR12	TREC	313	8
TR23	TREC	204	6
TR31	TREC	927	7
TR45	TREC	690	10
20 NG S1	20 newsgroup	2,000	20
20 NG S2	20 newsgroup	2,000	20
20 NG S3	20 newsgroup	2,000	20
20 NG S4	20 newsgroup	2,000	20
20 NG S5	20 newsgroup	2,000	20
RE S1	Reuters-21578	4,645	59
RE S2	Reuters-21578	200	10

Table 4 K-means clustering algorithm—avg. Entropy measured for norm TF, CBT, TF IDF ICF and TF IDF term weighting schemes

Data sets	Term weighting schemes			
	Norm TF	TF-IDF	CBT	TF-IDE-ICF
TR11	0.8413	0.7905	0.8535	0.7749
TR12	0.9009	0.6834	0.6139	0.6261
TR23	1.0424	0.9246	0.839	0.8501
TR31	0.9379	1.2781	0.9657	1.0822
TR45	1.0787	1.3469	1.0443	1.1485
20 NG S1	2.4824	0.4037	0.2995	0.4475
20 NG S2	2.4954	0.5791	0.4164	0.801
20 NG S3	2.4727	0.9366	0.5082	0.7886
20 NG S4	2.4923	0.3138	0.2845	0.521
20 NG S5	2.7464	0.2983	0.3274	0.5665

Table 5 Bisecting K-means clustering algorithm—avg. Entropy, avg. F-measure and avg. Purity measured for the TF-IDF and CBT term weighting schemes

Data source	CBT			TF-IDF		
	Avg entropy	Avg F-measure	Avg purity	Avg entropy	Avg F-measure	Avg purity
TR11	1.3435	0.2067	0.5039	1.4102	0.2478	0.485
TR12	1.548	0.2256	0.3936	1.7344	0.1946	0.3514
TR23	1.3337	0.1803	0.4853	1.3351	0.1719	0.4853
TR31	1.2539	0.1473	0.515	1.4105	0.1407	0.4344
TR45	1.507	0.3003	0.4404	1.5922	0.2627	0.421
RE S1	2.0039	0.0504	0.414	2.0061	0.0519	0.4137
RE S2	1.9169	0.2663	0.2764	1.9981	0.2444	0.2518
20 NG	2.8548	0.09863	0.1079	2.2575	0.1894	0.2141

5 Conclusion

Importance of using inter- and intra- cluster components in the term weights using the average entropy measure. Since the K-means clustering algorithm is unstable and sensitive to initial centroids, we ran the algorithm 10 times with different random seed for the initial centroids on each run. We repeated this experiment for the four term weighting schemes on the data collections listed in Table 3.

We calculated the entropy for the term weighting schemes, as given in Eq. (2.13), for each run after the algorithm converged. Then, we computed the average of the entropies obtained in each run. Similarly, we computed the average F-measure and average purity measures. Table 4 shows the average entropy calculated for each data set. Table 5 shows the average entropy, average F-Measure and average purity measured for the TF-IDF and CBT term weighting schemes for the K-means clustering algorithm. Both experiments show that the results obtained from the K-means clustering algorithms with the CBT term weighting scheme have shown better results compared to the other term weighting schemes on each data set. According to the cluster-based term weighting scheme, a term is considered important to a cluster if it is unique to that cluster and occurs frequently within the documents of that cluster. The inter- and intra- cluster components try to identify these important terms by analyzing the term frequency distribution at three levels: document, cluster and collection. And our experimental results have shown that adding these cluster components in the term weighting scheme significantly improves the results on each data set. We believe that some of the deviations in the results are due to the clustering algorithms' lack of handling the noise in the data collection.

References

1. Tan, P.N., Steinbach, M., Kumar, V.: Introduction to Data Mining, vol. 8, 1st edn. Addison-Wesley, Boston (2005)
2. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: KDD Workshop on Text Mining, vol. 400, pp. 525–526. Department of Computer Science and Engineering University of Minnesota, Citeseer (2000)
3. Guha, Sudipto, Rastogi, Rajeev, Shim, Kyuseok: CURE: an efficient clustering algorithm for large databases. *ACM SIGMOD Rec.* **27**(2), 73–84 (1998)
4. Zhao, Ying, Karypis, George, Fayyad, Usama: Hierarchical clustering algorithms for document datasets. *Data Min. Knowl. Disc.* **10**(2), 141–168 (2005)
5. Cutting, D.R., Karger, D.R., Pedersen, J.O., Tukey, J.W.: Scatter/gather: a cluster-based approach to browsing large document collections. In: Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '92), pp. 318–329. ACM Press, New York (1992)
6. Chisholm, E., Kolda, T.G.: New term weighting formulas for the vector space method in information retrieval. Technical report (1999)
7. Singhal, Amit, Buckley, Chris, Mitra, Mandar, Mitra, Ar: Pivoted document length normalization, pp. 21–29. ACM Press, New York (1996)

8. Liu, X., Croft, W.B.: Cluster-based retrieval using language models. In: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'04), pp. 186–193. ACM, New York (2004)
9. Voorhees, E.M.: The cluster hypothesis revisited. In: Proceedings of the 8th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '85), pp. 188–196. ACM, New York (1985)
10. Salton, Gerard: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley, Boston (1989)
11. Cummins, Ronan, O'Riordan, Colm: Evolving general term weighting schemes for information retrieval: tests on larger collections. *Artif. Intell. Rev.* **24**, 277–299 (2005)
12. Jung, Y., Park, H., Du, D.Z.: An effective term weighting scheme for information retrieval (2000)
13. Manning, C.D., Raghavan, P., Schütze, H.: Introduction to information retrieval, Chapter 16, p. 496. Cambridge University Press, Cambridge (2008)
14. David MacKay, J.C.: Information Theory Inference and Learning Algorithms. Cambridge University Press, Cambridge (2002)
15. Salton, Gerard, Buckley, Christopher: Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.* **24**(5), 513–523 (1988)
16. Text REtrieval Conference (TREC) (1999)
17. Lewis, D.D.: The reuters-21578 text categorization test collection (1999)
18. Zhou, X., Zhang, X., Hu, X.: Dragon toolkit: incorporating auto-learned semantic knowledge into large-scale text retrieval and mining. In: 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), pp. 197–201. IEEE (2007)
19. Hettich, S., Bay, S.D.: Reuters transcribed subset (1999)
20. Porter, M.F.: An algorithm for suffix stripping. *Program* **14**(3), 130–137 (1980)
21. Murugesan, K.: Cluster-based term weighting and document ranking models. kentucky (2011)

Performance of Poled Organic Polymeric Films by ANOVA in Terms of SHG Conversion Efficiency

Renu Tyagi, Yuvraj Singh Negi and Millie Pant

Abstract In this work we have studied the influence of applied DC electric field on the SHG conversion efficiency of poled organic polymeric films. Regression analysis is carried out to get the desired relationship between the input and output parameters. Statistical testing of the models are performed with F-test to obtain the mathematical relationship between input and output parameters. To examine the goodness of fit of a model, the test for significance of regression model is performed and ANOVA is applied to the response data. Curve fitting analysis has been done to perform the accuracy of observed data.

Keywords Curve fitting · Statistical analysis · ANOVA · Polymeric film · SHG

1 Introduction

Organic conjugated polymers are increasingly studied for application in nonlinear optics [1–3]. Materials possessing a strong $\chi^{(2)}$ nonlinearity are particularly prized for second-order nonlinear applications such as electro-optic modulation and active doped polymer components are currently on the verge of being commercial products. These polymers have been considered interesting materials for applications in different fields such as electronics, photonics, corrosion protection, etc. [4–6]. They present numerous advantages: light-weight, high chemical inertness, high corrosion resistance, high performance low cost, possibility of transitions

R. Tyagi (✉) · Y. S. Negi
Department of Polymer Science and Process Engineering, Saharanpur Campus, IIT Roorkee,
Roorkee, India
e-mail: renutyagi80@gmail.com

M. Pant
Department of Applied Science and Engineering, Saharanpur Campus, IIT Roorkee,
Roorkee, India

from metallic to insulating or semiconducting states and vice versa. In particular, semiconducting polymers can be used in the fabrication of field effect transistors, solar cells, lasers [7–9] etc. Polymeric materials with large values of second and third order optical susceptibilities are potentially applicable to optical data storage, optical computing, imaging and dynamic holography applications [10, 11]. In addition their potential applications are extended also in the field of telecommunications. In order to produce second-order effects and ordered arrangement of molecules in films, they need to be polarized by using such as electric-field poling and all-optical poling [12]. In recent years, second harmonic generation (SHG) intensity has been induced in NLO polymer materials by electric field poling process [13].

On the other hand, PMMA is one of the most commonly used non-linear organic substrate materials, easy to prepare and compatible with the chromophore [12]. Under normal circumstances, it has no interaction with the chromophore molecules. The work done in this paper are twofold: (1) Experiments are carried out on mNA doped polymeric films and the influence of applied DC electric field on the SHG conversion efficiency of doped films is shown, (2) the effect of mNA doping concentration on the SHG intensity of PMMA films is studied with the help of regression analysis. In the next subsection we give some preliminary concepts about regression analysis.

1.1 Regression Analysis-Concept

Regression analysis is carried out to get the desired relationship between the input and output parameters. Statistical testing of the models are performed with F-test to obtain the mathematical relationship between input and output parameters. To examine the goodness of fit of a model, the test for significance of regression model is performed and ANOVA is applied to the response data. The following points are taken into consideration for determining the best fit of a model.

- Significance analysis is performed based on the P-values. The terms with P value ≤ 0.05 are significant.
- P-value for the model should also be significant, i.e., $p \leq 0.05$.
- P-value for *lack of fit* should be insignificant. The insignificant *lack of fit* (>0.05) is good for data fitness to the model.
- The model maximizing the correlation coefficients, “R-Square” values such as “R-square”, “adjusted R-square” and “predicted R-square”, is the best model.
- The “predicted R-square” value must be in reasonable agreement with the values of “adjusted R-square” and “R-square”.
- A difference of <0.2 between “R-square” and “predicted R-square” is an indication of a suitable model.

- The predictive power of the model, as evidenced by “predicted R-square” value, must be greater than 0.5. Larger the “predicted R-square” value, better the model.
- Moreover, a classical analysis of variance (ANOVA) using F-tests allows us to analyze the total response variation by identifying the parts corresponding to the sources of variation (regression model, pure experimental error) and to analyze the residuals in order to point out the possible *lack of fit* of the postulated model when replicates are available.

2 Experimental and Characterization Details

Materials Used

Commercial grade Polymethyl methacrylate and optical grade meta nitro aniline manufactured by Merck, Germany is used to prepare the films. Dimethyl Sulfoxide (Merck, India) is used as a solvent.

Film Sample Preparation

Tough, transparent and flexible films of PMMA doped with appropriate amount of m-NA were prepared by solution cast method in particular solvent. The films were dried in vacuum oven for 4–5 h at 100 °C to remove all residual solvent. Meta nitro aniline is incorporated into PMMA as dopant material from 2 to 18 wt.% of PMMA with step size of 2. It is observed that optical grade mNA can incorporate up to 18 wt.% of PMMA in polymer matrix without showing any visible aggregate of m-NA. After drying colored transparent films were obtained and were subjected to optical characterization. Doped films are denoted by PM-X where X indicates wt.% of mNA ($X = 2$ to 18 wt.%) in polymer. The compositions to make the films are given in the Table 1.

Poling Procedure

To align the NLO active m-NA molecule in the amorphous matrix a contact electrode poling of doped films was carried out. For this doped polymer films were heated from room temperature to 120 °C. To analyze the effect of applied poling voltage on the SHG intensity, the film samples were poled with an applied dc electric field (EF) of 4 and 6 kV/cm for 30 min at a given poling temperature and poled films are denoted by EF-4 and EF-6 respectively. The samples are then cooled to room temperature and the poling field is subsequently removed which results a system where the dipole moment of NLO molecules i.e. m-NA are aligned in the electric field direction within the polymer matrix.

Characterization of Doped Films

For second harmonic generation (SHG) signal, poled doped polymer film samples are illuminated with a Q-switched Nd: YAG laser (wavelength 1,064 nm, pulse width 40 ns and 1 pps) beam. The generated SHG signal at 532 nm carries information about the material's properties is separated from the fundamental wave by using filters. SHG signal is measured relative to an input signal and SHG

Table 1 ANOVA analysis of EF-4 film samples

Film sample	Observed R square	Adj R square	Predicted R square	F value	P value	Significance
PM-2	0.9905	0.9275	1	116.05	2.73E-13	Significant
PM-4	0.9901	0.9318	1	123.94	1.44E-13	Significant
PM-6	0.987	0.9383	1	137.85	4.49E-14	Significant
PM-8	0.9871	0.9367	1	134.35	3.69E-14	Significant
PM-10	0.9872	0.9380	1	137.24	1.9E-14	Significant
PM-12	0.9893	0.9418	1	146.52	9.29E-15	Significant
PM-14	0.9881	0.9389	1	139.45	8.09E-15	Significant
PM-16	0.9856	0.9387	1	139.03	5.18E-15	Significant
PM-18	0.9857	0.9398	1	141.70	4.05E-15	Significant

conversion efficiency is calculated as the ratio of output signal to input signal multiplied by 100. SHG signal conversion efficiency is analyzed as a function of input laser energy and concentration of mRNA in polymer matrix.

3 Results and Discussion

In this paper SHG conversion efficiency is measured as the ratio of output SHG power to input power multiplied by 100 and is represented in terms of percentage (%). The obtained results are analyzed by ANOVA to establish relative significance.

ANOVA results for SHG conversion efficiency of film samples poled at 4KV/cm electric field (EF-4) and 6KV/cm electric field (EF-6) are shown in Tables 1 and 2 respectively. In these tables F- value is the variance ratio and p-value is the probability. In ANOVA the F-value and p-value determines whether an effect is significant or insignificant. The significant effect is observed for all parameters. The correlation coefficients R-square, adjusted R-square and predicted R-square are in reasonable agreement. Non linear models for film samples EF-4 and EF-6 are tabulated in Tables 3 and 4.

The SHG conversion efficiency versus laser input power for EF-4 film samples are shown in Fig. 1a and for EF-6 film samples are shown in Fig. 1b. Graph shows that SHG conversion efficiency increases as the density of mRNA molecules increases in polymer matrix. Higher electric field strength produces higher SHG conversion efficiency (Fig. 1b).

Curve fitting plots for EF-4 film samples are represented in Fig. 2 and for EF-6 film samples are represented in Fig. 3. It has been seen that observed SHG conversion efficiency are nearer to predicted SHG conversion efficiency.

Table 2 ANOVA analysis of EF-6 film samples

Film sample	Observed R square	Adj R square	Predicted R square	F value	P value	Significance
PM-2	0.9919	0.9656	1	253.35	6.26E-15	Significant
PM-4	0.9998	0.9690	1	282.89	2.37E-15	Significant
PM-6	0.9933	0.9634	1	239.99	2.78E-15	Significant
PM-8	0.9903	0.9565	1	198.96	4.5E-15	Significant
PM-10	0.9876	0.9398	1	141.58	1.22E-14	Significant
PM-12	0.9863	0.9356	1	131.83	1.39E-14	Significant
PM-14	0.9875	0.9361	1	133.02	9.26E-15	Significant
PM-16	0.9866	0.9402	1	142.43	5.4E-15	Significant
PM-18	0.9867	0.9431	1	150.24	3.2E-15	Significant

Table 3 Non linear models for EF-4 film samples

Film sample	Predicted model	Actual model
PM-2	$-7E-15x^2 + 1.0069x + 42.674$	$-1.073x^2 + 2.0866x + 40.936$
PM-4	$7E-15x^2 + 1.0038x + 44.574$	$-0.1026x^2 + 2.0364x + 42.911$
PM-6	$0.9626x + 46.913$	$-0.0891x^2 + 1.8593x + 45.469$
PM-8	$4E-15x^2 + 0.9667 + 48.902$	$-0.091x^2 + 1.8829x + 47.427$
PM-10	$1E-14x^2 + 0.9368x + 50.935$	$-0.0872x^2 + 1.8139x + 49.523$
PM-12	$0.9215x + 53.044$	$-0.0843x^2 + 1.7703x + 51.677$
PM-14	$4E-15x^2 + 0.9183x + 55.116$	$-0.0854x^2 + 1.777x + 53.732$
PM-16	$0.899x + 57.151$	$-0.0813x^2 + 1.7166x + 55.835$
PM-18	$0.9079x + 58.961$	$-0.0812x^2 + 1.7248x + 57.645$

Table 4 Non linear models for EF-6 film samples

Film sample	Predicted model	Actual model
PM-2	$4E-15x^2 + 1.0143x + 46.645$	$-0.068x^2 + 1.6989x + 45.543$
PM-4	$-4E-15x^2 + 0.9927x + 48.78$	$-0.0631x^2 + 1.6273x + 47.758$
PM-6	$-4E-15x^2 + 0.9737x + 50.91$	$-0.0695x^2 + 1.6731x + 49.783$
PM-8	$-4E-15x^2 + 0.9755x + 52.761$	$-0.0744x^2 + 1.7245x + 51.555$
PM-10	$0.9668x + 54.738$	$-0.0855x^2 + 1.8576x + 53.303$
PM-12	$-7E-15x^2 + .9831x + 56.726$	$-0.0929x^2 + 1.9179x + 55.221$
PM-14	$4E-15x^2 + 0.9717x + 58.725$	$-0.0926x^2 + 1.9031x + 57.225$
PM-16	$7E-15x^2 + .9732x + 60.806$	$-0.0878x^2 + 1.8563x + 59.834$
PM-18	$0.9679x + 62.869$	$-0.0843x^2 + 1.8166x + 61.502$

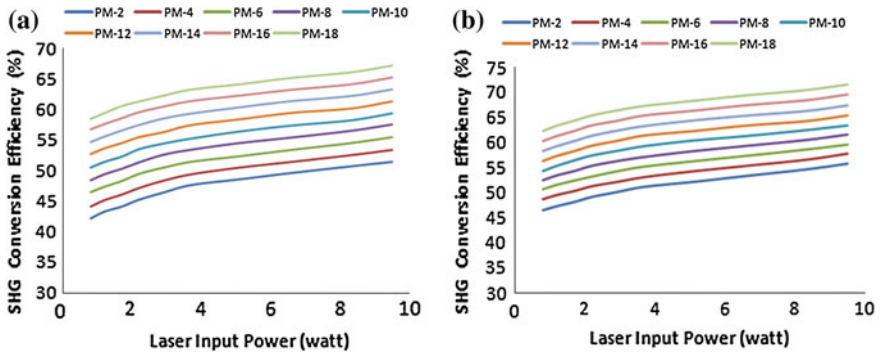


Fig. 1 Relation of SHG conversion efficiency with laser input power for EF-4 film samples (a) and for EF-6 film samples (b)

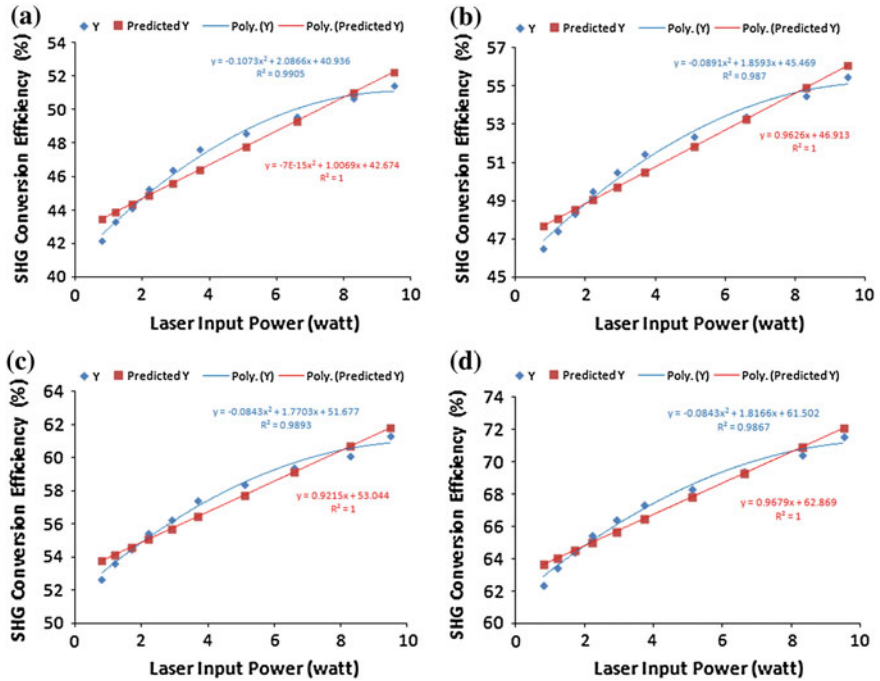


Fig. 2 Curve fitting plots for EF-4 film samples. (a) PM-2. (b) PM-6. (c) PM-12. (d) PM-18

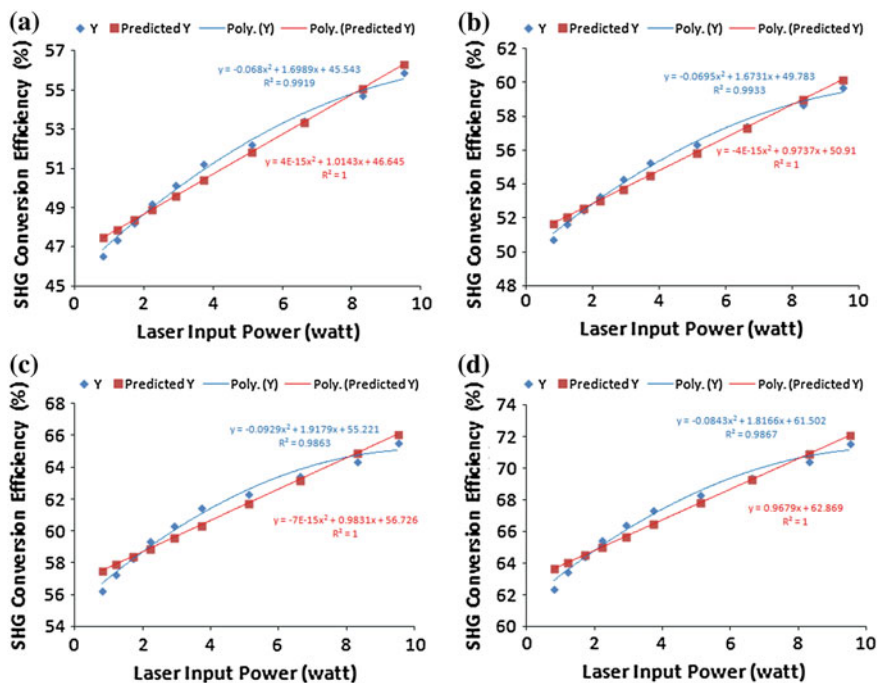


Fig. 3 Curve fitting plots for EF-6 film samples. (a) PM-2. (b) PM-6. (c) PM-12. (d) PM-18

4 Conclusion

In present paper we have studied the statistical analysis of guest host systems based on m-NA as guest and PMMA as host. The regression analysis shows the excellent R square value which indicates a significant correlation between input and output parameters. The possible lack of fit is insignificant. Curve fitting lines are closer to observed lines. It is observed that prepared optical grade films results in more than 40 % SHG conversion efficiency. Further rising of poling field strength increases this conversion efficiency. In brief we can conclude that PMMA/m-NA system show excellent NLO property as compared to other systems.

References

1. Raua, I., Atysa, P.A., Cholleta, P.-A., Kajzar, F., Zamboni, R.: Conjugated polymers oriented organic thin films for nonlinear optics. *Mol. Cryst. Liq. Cryst.* **446**, 23–45 (2006)
2. Kajzar, F., Messier, J.: Cubic non linear optical effects in conjugated polymers. *Poly. J.* **19**, 275–284 (1987)
3. Sunitha, M.S., Adhikari, M.S., Vishnumurthy, K.A.: Non linear optical studies of a D-A type conjugated polymer containing heterocyclic moieties. *Int. J. Appl. Phys. Math.* **2**, 436–438 (2012)

4. Jazbinsek, M., Mutter, L., Gunter, P.: Photonic applications with the organic non linear optical crystal DAST. *IEEE J. Sel. Top. Quantum Electron.* **14**, 1297–1311 (2008)
5. Salaneck, W.R., Friend, R.H., Bredas, J.L.: Electronic structure of conjugated polymers: Consequences of electron: lattice coupling. *Phys. Rep.* **319**, 231–251 (1999)
6. AlSalhi, M.S., Alam, J., Dass, L.A., Raja, M.: Recent advances in conjugated polymers for light emitting devices. *Int. J. Mol. Sci.* **12**, 2036–2054 (2011)
7. Sasaki, K.: Organic waveguide devices for photonic use. *J. Kor. Phys. Soc.* **31**, 439–442 (1997)
8. Yoshino, K., Ohmori, Y., Fujii, A., Ozaki, M.: Organic electronic devices based on polymeric material and tunable photonic crystal. *Japanese J. Appl. Phys.* **46**, 5655–5673 (2007)
9. Hide, F., Diaz-Garcia, M.A., Schwartz, B.J., Heeger, A.J.: New developments in the photonic applications of conjugated polymers. *Acc. Chem. Res.* **30**, 430–436 (1997)
10. Sudhahar, S., Kumar, M.K., Silambarasan, A., Muralidharan, R., Kumar, R.M.: Studies on structural, spectral, and optical properties of organic nonlinear optical single crystal: 2-Amino-4,6-dimethylpyrimidinium p-Hydroxybenzoate. *J. Mater.* **2013**, 1–7 (2013)
11. Nadtoka, O.: Nonlinear optical effects in polymeric azoesters. *Chem. Chem. Tech.* **4**, 185–190 (2010)
12. Virkki, M., Kauranen, M., Priimagi, A.: Different chromophore concentration dependence of photoinduced birefringence and second-order susceptibility in all optical poling. *Appl. Phys. Lett.* **99**, 183309 (1–3) (2011)
13. Ostroverkhova, O., Stickrath, A., Singer, K.D.: Electric field-induced second harmonic generation studies of chromophore orientational dynamics in photorefractive polymers. *J. Appl. Phys.* **91**, 9481–9486 (2002)

Performance Analysis of Bio-Inspired Techniques

Samiksha Goel, Arpita Sharma and V. K. Panchal

Abstract Increasing popularity of nature inspired meta-heuristics and novel additions in the pool of these techniques at a rapid pace results in a need to categorize and explore these meta-heuristics from different point of views. This paper attempts to compare three broad categories of bio-inspired techniques-Swarm Intelligence methods, Evolutionary techniques and Ecology based approaches via three renowned algorithms, Ant Colony Optimization (ACO), Genetic Algorithm (GA) and Biogeography Based Optimization (BBO) that fall under three categories respectively, based on few varied characteristics. Six benchmark functions are considered for comparison. The paper further suggests a taxonomy of nature inspired methods based on the source of inspiration.

Keywords: Nature inspired meta-heuristics · Bio-inspired methods · ACO · BBO · GA

1 Introduction

Nature is incessantly reckoned as a source of inspiration for human community in diverse ambit. Numbers of optimization algorithms have been developed stirred by the optimization approaches espoused by the nature [1]. These methods are often

S. Goel (✉) · A. Sharma
Department of Computer Science, Delhi University, Delhi, India
e-mail: samiksha.goel@gmail.com

A. Sharma
e-mail: asharma@ddu.du.ac.in

V. K. Panchal
DTRL, DRDO, Delhi, India
e-mail: vkpans@ieee.com

entitled as nature inspired meta-heuristics impersonating certain triumphant distinctiveness of the nature. In spite of the verity that the preeminent strategy is yet to be discovered, or may not be present at all, this suite of techniques have transpired remarkably in the last two decades [2]. They are often considered as astonishingly proficient in solving difficult and challenging real-world optimization problems, which may even sometimes required to deal with NP hard problems. They have the knack to portray and resolve intricate relationships from essentially extremely simple initial conditions and rules with modest or no knowledge of the search space. The span of this area is certainly enormous leading to an innovative era in next generation computing, modeling and algorithm engineering. An immense literature of nature inspired approaches exists to decipher imposing range of problems [3].

A group of algorithms of this vast research area is inspired by biological mechanisms occurring in the nature and its components and are fall under the umbrella of bio-inspired meta-heuristics. Bio inspired algorithms are going to be a new-fangled transformation in computer science. The key feature is the equilibrium between solution diversity and solution speed. Ideally, the aim is to find the global best solution with the minimum computing effort. This study is an attempt to compare the broad sub-categories of bio-inspired intelligence methods via three well-known and well-established algorithms, Ant Colony Optimization (ACO) [4], Biogeography Based Optimization (BBO) [5] and Genetic Algorithm (GA) [6] based on assorted aspects. Study is supported by exploiting six benchmark functions available for global optimization. In addition to this a better taxonomy is suggested for the wide range of nature inspired intelligence algorithms based on their source of inspiration.

The rest of the paper is organized as follows. [Section 2](#) highlights the taxonomy of the nature inspired techniques focusing more towards bio-inspired approaches. The next section presents the brief overview of the three approaches, ACO, BBO and GA used in the study. The comparison of these algorithms based on various aspects will be carried out in [Sect. 4](#). [Section 5](#) illustrates the experiments performed using the six benchmark functions. Analysis of the results obtained, are also presented in the same section. The last section is dedicated to the conclusion with some suggestions for the future work.

2 Taxonomy of Nature Inspired Meta-heuristics

The vast literature of Nature inspired techniques (NIT) successfully marked the presence of more than 40 innovative algorithms under this category [2]. Numerous researchers have tried to classify them based on assorted facets. It is truly an exigent errand to categorize these algorithms analytically. The most recent taxonomy recommended by Yang [3] considered the higher level grouping of these techniques based on broad source of inspiration areas of biology, physics or chemistry. He has also conveyed that from the set theory point of view, SI-based

algorithms are a subset of bio-inspired algorithms, which itself is a subset of nature-inspired algorithms. This relationship can be easily represented as [2],

$$\begin{aligned} \text{Swarm Intelligence Based Algorithms} &\subset \text{Bio – Inspired Algorithms} \\ &\subset \text{Nature Inspired Algorithms} \end{aligned}$$

He has further specified that some algorithms do not fall under bio inspired category but still they are developed by mimicking certain physical and/or chemical laws. Thus they are specified under a sub category of NIT satisfying the following relation [2].

$$\text{Physics \& Chemistry Based Algorithms} \begin{cases} \not\subset \text{Bio – inspired Algorithm} \\ \subset \text{Nature Inspired Algorithm} \end{cases}$$

Inspired by the above statements, we have proposed a taxonomy based on the source of inspiration as illustrated in Fig. 1. Initially the broad categories of bio-inspired methods are evolutionary techniques and swarm intelligence mechanisms. However, it has been identified that methods inspired by human mechanisms also spur a large number of approaches, which are nothing but the bio-inspired methods. Recently few approaches have been developed which have proved their competence in this pool but unable to find its place in any of these categories. Hence, in the taxonomy we have proposed a new sub-category associating the methods inspired by adopting some of the ecological mechanisms. A recent addition to this category is Biogeography Based optimization. This paper is an attempt to compare the three broad categories—Swarm Intelligence methods, Evolutionary techniques and Ecology based approaches via three algorithms namely ACO, GA and BBO that fall under these groups respectively. This comparison is an effort in support of the addition of a new subcategory under bio-inspired methods. The Swarm Intelligence methods and Evolutionary techniques have been chosen because the Ecology based techniques are thought of having the competencies of both these categories.

3 Brief Overview of ACO, BBO and GA Algorithms

This section discusses three main algorithms selected for comparison with their generalized pseudo codes.

3.1 Ant Colony Optimization

ACO is one of the most booming swarm based algorithms proposed by Dorigo in 1999 [4]. It is inspired by the foraging behavior of ants and based on the phenomena known as stigmergy. The most fascinating facet of the collaborative behavior of

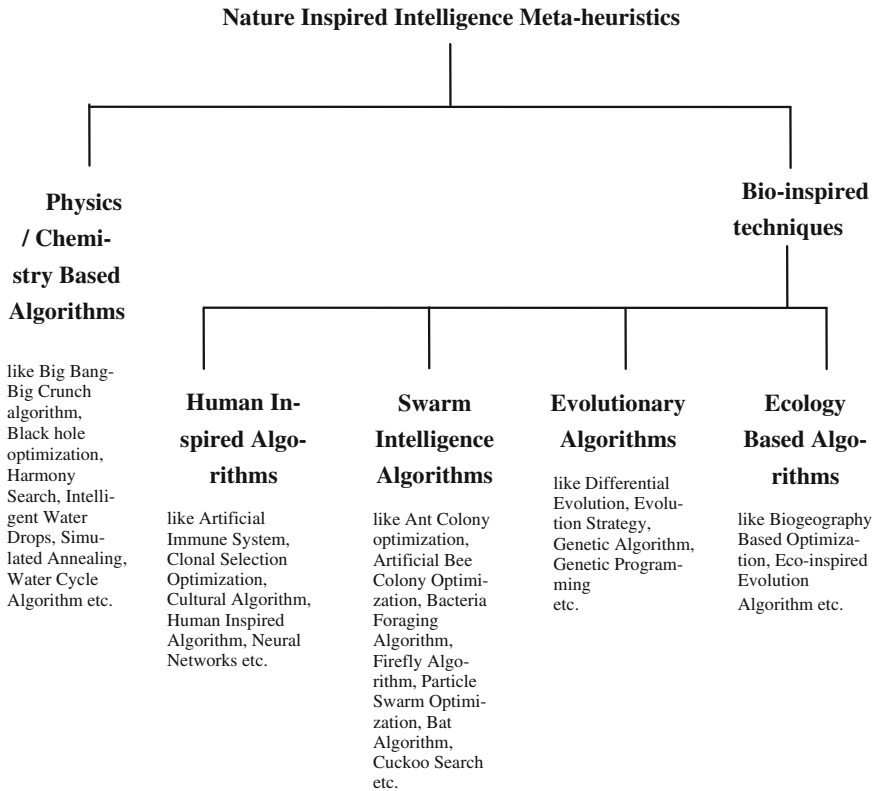


Fig. 1 Taxonomy of nature inspired meta-heuristics based on the source of information

numerous ant species is their knack to locate the shortest paths between the ants’ nest and the food sources by tracing pheromone trails: stronger the pheromone trail, higher its desirability. The pseudo code [4] is illustrated below.

BEGIN

SET parameters, INITIALIZE and LOCATE individuals of the population randomly in search space. Also INITIALIZE pheromone values

REPEAT UNTIL (TERMINATION CONDITION is satisfied)

EVALUATE the fitness of each ant solution

FOR each individual

UPDATE individual’s knowledge, position and pheromone values by using the knowledge of other individuals

END LOOP

END LOOP

END

3.2 Biogeography Based Optimization

BBO is a global optimization algorithm developed by Simon in 2008 [5]. It is inspired by mathematical models of biogeography. Each possible solution corresponds to an island and their features that characterize habitability are termed as suitability index variables (SIV). The fitness of each solution is called its habitat suitability index (HSI) and depends on many features of the habitat. High-HSI solutions tend to share their features with low by emigrating solution features to other habitats. Low-HSI solutions accept a lot of new features from high-HSI solutions by immigration from other habitats. Immigration and emigration tend to improve the solutions and thus evolve a solution to the optimization problem [6]. The pseudo code [5] is discussed below.

```

BEGIN
  FOR each Habitat  $H_i$ 
    FOR each Habitat feature  $s$ 
      Select Habitat  $H_i$  with probability proportional to  $\lambda_i$  (immigration rate)
      If Habitat  $H_i$  is selected
        Select  $H_j$  with probability proportional to  $\mu_i$  (emigration rate)
        If  $H_j$  is selected
           $H_i(s) \leftarrow H_j(s)$ 
    END LOOP
    Next Habitat feature
    Mutate  $H_i$  (Based on mutation probability)
  END LOOP
END

```

3.3 Genetic Algorithms

GA is an evolutionary based stochastic optimization algorithm with a global search potential proposed by Holland [7]. GA is inspired by the evolutionary ideas of natural selection. They follow the principles of Charles Darwin Theory of survival of the fittest. Population of solution (chromosome) is initialized randomly. Based on the fitness function, the best chromosomes are selected into the mating pool, where cross over and mutation are carried out to obtain new set of solutions (offspring). The three principal genetic operators in GA involve selection, cross-over, and mutation. The pseudo code [7] of the algorithm is as follows,

BEGIN

 INITIALIZE population of n chromosomes with random candidate solutions

 EVALUATE each candidate with the fitness function $f(x)$

 REPEAT UNTIL (TERMINATION CONDITION is satisfied)

 SELECT two parent chromosomes based on the fitness values

 CROSSOVER pair of parents to form new offspring using crossover probability

 MUTATE the resulting offspring with mutation probability

 EVALUATE the new candidates using $f(x)$

 SELECT individuals for next generation based on the survival of fittest

 END LOOP

END

4 Comparison of Bio-inspired Techniques

This section highlights the comparative study carried out for the three broad categories of bio-inspired techniques—Swarm Intelligence methods [8, 9], Evolutionary techniques [10] and Ecology based approaches. For this we have selected three well-known and well-established algorithms, Ant Colony Optimization, Genetic Algorithm and Biogeography Based Optimization one from each group and compare them on the basis of various characteristic [11, 12]. Table 1 illustrates the summarized details of the assessment.

5 Experiments and Analysis

The six benchmark functions [13] have been used to compare the three algorithms on various nodes as represented in the Table 2 with their properties.

Each algorithm is run for the above mentioned six benchmark functions for 50 generations considering 20 probable solutions in one generation. The results obtained have been illustrated in Table 3 with their initial and final best and mean minimum. To analyze the flow of convergence, i.e., the search strategy, initial point is considered as same for each of the algorithm.

The Table 3 shows that the BBO converges quite fast towards the solution as compared to the other two algorithms. Also, the trend followed by the BBO is quite similar to GA in some functions. However, the ACO convergence strategy is totally different. The next Table 4 shows the results obtained by 100 monte carlo simulations run for 100 generations. Results clearly state that BBO outperforms the other two methods in finding the solutions. Also, the average time taken by

Table 1 Comparison of ACO, BBO and GA based on various factors

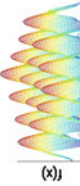
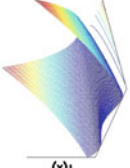
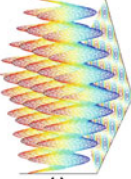
Factors	ACO	BBO	GA
Source of motivation	Inspired by the collective decisions taken by the ants in search of their food source by adopting shortest path to the destination	Motivated by the movement of species from one island to another in search of a more suitable habitat in the biogeography of the nature	Based on the evolutionary ideas of natural selection and genetics. It is based on the Darwinian evolution concept to breed good solution
Design principles	N “unintelligent” agents in the swarm are cooperating by working in a decentralized manner to achieve some purposeful behavior and achieve some goal. Based on the foraging behavior of the ants. The pheromone deposited by one ant acts as a hint for other members of the colony to follow a particular path	The species are emigrated or immigrated from one island to another depending on their fitness values. Each habitat is validated by Habitat Suitability Index (HSI), a fitness function which is dependent on the Suitability Index Variables (SIVs) representing the features of the habitat. The individuals with high HSI is likely to share features with others and less likely to accept features from others	The main design of this approach is based on the following three operators: <ul style="list-style-type: none"> • Selection-mimics the “survival of the fittest” • Recombination-mimics “mating” in the biological population • Mutation-mimics “diversity” in the population
Flow of knowledge	Indirect sharing of knowledge via environment in the form of pheromone values	Direct sharing by emigration and immigration of species between habitats	Knowledge shared directly via mutation and migration of species
Representation of solutions	<ul style="list-style-type: none"> • Position of ants represent the candidate solution • Solutions have the property to club together • Solution changes indirectly based on the pheromone values • Original population is not discarded from one generation to another 	<ul style="list-style-type: none"> • N habitats represents the N candidate solutions • Solutions change directly by sharing their attributes • Original population is not discarded after each generation • Solutions donot have property to group together 	<ul style="list-style-type: none"> • Each chromosome corresponds to a probable solution • Solutions which are having low fitness values are discarded in each generation • Solutions have no inherent property of clustering

(continued)

Table 1 (continued)

Factors	ACO	BBO	GA
Time complexity	Have high computation cost. However, its is problem dependent Experimental section shows the computation cost incurred by each of the three algorithms on the various benchmark functions	Being a novel approach still need to be explored to identify its actual strength	Have strong global optimization ability but with low computation ability
Operators involved	Pheromone Update and Measure, trail evaporation	Migration (emigration and immigration), mutation	Crossover, mutation, selection, Inversion, Gene Silencing
Control parameters	Number of ants, iterations, pheromone evaporation rate, amount of reinforcement	Number of habitats (population size), maximum migration and mutation rate, elitism	Population size, maximum generations, cross over probability, mutation probability, length of chromosome, chromosome encoding
Exploration tactic	Parallel approach is followed as a number of ants initiate the system by searching the destination food source (target solutions) in various directions	Intrinsically parallel approach is followed Multiple species explored the solution space in multiple directions at once to find the optimal solutions	Parallel approach by implementing hill climbing heuristic method, i.e., it will preserve multiple solutions, eradicates unpromising solutions and provides reasonable solutions

Table 2 List of Benchmark functions used for comparison

Benchmark functions	Properties	Range of dimensions	Function and its plot in 2D
Rastrigin	Continuous, differentiable, separable, scalable, multimodal	± 5.12	 $f(x) = 10n + \sum_{i=1}^n [x_i^2 - 10 \cos(2\pi x_i)]$
Rosenbrock	Continuous, differentiable, non-separable, scalable, unimodal	± 2.048	 $f(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2]$
Griewank	Continuous, differentiable, non-separable, scalable, multimodal	± 600	 $f(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$

(continued)

Table 2 (continued)

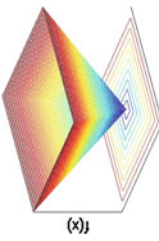
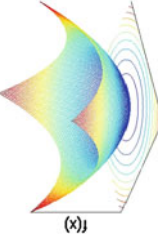
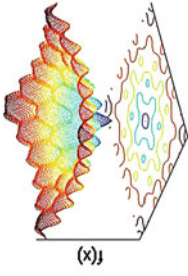
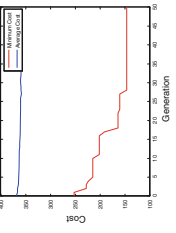
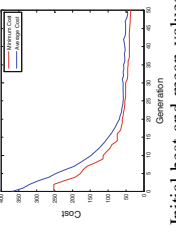
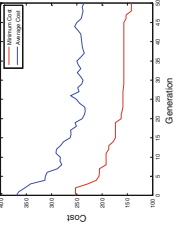
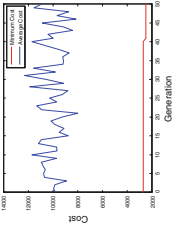
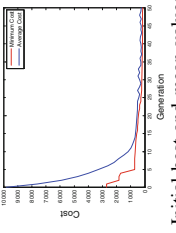
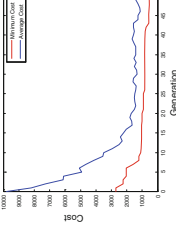
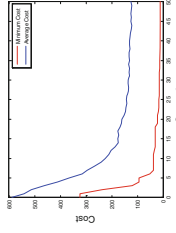
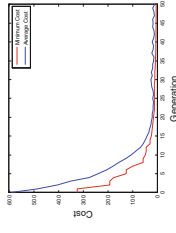
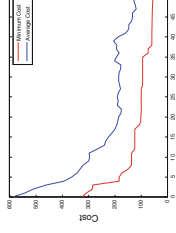
Benchmark functions	Properties	Range of dimensions	Function and its plot in 2D
Schwefel 2.21	Continuous, non-differentiable, separable, scalable, unimodal	± 100	 $f(x) = \sum_{i=1}^n \left(\sum_{j=1}^i x_j \right)^2$
Sphere	Continuous, differentiable, separable, scalable, multimodal	$\pm 5, 12$	 $f(x) = \sum_{i=1}^n x_i^2$
Ackley	Continuous, differentiable, non-separable, scalable, multimodal	± 30	 $f(x) = -a \cdot \exp \left(-b \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} - \exp \left(\frac{1}{n} \sum_{i=1}^n \cos(c x_i) \right) + a + \exp(1) \right)$

Table 3 Illustrating the graphs of each of the algorithm with their initial and final best and mean minimum values

	ACO	BBO	GA
Rastrigin	 <p>Initial best and mean values—252.2727 and 368.4794 Final best and mean values—146.1364 and 359.0697</p>	 <p>Initial best and mean values—252.2727 and 368.4794 Final best and mean values—36.4478 and 46.4925</p>	 <p>Initial best and mean values—252.2727 and 368.4794 Final best and mean values—141.9662 and 235.125</p>
Rosenbrock	 <p>Initial best and mean values—2699.1972 and 9797.4428 Final best and mean values—2479.3969 and 10888.3219</p>	 <p>Initial best and mean values—2699.1972 and 9797.4428 Final best and mean values—144.696 and 242.488</p>	 <p>Initial best and mean values—2699.1972 and 9797.4428 Final best and mean values—530.1439 and 1348.4363</p>
Griewank	 <p>Initial best and mean values—324.2704 and 589.3174 Final best and mean values—10.9194 and 124.613</p>	 <p>Initial best and mean values—324.2704 and 589.3174 Final best and mean values—4.2095 and 9.7893</p>	 <p>Initial best and mean values—324.2704 and 589.3174 Final best and mean values—55.3248 and 119.6696</p>

(continued)

Table 3 (continued)

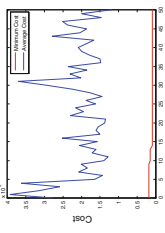
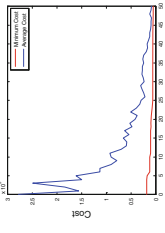
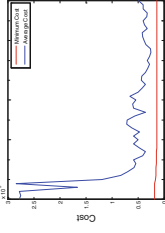
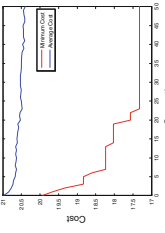
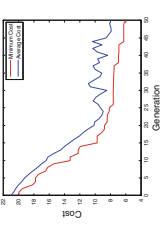
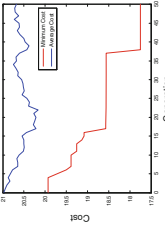
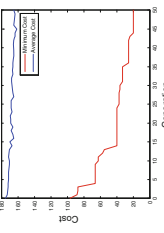
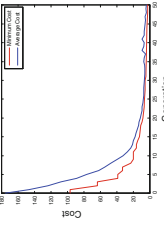
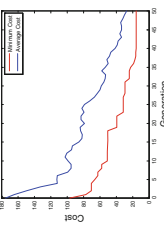
	ACO	BBO	GA
Schwefel 2.21	 <p>Initial best and mean values—18121.7237 and 277492.0243 Final best and mean values—8286.2811 and 120006.7654</p>	 <p>Initial best and mean values—18121.7237 and 277492.0243 Final best and mean values—4677.1871 and 8652.6547</p>	 <p>Initial best and mean values—18121.7237 and 277492.0243 Final best and mean values—13342.975 and 48400.2976</p>
	 <p>Initial best and mean values—19.9281 and 20.9711 Final best and mean values—17.326 and 20.3929</p>	 <p>Initial best and mean values—19.9281 and 20.9711 Final best and mean values—5.9785 and 7.9114</p>	 <p>Initial best and mean values—19.9281 and 20.9711 Final best and mean values—17.7507 and 20.6775</p>
	 <p>Initial best and mean values—96.8048 and 174.8175 Final best and mean values—19.8246 and 165.2299</p>	 <p>Initial best and mean values—96.8048 and 74.8175 Final best and mean values—2.4491 and 3.6849</p>	 <p>Initial best and mean values—96.8048 and 74.8175 Final best and mean values—15.3295 and 26.9306</p>

Table 4 BEST values obtained by 100 monte carlo simulations run for 100 generations

	Rastrigin	Rosenbrock	Griewank	Schwefel 2.21	Ackley	Sphere
ACO	97.83731	808.7979	2.425295	1844.236	8.897557	6.500687
BBO	8.087152	21.06172	1.821845	1367.122	4.069425	0.151179
GA	75.74998	97.1302	4.047823	1057.101	10.50392	3.194911

Table 5 Average CPU time taken by each of the algorithm during the 100 monte carlo simulations

	Rastrigin	Rosenbrock	Griewank	Schwefel 2.21	Ackley	Sphere
ACO	3.6	2.7	16.6	5.1	3.3	4.3
BBO	1.3	1.9	3.9	3.0	1.6	2.7
GA	3.2	2.8	2.8	4.3	2.3	2.2

each of the algorithm during the monte carlo simulations has been specified in the Table 5. It shows that the time taken by BBO algorithm is least followed by GA and ACO.

6 Conclusion

The last two decades has experienced an astonishing growth in the nature inspired meta-heuristics. Its frontiers have been expanding in multiple directions resulting in the development of novel nature inspired computing techniques at rapid pace. This paper presents a comparative study of ACO, BBO and GA-the three bio-inspired mechanisms based on the various characteristics and six well-known benchmark functions. The paper validates the better performance of BBO which combines the features of swarm techniques and evolutionary approach, thereby re-enforcing our belief to keep it under separate category of ecology based algorithms. A more detailed comparison can be carried out by considering different combinations of parameters involved. More techniques can be used for comparison purpose to further support the statement.

References

1. Kari, L., Rozenberg, G.: The many facets of natural computing. *Commun. ACM* **51**(10), 72–83 (2008)
2. Fister Jr, I., et al.: A brief review of nature-inspired algorithms for optimization. *arXiv preprint arXiv: 1307.4186* (2013)
3. Yang, X.S.: *Nature-Inspired Metaheuristic Algorithms*. Luniver Press (2010)
4. Dorigo, M., Di Caro, G.: Ant colony optimization: a new meta-heuristic. In: *Proceedings of the congress on Evolutionary Computation*, vol. 2. IEEE (1999)

5. Simon, D.: Biogeography-based optimization. *IEEE Trans. Evol. Comput.* **12**(6), 702–713 (2008)
6. Simon, D., Ergezer, M., Du, D.: Population distributions in biogeography-based optimization algorithms with elitism. In: *IEEE Conference on Systems, Man, and Cybernetics*, pp. 1017–1022, Oct 2009
7. Holland, J.H.: Genetic algorithms and the optimal Allocation of trials. *SIAM J. Comput.* **2**(2), 88–105 (1973)
8. Engelbrecht, A.P.: *Fundamentals of Computational Swarm Intelligence*. Wiley, Hoboken (2006)
9. Bonabeau, E., Dorigo, M., Theraulaz, G.: *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, London (1999)
10. Back, T., Hammel, U., Schwefel, H.: Evolutionary computation: comments on the history and current state. *IEEE Trans. Evol. Comput.* **1**, 3–17 (1997)
11. Vesterstrom, J., Thomsen, R.: A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In *IEEE Congress on Evolutionary Computation*, vol. 2, pp. 1980–1987 (2004)
12. Goel, S., Sharma, A., Panchal, V.K.: Biogeography based optimization, an evolutionary technique or a swarm based approach? In: *5th International Multiconference on Intelligent Systems, Sustainable New and Renewable Energy Technology & Nanotechnology (IISN-2011)*, pp 220–225
13. Chattopadhyay, R.: A study of test functions for optimization algorithms. *J. Opt. Theory Appl.* **8**, 231–236 (1971)

On a Class of Nondifferentiable Multiobjective Continuous Programming Problems

Iqbal Husain and Vikas Kumar Jain

Abstract Fritz John and Karush-Kuhn-Tucker type optimality conditions for a nondifferentiable multiobjective variational problem with equality and inequality constraints are obtained. Using Karush-Kuhn-Tucker type optimality conditions, a Wolfe type second-order nondifferentiable multiobjective dual variational problem is constructed. Various duality results for the pair of Wolfe type second-order dual variational problems are proved under second-order pseudoinvexity. A pair of Wolfe type dual variational problems having equality and inequality constraints with natural boundary values is also formulated to prove various duality results. Finally, the linkage between our results and those of their static counterparts existing in the literature is briefly outlined.

Keywords Multiobjective variational problems · Wolfe type second-order duality · Second-order generalized invexity · Multiobjective nonlinear programming problems

1 Introduction

Second-order duality in mathematical programming has been widely researched. As second-order dual to a constrained optimization problem gives a tighter bound and hence enjoys computational advantage over the first-order dual to the problem. Mangasarian [1] was the first to second-order duality in non-linear programming.

I. Husain (✉) · V. K. Jain
Department of Mathematics, Jaypee University of Engineering and Technology, Guna,
Madhya Pradesh, India
e-mail: ihusain11@yahoo.com

V. K. Jain
e-mail: jainvikas13@yahoo.com

Motivated with analysis of Mangasarian [1], Chen [2] presented Wolfe type second-order dual to a class of constrained variational problems under an involved invexity like conditions. Later Husain et al. [3] introduced second-order invexity and generalized invexity and presented a Mond-Weir type second-order dual to the problem of [2] in order to relax implicit invexity requirements to the generalized second order invexity.

Multiobjective optimization has applications in various fields of science that includes engineering, economics and logistics where optimal decisions need to be taken in the presence of trade-off between two or more conflicting objectives. Motivated with facts that multiobjective optimization models represent a variety of real life problems, Husain and Jain [4] formulated a multiobjective version of the variational problem considered by Chen [2] with equality and inequality constraints which represent more realistic problems than those variational problems with an inequality constraint only. In [4], they studied optimality and duality for their variational problem. In this paper, optimality conditions for a multiobjective variational problem which contains a term of square root of certain quadratic form in each component of the vector valued objective function of the problem, are obtained and Wolfe type second-order duality is investigated for this problem. Finally, our results are shown to be the dynamic generalization of those of non-differentiable multiobjective mathematical programming problems, already studied in the literature.

2 Definitions and Related Pre-requisites

Let $I = [a, b]$ be a real interval, A and $\phi : I \times R^n \times R^n \rightarrow R^m$ be twice continuously differentiable functions. In order to consider $\phi(t, x(t), \dot{x}(t))$, where $x : I \rightarrow R^n$ is differentiable with derivative \dot{x} , denoted by ϕ_x and $\phi_{\dot{x}}$, the first order derivatives of ϕ with respect to $x(t)$ and $\dot{x}(t)$, respectively, that is,

$$\phi_x = \left(\frac{\partial \phi}{\partial x^1}, \frac{\partial \phi}{\partial x^2}, \dots, \frac{\partial \phi}{\partial x^n} \right)^T, \quad \phi_{\dot{x}} = \left(\frac{\partial \phi}{\partial \dot{x}^1}, \frac{\partial \phi}{\partial \dot{x}^2}, \dots, \frac{\partial \phi}{\partial \dot{x}^n} \right)^T$$

Further Denote by ϕ_{xx} and ψ_x the $n \times n$ Hessian matrix and $m \times n$ Jacobian matrix respectively. The symbols $\phi_{\dot{x}}, \phi_{\dot{x}\dot{x}}, \phi_{x\dot{x}}$ and $\psi_{\dot{x}}$ have analogous representations.

Designate by X , the space of piecewise smooth functions $x : I \rightarrow R^n$, with the norm $\|x\| = \|x\|_\infty + \|Dx\|_\infty$, where the differentiation operator D is given by

$$u = Dx \Leftrightarrow x(t) = \int_a^t u(s)ds,$$

Thus $\frac{d}{dt} = D$ except at discontinuities.

We incorporate the following definitions which are required for the derivation of the duality results.

Definition 1 (Second-order Invex): If there exists a vector function $\eta = \eta(t, x, \bar{x}) \in R$ where $\eta : I \times R^n \times R^n \rightarrow R^n$ and with $\eta = 0$ at $t = a$ and $t = b$ such that for a scalar function $\phi(t, x, \dot{x})$, the functional $\int_I \phi(t, x, \dot{x}) dt$ where $\phi : I \times R^n \times R^n \rightarrow R$ satisfies

$$\begin{aligned} & \int_I \phi(t, x, \dot{x}) dt - \int_I \left\{ \phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2} \beta(t)^T G \beta(t) \right\} dt \\ & \geq \int_I \left\{ \eta^T \phi_x(t, \bar{x}, \dot{\bar{x}}) + (D\eta)^T \phi_{\dot{x}}(t, \bar{x}, \dot{\bar{x}}) + \eta^T G \beta(t) \right\} dt \end{aligned}$$

then $\int_I \phi(t, x, \dot{x}) dt$ is second-order invex with respect to η , where $G = \phi_{xx} - 2D\phi_{x\dot{x}} + D^2\phi_{\dot{x}\dot{x}} - D^3\phi_{\ddot{x}\ddot{x}}$ and $\beta \in C(I, R^n)$ the space of n-dimensional continuous vector functions.

Definition 2 (Second-order Pseudoinvex): If the functional $\int_I \phi(t, x, \dot{x}) dt$ satisfies

$$\begin{aligned} & \int_I \left\{ \eta^T \phi_x + (D\eta)^T \phi_{\dot{x}} + \eta^T G \beta(t) \right\} dt \geq 0 \\ & \Rightarrow \int_I \phi(t, x, \dot{x}) dt \geq \int_I \left\{ \phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2} \beta^T(t) G \beta(t) \right\} dt \end{aligned}$$

then $\int_I \phi(t, x, \dot{x}) dt$ is said to be second-order pseudoinvex with respect to η .

Definition 3 (Second-order strict-pseudoinvex): If the functional $\int_I \phi(t, x, \dot{x}) dt$ satisfies

$$\begin{aligned} & \int_I \left\{ \eta^T \phi_x + (D\eta)^T \phi_{\dot{x}} + \eta^T G \beta(t) \right\} dt \geq 0 \\ & \Rightarrow \int_I \phi(t, x, \dot{x}) dt > \int_I \left\{ \phi(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2} \beta(t)^T G \beta(t) \right\} dt \end{aligned}$$

then $\int_I \phi(t, x, \dot{x}) dt$ is said to be second-order pseudoinvex with respect to η .

The following inequality will also be required in the forthcoming analysis of the research:

It states that

$$x(t)^T B(t) z(t) \leq (x(t)^T B(t) x(t))^{1/2} (z(t)^T B(t) z(t))^{1/2}$$

with equality in above if $B(t)x(t) - q(t)z(t) = 0$, for some $q(t) \in R, t \in I$.

Throughout the analysis of this research, the following conventions for the inequalities will be used:

If $d, s \in R^n$ with $d = (d^1, d^2, \dots, d^n)$ and $s = (s^1, s^2, \dots, s^n)$, then

$$\begin{aligned}
 d \geq s &\Leftrightarrow d^i \geq s^i, \quad (i = 1, 2, \dots, n) \\
 d \geq s &\Leftrightarrow d \geq s \quad \text{and} \quad d \neq s \\
 d > s &\Leftrightarrow d^i > s^i, \quad (i = 1, 2, \dots, n)
 \end{aligned}$$

3 Optimality Conditions

We consider the following nondifferentiable multiobjective variational problems containing terms of square root functions.

(NWVEP): Minimize

$$\left(\int_I \left\{ f^1(t, x(t), \dot{x}(t)) + (x(t)^T B^1(t)x(t))^{1/2} \right\} dt, \dots, \int_I \left\{ f^p(t, x(t), \dot{x}(t)) + (x(t)^T B^p(t)x(t))^{1/2} \right\} dt \right)$$

subject to

$$x(a) = 0, \quad x(b) = 0 \tag{1}$$

$$g(t, x(t), \dot{x}(t)) \leq 0, \quad t \in I \tag{2}$$

$$h(t, x(t), \dot{x}(t)) = 0, \quad t \in I \tag{3}$$

where

- (i) $f^i : I \times R^n \times R^n \rightarrow R, i \in K = \{1, 2, \dots, p\}, g : I \times R^n \times R^n \rightarrow R^m$ and $h : I \times R^n \times R^n \rightarrow R^k$ are assumed to be continuously differentiable functions, and
- (ii) for each $t \in I, i \in K = \{1, 2, \dots, p\}, B^i(t)$ is an $n \times n$ positive semi definite (symmetric) matrix, with $B^i(\cdot)$ continuous on i .

If $B^i(t) = 0$ for all i and t , then the above problem reduces to the problem of [4].

In order to obtain optimality condition for the problem (NWVEP) we consider the following nondifferentiable single objective continuous programming problem considered by Chandra et al. [5]:

$$\begin{aligned}
 \text{(CP):} \quad &\text{Minimize} \quad \int_I \phi(t, x(t), \dot{x}(t)) dt \\
 &\text{Subject to} \quad x(a) = 0, \quad x(b) = 0 \\
 &\quad \quad \quad g(t, x(t), \dot{x}(t)) \leq 0, \quad h(t, x(t), \dot{x}(t)) = 0, \quad t \in I
 \end{aligned}$$

where $\phi : I \times R^n \times R^n \rightarrow R$ is a continuous differentiable functions, and g and h as the same as given for (NWVEP).

Chandra et al. [5] derive the following Fritz John type optimality conditions:

Lemma 1 (Fritz John type optimality conditions): If \bar{x} is an optimality solution of the problem (NWVEP), and if $h_x(\cdot, \bar{x}(t), \dot{\bar{x}}(t))$ maps X onto a closed subspace of $C(I, R^k)$, then there exist Lagrange multipliers $\tau \in R$, piecewise smooth $\bar{y} : I \rightarrow R^m$ and $\bar{z} : I \rightarrow R^k$ such that

$$\begin{aligned} & \tau f_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{y}(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{z}(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t)) \\ & - D(\tau f_{\dot{x}}(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{y}(t)^T g_{\dot{x}}(t, \bar{x}(t), \dot{\bar{x}}(t)) \\ & + \bar{z}(t)^T h_{\dot{x}}(t, \bar{x}(t), \dot{\bar{x}}(t))) = 0, \quad t \in I \\ & \bar{x}(t)^T B(t)w(t) = (\bar{x}(t)^T B(t)\bar{x}(t))^{1/2}, \quad t \in I \\ & w(t)^T B(t)w(t) \leq 1, \quad t \in I \\ & y(t)^T g(t, \bar{x}(t), \dot{\bar{x}}(t)) = 0, \quad t \in I \\ & (\tau, y(t)) \geq 0, (\tau, y(t), z(t)) \neq 0, \quad t \in I \end{aligned}$$

The above Fritz-John type necessary optimality conditions if $\tau = 1$ (then \bar{x} may called normal). For $\tau = 1$, it suffice to assume that the Robinson conditions [5] holds or Slater’s condition [5] holds, i.e. for some $v \in X$ and all $t \in I$,

$$t \in Ig(t, \bar{x}(t), \dot{\bar{x}}(t)) + g_x(t, \bar{x}(t), \dot{\bar{x}}(t))v(t) + g_{\dot{x}}(t, \bar{x}(t), \dot{\bar{x}}(t))\dot{v}(t) > 0.$$

The following lemma relates an efficient of (NWVEP) with an optimal solution of p -single objective variational problems.

Lemma 2 (Chankong and Haimes): A point $\bar{x}(t) \in X$ is an efficient solution of (NWVEP) if and only $\bar{x}(t)$ is an optimal solution (NWVEP_r) for each $r \in K = \{1, 2, \dots, p\}$

$$\begin{aligned} \text{(NWVEP}_r\text{):} \quad & \text{Minimize} \quad \int_I \left(f^r(t, x(t), \dot{x}(t)) + (x(t)^T B^r(t)x(t))^{1/2} \right) dt \\ & \text{subject to} \\ & x(a) = 0, \quad x(b) = 0 \\ & g(t, x(t), \dot{x}(t)) \leq 0, h(t, x(t), \dot{x}(t)) = 0, \quad t \in I \\ & \int_I \left(f^i(t, x(t), \dot{x}(t)) + (x(t)^T B^i(t)x(t))^{1/2} \right) dt \\ & \leq \int_I \left(f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2} \right) dt, \\ & i \in K_r = K - \{r\} \end{aligned}$$

Since the variational problem (NWVEP) does not contain integral inequality, it can easily be shown that Lemma 2 still remains valid for the constraint without integral sign. That is,

$$f^i(t, x(t), \dot{x}(t)) + (x(t)^T B^i(t)x(t))^{1/2} \leq f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2},$$

$$i \in K_r$$

Theorem 1 (Fritz John type optimality conditions): Let $\bar{x}(t)$ be an efficient solution of (NWVEP) and if $h_x(\cdot, \bar{x}(t), \dot{\bar{x}}(t))$ map X onto a closed subspace of $C(I, R^k)$. Then there exist $\bar{\lambda}^i \in R$ and piecewise smooth functions $\bar{y} : I \rightarrow R^m$, $\bar{z} : I \rightarrow R^k$ and $w^i : I \rightarrow R^m, i \in K$ such that

$$\sum_{i=1}^p \lambda^i (f_x^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + B^i(t)w^i(t) - Df_{\dot{x}}^i(t, \bar{x}(t), \dot{\bar{x}}(t)))$$

$$+ y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t))$$

$$- D(y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t))) = 0, \quad t \in I$$

$$\bar{x}(t)^T B^i(t)w^i(t) = (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2}, (w^i(t)^T B^i(t)w^i(t))^{1/2} \leq 1, \quad i \in K, t \in I$$

$$(\bar{\lambda}^1, \bar{\lambda}^2, \dots, \bar{\lambda}^p, y(t)) \geq 0, (\bar{\lambda}^1, \bar{\lambda}^2, \dots, \bar{\lambda}^p, y(t), z(t)) \neq 0, \quad t \in I$$

Proof: Since $\bar{x}(t)$ is an efficient solution of (NWVEP), by Lemma 1, $\bar{x}(t)$ is an optimal solution of (NWVEP_r) for each $r \in K$ and hence of (NWVEP₁). Hence by Lemma 2, there exist $\bar{\lambda}^i \in R, i \in K$ and piecewise smooth functions $\bar{y} : I \rightarrow R^m$, $\bar{z} : I \rightarrow R^k$ such that

$$\sum_{i=1}^p \lambda^i (f_x^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + B^i(t)w^i(t) - Df_{\dot{x}}^i(t, \bar{x}(t), \dot{\bar{x}}(t)))$$

$$+ y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t))$$

$$- D(y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t))) = 0, \quad t \in I$$

$$\bar{x}(t)^T B^i(t)w^i(t) = (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2}, (w^i(t)^T B^i(t)w^i(t))^{1/2} \leq 1, i \in K, \quad t \in I$$

$$(\bar{\lambda}^1, \bar{\lambda}^2, \dots, \bar{\lambda}^p, y(t)) \geq 0, \quad t \in I$$

which give the required optimality conditions.

Theorem 2 (Karush-Kuhn-Tucker necessary optimality conditions): Assume that

- (i) $\bar{x}(t)$ be an efficient solution of (NWVEP) and
- (ii) for each $r \in K$, the constraints of (NWVEP_r) satisfy Slater's [5] or Robinsons [5] condition at $\bar{x}(t)$.

Then there exist $\bar{\lambda} \in R^p$ and piecewise smooth functions $\bar{y} : I \rightarrow R^m, \bar{z} : I \rightarrow R^k$ and $w^i : I \rightarrow R^m, i \in K$ such that

$$\begin{aligned} & \sum_{i=1}^p \lambda^i (f_x^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + B^i(t)w^i(t) - Df_{\bar{x}}^i(t, \bar{x}(t), \dot{\bar{x}}(t))) \\ & + y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t)) \\ & - D(y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, \bar{x}(t), \dot{\bar{x}}(t))) = 0, \quad t \in I \\ y(t)^T g(t, \bar{x}(t), \dot{\bar{x}}(t)) = 0, y(t) \geq 0, t \in I, \lambda > 0, \quad \sum_{i=1}^p \lambda^i = 1. \\ \bar{x}(t)^T B^i(t)w^i(t) = (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2}, (w^i(t)^T B^i(t)w^i(t))^{1/2} \leq 1, \quad i \in K \end{aligned}$$

Proof: Since $\bar{x}(t)$ is an efficient solution of (NWVEP), by Lemma 1 $\bar{x}(t)$ is an optimal solution of (NWVEP_r), by the Karush-Kuhn-Tucker conditions given earlier, for each $r \in K$, there exist $\bar{v}_r^i \in R, r \in K_r$ and piecewise smooth functions $\mu_r^j \in R, j = 1, 2, \dots, m, \delta_r^l \in R, l = 1, 2, \dots, k$ such that

$$\begin{aligned} & f_x^r(t, \bar{x}, \dot{\bar{x}}) + B^i(t)w^i(t) - Df_x^r(t, \bar{x}, \dot{\bar{x}}) \\ & + \sum_{i \in K_r} \bar{v}_r^i (f_x^i(t, \bar{x}, \dot{\bar{x}}) + B^i(t)w^i(t) - Df_x^i(t, \bar{x}, \dot{\bar{x}})) \\ & \sum_{j=1}^m (\mu_r^j g_x^j(t, \bar{x}, \dot{\bar{x}}) - D(\mu_r^j g_x^j(t, \bar{x}, \dot{\bar{x}}))) \\ & + \sum_{l=1}^k (\delta_r^l h_x^l(t, \bar{x}, \dot{\bar{x}}) - D(\delta_r^l h_x^l(t, \bar{x}, \dot{\bar{x}}))) = 0 \\ & \sum_{j=1}^m \mu_r^j(t) g^j(t, \bar{x}, \dot{\bar{x}}) = 0, \bar{v}_r^i \geq 0, i \in K_r, y(t) \geq 0, \quad t \in I \end{aligned}$$

Summing over $i \in K$, we get

$$\begin{aligned} & \sum_{i=1}^p (\bar{v}_1^i + \bar{v}_2^i + \dots + \bar{v}_p^i) (f_x^i(t, \bar{x}, \dot{\bar{x}}) + B^i(t)w^i(t) - Df_x^i(t, \bar{x}, \dot{\bar{x}})) \\ & + \sum_{j=1}^m \left\{ (\mu_1^j(t) + \mu_2^j(t) + \dots + \mu_p^j(t)) g_x^j(t, \bar{x}, \dot{\bar{x}}) \right. \\ & \left. - D(\mu_1^j(t) + \mu_2^j(t) + \dots + \mu_p^j(t)) g_x^j(t, \bar{x}, \dot{\bar{x}}) \right\} \\ & + \sum_{l=1}^k \left\{ (\delta_1^l(t) + \delta_2^l(t) + \dots + \delta_p^l(t)) h_x^l(t, \bar{x}, \dot{\bar{x}}) \right. \\ & \left. - D(\delta_1^l(t) + \delta_2^l(t) + \dots + \delta_p^l(t)) h_x^l(t, \bar{x}, \dot{\bar{x}}) \right\} = 0 \\ & \sum_{j=1}^m (\mu_1^j(t) + \mu_2^j(t) + \dots + \mu_p^j(t)) g^j(t, \bar{x}, \dot{\bar{x}}) = 0, \quad t \in I \end{aligned}$$

where $\bar{v}_r^i = 1$ for each $i \in K$.

Equivalently,

$$\begin{aligned} & \sum_{i=1}^k \bar{v}^i (f_x^i(t, \bar{x}, \dot{\bar{x}}) + B^i(t)w^i(t) - Df_x^i(t, \bar{x}, \dot{\bar{x}})) \\ & + \sum_{j=1}^m \{(\mu^j g_x^j(t, \bar{x}, \dot{\bar{x}}) - D(\mu^j g_x^j(t, \bar{x}, \dot{\bar{x}})))\} \\ & + \sum_{l=1}^k (\delta^l h_x^l(t, \bar{x}, \dot{\bar{x}}) - D(\delta^l h_x^l(t, \bar{x}, \dot{\bar{x}}))) = 0 \end{aligned}$$

where $\bar{v}^i = 1 + \sum_{r \in K_r} v_r^i > 0, i \in K, \mu^j(t) = \sum_{r=1}^k \mu_r^j(t) \geq 0, t \in I, j = 1, 2, \dots, m$ and

$$\begin{aligned} \delta^l(t) &= \sum_{r=1}^k \delta_r^l(t) \geq 0, t \in I, \quad l = 1, 2, \dots, k. \\ \Rightarrow \lambda^i &= \frac{\bar{v}^i}{\sum_{r=1}^p \bar{v}^i}, \quad i \in K, \end{aligned}$$

$$y^j(t) = \frac{\mu^j(t)}{\sum_{r=1}^p \bar{v}^i}, \quad j = 1, 2, \dots, m,$$

$$z^l(t) = \frac{\delta^l(t)}{\sum_{r=1}^p \bar{v}^i}, \quad l = 1, 2, \dots, k.$$

We get

$$\begin{aligned} & \sum_{i=1}^p \lambda^i (f_x^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + B^i(t)w^i(t) - Df_x^i(t, \bar{x}(t), \dot{\bar{x}}(t))) \\ & + \sum_{j=1}^m (y^j(t)g_x^j(t, \bar{x}(t), \dot{\bar{x}}(t)) - D(y^j(t)g_x^j(t, \bar{x}(t), \dot{\bar{x}}(t)))) \\ & + \sum_{l=1}^k (z^l(t)h_x^l(t, \bar{x}(t), \dot{\bar{x}}(t)) - D(z^l(t)h_x^l(t, \bar{x}(t), \dot{\bar{x}}(t)))) = 0 \\ & \sum_{j=1}^m y^j(t)g^j(t, \bar{x}, \dot{\bar{x}}) = 0, \quad t \in I \end{aligned}$$

or

$$\begin{aligned} & \sum \lambda^i (f_x^i(t, \bar{x}, \dot{\bar{x}}) + B^i(t)w^i(t) - Df_x^i(t, \bar{x}, \dot{\bar{x}})) \\ & + (y(t)^T g_x(t, \bar{x}, \dot{\bar{x}}) + z(t)^T h_x(t, \bar{x}, \dot{\bar{x}}) \\ & - D(y(t)^T g_x(t, \bar{x}, \dot{\bar{x}}) + z(t)^T h_x(t, \bar{x}, \dot{\bar{x}})) = 0, \quad t \in I \\ & y(t)^T g(t, \bar{x}, \dot{\bar{x}}) = 0, y(t) = (y^1(t), y^2(t), \dots, y^m(t)) \geq 0, \quad t \in I \\ & \lambda = (\lambda^1, \lambda^2, \dots, \lambda^k) > 0, \quad \sum_{i=1}^k \lambda^i = 1 \end{aligned}$$

4 Wolfe Type Second-Order Duality

We construct the following problem as the dual to the problem (P):

(D): Maximize

$$\begin{aligned} & \left(\int_I (f^1(t, u(t), \dot{u}(t)) + u(t)^T B^1(t)w^1(t) + y(t)^T g(t, u(t), \dot{u}(t)) \right. \\ & \quad \left. + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^1 \beta(t)) dt, \dots, \right. \\ & \left. \int_I (f^p(t, u(t), \dot{u}(t)) + u(t)^T B^p(t)w^p(t) + y(t)^T g(t, u(t), \dot{u}(t)) \right. \\ & \quad \left. + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^p \beta(t)) dt \right) \end{aligned}$$

subject to

$$u(a) = 0 = u(b) \tag{4}$$

$$\begin{aligned} & \sum_{i=1}^p \lambda^i (f_u^i(t, u(t), \dot{u}(t)) + B^i(t)w^i(t) - Df_u^i(t, u(t), \dot{u}(t))) \\ & + y(t)^T g_u(t, u(t), \dot{u}(t))n + z(t)^T h_u(t, u(t), \dot{u}(t)) \\ & - D(y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t))) \\ & + H\beta(t) = 0, \quad t \in I \end{aligned} \tag{5}$$

$$w^i(t)^T B^i(t)w^i(t) \leq 1, \quad i \in K \tag{6}$$

$$y(t) \geq 0, \quad t \in I \tag{7}$$

$$\lambda > 0, \quad \sum_{i=1}^p \lambda^i = 1 \tag{8}$$

where

$$\begin{aligned}
 H^i &= f_{xx}^i - 2Df_{x\dot{x}}^i + D^2f_{\dot{x}\dot{x}}^i - D^3f_{\ddot{x}\ddot{x}}^i + (y(t)^T g_x + z(t)^T h_x)_x - 2D(y(t)^T g_x + z(t)^T h_x)_{\dot{x}} \\
 &\quad + D^2(y(t)^T g_{\dot{x}} + z(t)^T h_{\dot{x}})_{\dot{x}} - D^3(y(t)^T g_{\ddot{x}} + z(t)^T h_{\ddot{x}})_{\ddot{x}} \\
 H &= \lambda^T f_{xx} - 2D\lambda^T f_{x\dot{x}} + D^2\lambda^T f_{\dot{x}\dot{x}} - D^3\lambda^T f_{\ddot{x}\ddot{x}} + (y(t)^T g_x + z(t)^T h_x)_x \\
 &\quad - 2D(y(t)^T g_x + z(t)^T h_x)_{\dot{x}} + D^2(y(t)^T g_{\dot{x}} + z(t)^T h_{\dot{x}})_{\dot{x}} - D^3(y(t)^T g_{\ddot{x}} + z(t)^T h_{\ddot{x}})_{\ddot{x}}.
 \end{aligned}$$

Theorem 3 (Weak Duality): Let $x(t) \in C_p$ and $(u(t), \lambda, y(t), z(t), w^1(t), \dots, w^p(t)) \in C_D$ such that for $\int_I (f^i(t, \dots) + (\cdot)^T B^i(t)w^i(t) + y(t)^T g(t, \dots) + z(t)^T h(t, \dots))dt, i \in K$, is second-order pseudoinvex for all $w^i(t) \in R^n, i \in K$ with respect to η . Then

$$\begin{aligned}
 &\int_I \left(f^i(t, x(t), \dot{x}(t)) + (x(t)^T B^i(t)x(t))^{1/2} \right) dt \\
 &< \int_I \left(f^i(t, u(t), \dot{u}(t)) + u(t)^T B^i(t)w^i(t) + y(t)^T g(t, u(t), \dot{u}(t)) \right. \\
 &\quad \left. + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2}\beta(t)^T H^i\beta(t) \right) dt, i \in K = \{1, 2, \dots, p\}
 \end{aligned}$$

and

$$\begin{aligned}
 &\int_I \left(f^j(t, x(t), \dot{x}(t)) + (x(t)^T B^j(t)x(t))^{1/2} \right) dt \\
 &\leq \int_I \left(f^j(t, u(t), \dot{u}(t)) + u(t)^T B^j(t)w^j(t) + y(t)^T g(t, u(t), \dot{u}(t)) \right. \\
 &\quad \left. + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2}\beta(t)^T H^j\beta(t) \right) dt, \text{ for all } j \in K, j \neq i.
 \end{aligned}$$

cannot hold.

Proof: Suppose to the contrary that there is $\bar{x}(t)$ feasible for (VP) and $(u(t), \lambda, y(t), z(t), w^1(t), \dots, w^p(t))$ feasible for (DV) such that

$$\int_I \left(f^i(t, x(t), \dot{x}(t)) + (x(t)^T B^i(t)x(t))^{1/2} \right) dt < \int_I \left(f^i(t, u(t), \dot{u}(t)) + u(t)^T B^i(t)w^i(t) + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^i \beta(t) \right) dt, \text{ for some } i \in \{1, 2, \dots, p\}$$

and

$$\int_I \left(f^j(t, x(t), \dot{x}(t)) + (x(t)^T B^j(t)x(t))^{1/2} \right) dt \leq \int_I \left(f^j(t, u(t), \dot{u}(t)) + u(t)^T B^j(t)w^j(t) + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^j \beta(t) \right) dt, \text{ for all } j \in K, j \neq i.$$

Then, using $x(t)^T B^i(t)w^i(t) \leq (x(t)^T B^i(t)x(t))^{1/2} (w^i(t)^T B^i(t)w^i(t))^{1/2}$, $i \in K$

We have

$$\int_I \left(f^i(t, x(t), \dot{x}(t)) + x(t)^T B^i(t)w^i(t) + y(t)^T g(t, x(t), \dot{x}(t)) + z(t)^T h(t, x(t), \dot{x}(t)) \right) dt < \int_I \left(f^i(t, u(t), \dot{u}(t)) + u(t)^T B^i(t)w^i(t) + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^i \beta(t) \right) dt, \text{ } i \in K$$

$$\int_I \left(f^j(t, x(t), \dot{x}(t)) + x(t)^T B^j(t)w^j(t) + y(t)^T g(t, x(t), \dot{x}(t)) + z(t)^T h(t, x(t), \dot{x}(t)) \right) dt \leq \int_I \left(f^j(t, u(t), \dot{u}(t)) + u(t)^T B^j(t)w^j(t) + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) - \frac{1}{2} \beta(t)^T H^j \beta(t) \right) dt, \text{ for all } j \text{ with } j \neq i.$$

Since $\int_I (f^i(t, \dots) + (\cdot)^T B^i(t)w^i(t) + y(t)^T g(t, \dots) + z(t)^T h(t, \dots)) dt$, is second-order pseudoinvex for all $i \in K$, then

$$\int_I \left[\eta^T (f_u^i + B^i(t)w^i(t) + y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t))) + \left(\frac{d\eta}{dt} \right)^T (f_u^i + y(t)^T g_u + z(t)^T h_u) + \eta^T H^i \beta(t) \right] < 0, \quad \text{for all } i \in K,$$

and

$$\int_I \left[\eta^T (f_u^j + B^j(t)w^j(t) + y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t))) + \left(\frac{d\eta}{dt} \right)^T (f_u^j + y(t)^T g_u + z(t)^T h_u) + \eta^T H^j \beta(t) \right] \leq 0, \quad j \neq i.$$

Thus

$$\begin{aligned} 0 &> \int_I \left[\eta^T \left(\sum_{i=1}^p \lambda^i (f_u^i + B^i(t)w^i(t)) + y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t)) \right) + \left(\frac{d\eta}{dt} \right)^T (\lambda^T f_u + y(t)^T g_u + z(t)^T h_u) + \eta^T H \beta(t) \right] dt \\ &= \int_I \left[\eta^T \left(\sum_{i=1}^p \lambda^i (f_u^i + B^i(t)w^i(t)) + y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t)) \right) \right] dt \\ &\quad - D(\lambda^T f_u + y(t)^T g_u + z(t)^T h_u) + H \beta(t) dt + \eta^T (\lambda^T f_u + y(t)^T g_u + z(t)^T h_u) \Big|_{t=a}^{t=b} \end{aligned}$$

Using $\eta = 0$, at $t = a$ and $t = b$, we obtain,

$$\begin{aligned} \Rightarrow \int_I \eta^T \left[\sum_{i=1}^p \lambda^i (f_u^i + B^i(t)w^i(t)) + y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t)) - D(\lambda^T f_u + y(t)^T g_u + z(t)^T h_u) + H \beta(t) \right] dt < 0 \end{aligned}$$

contradicting the constraint of (NWVED). Thus the validity of the conclusion of the theorem follows.

Theorem 4 (Strong Duality): Let $\bar{x}(t)$ be efficient and normal solution for (NWVEP), then there exist $\lambda \in R^k$ and piecewise smooth $y : I \rightarrow R^m, z : I \rightarrow R^l$, and $w^i : I \rightarrow R^n, i \in K$ such that $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta = 0)$ is feasible for (WVED) and the two objective functionals are equal. Furthermore, if the hypothesis of theorem hold, then $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta)$ is efficient for the problem (NWVED).

Proof: By Theorem 2, there exist $\lambda = (\lambda^1, \lambda^2, \dots, \lambda^p) \in R^p$ and piecewise smooth $y : I \rightarrow R^m, z : I \rightarrow R^l$ and $w^i : I \rightarrow R^l, i = 1, 2, \dots, p$ such that

$$\begin{aligned} & \sum_{i=1}^p \lambda^i (f_x^i(t, \bar{x}, \dot{\bar{x}}) + \bar{x}(t)B^i(t)w^i(t) - Df_{\bar{x}}^i(t, \bar{x}, \dot{\bar{x}})) \\ & + y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, x(t), \dot{x}(t)) \\ & - D(y(t)^T g_x(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h_x(t, x(t), \dot{x}(t))) = 0, \quad t \in I \\ & y(t)^T g(t, \bar{x}(t), \dot{\bar{x}}(t)) = 0, y(t) \geq 0, t \in I, \lambda > 0, \quad \sum_{i=1}^p \lambda^i = 1. \\ & \bar{x}(t)^T B^i(t)w^i(t) = (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2}, w^i(t)^T B^i(t)w^i(t) \leq 1, \quad t \in I, i \in K \end{aligned}$$

Thus $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta = 0)$ is feasible for (NWVED) and for all $i \in K$.

$$\begin{aligned} & \int_I (f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2}) dt \\ & < \int_I (f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{x}(t)^T B^i(t)\bar{w}^i(t) \\ & + y(t)^T g(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h(t, \bar{x}(t), \dot{\bar{x}}(t))) dt, \quad t \in I, i \in K \end{aligned}$$

This implies that, the objective functional values are equal.

If $(\bar{x}, \lambda, y, z, w^1, w^2, \dots, w^p, \beta = 0)$ is not efficient solution of (NWVED), then there exists feasible $(u^*, \lambda^*, y^*, z^*, w^1, \dots, w^p, \beta^*)$ for (NWVED) such that

$$\begin{aligned} & \int_I (f^i(t, u^*, \dot{u}^*) + u^*(t)^T B^i(t)w^i(t) + y^*(t)^T g(t, u^*, \dot{u}^*) \\ & + z^*(t)^T h(t, u^*, \dot{u}^*) - \frac{1}{2} \beta^*(t)^T H^i \beta^*(t)) dt \\ & > \int_I (f^i(t, \bar{x}, \dot{\bar{x}}) + \bar{x}(t)^T B^i(t)w^i(t) + y^*(t)^T g(t, \bar{x}, \dot{\bar{x}}) \\ & + z^*(t)^T h(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2} \beta^*(t)^T H^i \beta^*(t)) dt, \quad i \in K. \end{aligned}$$

and

$$\begin{aligned} & \int_I (f^j(t, u^*, \dot{u}^*) + u^*(t)^T B^j(t) w^i(t) + y^*(t)^T g(t, u^*, \dot{u}^*) \\ & \quad + z^*(t)^T h(t, u^*, \dot{u}^*) - \frac{1}{2} \beta^*(t)^T H^{*j} \beta^*(t)) dt \\ & \geq \int_I (f^j(t, \bar{x}, \dot{\bar{x}}) + \bar{x}(t)^T B^j(t) w^j(t) + y^*(t)^T g(t, \bar{x}, \dot{\bar{x}}) \\ & \quad + z^*(t)^T h(t, \bar{x}, \dot{\bar{x}}) - \frac{1}{2} \beta^*(t)^T H^j \beta^*(t)) dt, \quad j \neq i \end{aligned}$$

Since $\int_I (f^i(t, \dots) + (\cdot)^T B^i(t) w^i(t) + y(t)^T g(t, \dots) + z(t)^T h(t, \dots)) dt$, is second-order pseudoinvex with respect to η , as earlier

$$\begin{aligned} & \int_I \eta^T [(f_u^i(t, u^*, \dot{u}^*) + B^i(t) w^{*i} - Df_u^i(t, u^*, \dot{u}^*)) \\ & \quad + y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h_u(t, u^*, \dot{u}^*) \\ & \quad - D(y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h_u(t, u^*, \dot{u}^*) + H^* \beta(t))] dt < 0, \quad i \in K \\ & \int_I \eta^T [(f_u^j(t, u^*, \dot{u}^*) + B^j(t) w^{*i} - Df_u^j(t, u^*, \dot{u}^*)) \\ & \quad + y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h_u(t, u^*, \dot{u}^*) \\ & \quad - D(y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h_u(t, u^*, \dot{u}^*) + H^* \beta(t))] dt \leq 0, \quad j \neq i. \end{aligned}$$

Thus $\int_I \eta^T \left[\sum_{i=1}^p \lambda^i (f_u^i(t, u^*, \dot{u}^*) + B^i(t) w^{*i} - Df_u^i(t, u^*, \dot{u}^*)) + y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h(t, u^*, \dot{u}^*) - D(y^*(t)^T g_u(t, u^*, \dot{u}^*) + z^*(t)^T h_u(t, u^*, \dot{u}^*) + H^* \beta(t)) \right] dt < 0$ contradicting the feasibility of $(u^*, \lambda^*, y^*, z^*, w^{*1}, w^{*2}, \dots, w^{*p}, \beta^*)$ for (NWVED). Thus $(\bar{x}, \lambda, y, z, w^1, w^2, \dots, w^p, \beta = 0)$ is efficient for the dual (NWVED).

Below is the Mangasarian [1] type Strict-Convex duality theorem:

Theorem 5 (Strict-Convex duality): Let $\bar{x}(t)$ and $(\bar{u}(t), \lambda, y(t), z(t), w^1(t), \dots, w^p(t))$ be efficient solutions for the problems (NWVEP) and (NWVED) such that

$$\begin{aligned}
 & \sum_{i=1}^p \lambda^i \int_I (f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{x}(t)^T B^i(t) w^i(t)) dt \\
 &= \int_I \left(\sum_{i=1}^p \lambda^i (f^i(t, \bar{u}(t), \dot{\bar{u}}(t)) + \bar{u}(t)^T B^i(t) w^i(t)) \right. \\
 & \quad \left. + y(t)^T g(t, \bar{u}(t), \dot{\bar{u}}(t)) + z(t)^T h(t, \bar{u}(t), \dot{\bar{u}}(t)) - \frac{1}{2} \beta(t)^T H \beta(t) \right) dt
 \end{aligned} \tag{9}$$

If $\int_I \left(\sum_{i=1}^p \lambda^i (f^i(t, \dots) + (\cdot)^T B^i(t) w^i(t)) + y(t)^T g(t, \dots) + z(t)^T h(t, \dots) \right) dt$ is second-order strictly pseudoinvex with respect to η , then $\bar{x}(t) = \bar{u}(t), t \in I$.

Proof: Suppose $\bar{x}(t) \neq \bar{u}(t)$, for $t \in I$. By second-order strict pseudoinvexity with respect to η , (9) yields,

$$\begin{aligned}
 0 &> \int_I \left\{ \eta^T \left(\sum_{i=1}^p \lambda^i (f_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) + B^i(t) w^i(t) - Df_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) \right. \right. \\
 & \quad \left. \left. + y(t)^T g_u(t, \bar{u}(t), \dot{\bar{u}}(t)) + z(t)^T h_u(t, \bar{u}(t), \dot{\bar{u}}(t))) \right. \right. \\
 & \quad \left. \left. + \left(\frac{d\eta}{dt} \right)^T (\lambda f_{\bar{u}}^i + y(t)^T g_{\bar{u}} + z(t)^T h_{\bar{u}}(t, u, \dot{u})) + \eta^T H \beta(t) \right) dt \right. \\
 &= \int_I \eta^T \left\{ \sum_{i=1}^p \lambda^i (f_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) + B^i(t) w^i(t) - Df_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) \right. \\
 & \quad \left. + y(t)^T g_u(t, \bar{u}(t), \dot{\bar{u}}(t)) + z(t)^T h_u(t, \bar{u}(t), \dot{\bar{u}}(t))) \right. \\
 & \quad \left. - D(f_{\bar{u}}^i + y(t)^T g_{\bar{u}} + z(t)^T h_{\bar{u}}(t, u, \dot{u})) + H \beta(t) \right\} dt \\
 & \quad \left. + \eta^T (\lambda f_{\bar{u}}^i + y(t)^T g_{\bar{u}} + z(t)^T h_{\bar{u}}(t, u, \dot{u})) \Big|_{t=a}^{t=b} \right.
 \end{aligned}$$

Using $\eta = 0$, at $t = a$ and $t = b$, we have

$$\begin{aligned}
 & \int_I \eta^T \left\{ \sum_{i=1}^p \lambda^i (f_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) + B^i(t) w^i(t) - Df_{\bar{u}}^i(t, \bar{u}(t), \dot{\bar{u}}(t)) \right. \\
 & \quad \left. + y(t)^T g_u(t, \bar{u}(t), \dot{\bar{u}}(t)) + z(t)^T h_u(t, \bar{u}(t), \dot{\bar{u}}(t))) \right. \\
 & \quad \left. - D(f_{\bar{u}}^i + y(t)^T g_{\bar{u}} + z(t)^T h_{\bar{u}}(t, u, \dot{u})) + H \beta(t) \right\} dt < 0.
 \end{aligned}$$

contradicts the equality constraint of the dual variational problem (NWVED). Hence

$$\bar{x}(t) = \bar{u}(t), t \in I.$$

The following is the Huard [6] type converse duality:

Theorem 6 (Converse duality): Let $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta(t))$ be an efficient solution of (NWVED) for which

$$\begin{aligned} (C_1) \quad & \text{H is non - singular} \\ (C_2) \quad & \left[\begin{array}{l} (\sigma(t)^T H^i \sigma(t))_{\bar{x}} - D(\sigma(t)^T H^i \sigma(t))_{\bar{x}} + D^2(\sigma(t)^T H^i \sigma(t))_{\bar{x}} \\ - D^3(\sigma(t)^T H^i \sigma(t))_{\bar{x}} + D^4(\sigma(t)^T H^i \sigma(t))_{\bar{x}} \dots \end{array} \right] \\ & + 2 \left[\begin{array}{l} \sigma(t)^T (H\sigma(t))_{\bar{x}} - \sigma(t)^T D(H\sigma(t))_{\bar{x}} + \sigma(t)^T D^2(H\sigma(t))_{\bar{x}} \\ - \sigma(t)^T D^3(H\sigma(t))_{\bar{x}} + \sigma(t)^T D^4(H\sigma(t))_{\bar{x}} \dots \end{array} \right] = 0 \Rightarrow \sigma(t) = 0, \end{aligned}$$

where $\sigma(t)$ is a vector function.

Then $\bar{x}(t)$ is feasible for (NWVEP) and the two objectives functional have the same value. Also, if the weak duality theorem holds for all feasible of (NWVEP) and (NWVED), then $\bar{x}(t)$ is efficient.

Proof: Since $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta(t))$ is an efficient solution of (NWVED), there exists $\alpha, \xi \in R^p, \delta \in R$ and piecewise smooth $\theta : I \rightarrow R^n, \eta : I \rightarrow R^m, q^i : I \rightarrow R, i \in K$ such that following Fritz John conditions (Theorem 1) are satisfied at $(\bar{x}, \lambda, y, z, w^1, \dots, w^p, \beta(t))$:

$$\begin{aligned} & \sum_{i=1}^p \alpha^i [(f_x^i + B^i(t)w^i(t) + y^T g_x + z^T h_x) - D(f_x^i + y^T g_x + z^T h_x) \\ & - \frac{1}{2}(\beta(t)^T H^i \beta(t))_{\bar{x}} + \frac{1}{2}D(\beta(t)^T H^i \beta(t))_{\bar{x}} - \frac{1}{2}D^2(\beta(t)^T H^i \beta(t))_{\bar{x}} \\ & + \frac{1}{2}D^3(\beta(t)^T H^i \beta(t))_{\bar{x}} - \frac{1}{2}D^4(\beta(t)^T H^i \beta(t))_{\bar{x}}] \\ & + \theta(t)^T [H + (H\beta)_{\bar{x}} - D(H\beta)_{\bar{x}} + D^2(H\beta)_{\bar{x}} - D^3(H\beta)_{\bar{x}} + D^4(H\beta)_{\bar{x}}] = 0 \end{aligned} \tag{10}$$

$$\begin{aligned} & (\alpha^T e) \left[g^j - \frac{1}{2} \beta(t)^T g_{xx}^j \beta(t) \right] + \theta(t)^T [g_x^j - Dg_x^j \\ & + (g_{xx}^j - 2Dg_{xx}^j + D^2g_{xx}^j - D^3g_{xx}^j) \beta(t)] + \eta^i(t) = 0 \end{aligned} \tag{11}$$

$$\begin{aligned} & (\alpha^T e) \left[h - \frac{1}{2} \beta(t)^T h_{xx} \beta(t) \right] + \theta(t)^T [h_x - Dh_x \\ & + (h_{xx} - 2Dh_{xx} + D^2h_{xx} - D^3h_{xx}) \beta(t)] = 0 \end{aligned} \tag{12}$$

$$(\theta(t) - (\alpha^T e)\beta)H = 0 \tag{13}$$

$$\theta(t)^T(f_x^i + B^i(t) - Df_x^i - H^i\beta(t)) + \zeta^i + \delta^i = 0 \tag{14}$$

$$\alpha^i B^i(t)\bar{x}(t) + \theta B^i(t) - 2q^i(t)B^i(t)w^i(t) = 0, \quad t \in I, i \in K \tag{15}$$

$$q^i(t)(1 - w^i(t)^T B^i(t)w^i(t)) = 0 \tag{16}$$

$$\eta(t)^T y(t) = 0, \quad t \in I \tag{17}$$

$$\zeta^i \lambda = 0 \tag{18}$$

$$\delta \left(\sum_{i=1}^p \lambda^i - 1 \right) = 0 \tag{19}$$

$$(\alpha, \eta, \zeta, q(t), \delta) \geq 0 \tag{20}$$

$$(\alpha, \theta(t), \eta, \zeta, q(t)\delta) \neq 0 \tag{21}$$

Since $\lambda > 0$, (18) implies $\zeta = 0$.

Since H is non singular, (13) implies

$$\theta(t) = (\alpha^T e)\beta(t), \quad t \in I \tag{22}$$

Using the equality constraint of the dual and (22), we

$$\begin{aligned} & -\sum \alpha^i 0 \left[\begin{aligned} & H\beta(t)^T - \frac{1}{2}(\beta(t)^T H^i \beta(t))_{\bar{x}} + \frac{1}{2}D(\beta(t)^T H^i \beta(t))_{\bar{x}} - \frac{1}{2}D^2(\beta(t)^T H^i \beta(t))_{\bar{x}} \\ & + \frac{1}{2}D^3(\beta(t)^T H^i \beta(t))_{\bar{x}} - \frac{1}{2}D^4(\beta(t)^T H^i \beta(t))_{\bar{x}} \end{aligned} \right] \\ & + (\alpha^T e) \left[\begin{aligned} & H\beta(t)^T + \beta(t)^T (H\beta)_{\bar{x}} - \beta(t)^T D(H\beta)_{\bar{x}} + \beta(t)^T D^2(H\beta)_{\bar{x}} \\ & - \beta(t)^T D^3(H\beta)_{\bar{x}} + \beta(t)^T D^4(H\beta)_{\bar{x}} \end{aligned} \right] = 0 \\ & \left[\begin{aligned} & (\beta(t)^T H^i \beta(t))_{\bar{x}} - D(\beta(t)^T H^i \beta(t))_{\bar{x}} + D^2(\beta(t)^T H^i \beta(t))_{\bar{x}} \\ & - D^3(\beta(t)^T H^i \beta(t))_{\bar{x}} + D^4(\beta(t)^T H^i \beta(t))_{\bar{x}} \end{aligned} \right] \\ & + 2 \left[\begin{aligned} & \beta(t)^T (H\beta)_{\bar{x}} - \beta(t)^T D(H\beta)_{\bar{x}} + \beta(t)^T D^2(H\beta)_{\bar{x}} \\ & - \beta(t)^T D^3(H\beta)_{\bar{x}} + \beta(t)^T D^4(H\beta)_{\bar{x}} \end{aligned} \right] = 0 \end{aligned}$$

This because of the hypothesis (C₂), yields $\beta(t) = 0, t \in I$

Using $\beta(t) = 0, t \in I$ in (22), we have $\theta(t) = 0, t \in I$

Let $\alpha^i = 0, i \in K$, then (11) and (14) respectively give $\eta^i = 0$ and $\delta^i = 0, i \in K$.

Consequently (15) and (16) imply $q^i(t) = 0, t \in I, i \in K$

Thus $(\alpha, \theta(t), \eta, \zeta, q(t), \delta) = 0, i \in K$, where $q(t) = (q^1(t), q^2(t), \dots, q^p(t))$, contradicting the Fritz John condition (21). Hence $\alpha^i > 0, i \in K$.

Using $\theta(t) = 0, \alpha > 0$ and $\beta(t) = 0, t \in I$, from (11), we have

$$g^j = -\frac{\eta^j(t)}{(\alpha^T e)} \leq 0, \quad t \in I, \tag{23}$$

yielding

$$g(t, x, \dot{x}) \leq 0, \quad t \in I$$

The relations (23) and (12) respectively imply that

$$y(t)^T g(t, x, \dot{x}) = 0, \quad t \in I \text{ and } z(t)^T h(t, x, \dot{x}) = 0 \tag{24}$$

Also $g(t, x, \dot{x}) \leq 0$ and $h(t, x, \dot{x}) = 0, t \in I$ imply that \bar{x} is feasible for (NWVEP).

Using $\theta(t) = 0, t \in I, \alpha^i > 0, i \in K$, (15) implies

$$B^i(t)\bar{x}(t) = \left(\frac{2q^i(t)}{\alpha^i}\right)B^i(t)w^i(t) = 0, \quad t \in I, i \in K \tag{25}$$

This yields the equality in the Schwartz inequality. That is,

$$\bar{x}(t)^T B(t)w^i(t) = (\bar{x}(t)^T B(t)\bar{x}(t))^{1/2} (w^i(t)^T B(t)w^i(t))^{1/2}, \quad t \in I, i \in K \tag{26}$$

If $q^i(t) > 0, \forall i \in K, t \in I$, then (16) gives $w^i(t)^T B(t)w^i(t) = 1, t \in I$ and so (26) gives $\bar{x}(t)^T B(t)w^i(t) = (\bar{x}(t)^T B(t)\bar{x}(t))^{1/2}, t \in I, i \in K$

If $q^i(t) = 0, \forall i \in K, t \in I$, then (15) gives $B^i(t)\bar{x}(t) = 0, \forall i \in K, t \in I$

So we will still have

$$\bar{x}(t)^T B(t)w^i(t) = (\bar{x}(t)^T B(t)\bar{x}(t))^{1/2}, \quad t \in I, i \in K \tag{27}$$

In view of (24) and (26) we have

$$\int_I \left\{ f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + (\bar{x}(t)^T B^i(t)\bar{x}(t))^{1/2} \right\} dt = \int_I \left(f^i(t, \bar{x}(t), \dot{\bar{x}}(t)) + \bar{x}(t)^T B^i(t)w^i(t) + y(t)^T g(t, \bar{x}(t), \dot{\bar{x}}(t)) + z(t)^T h(t, \bar{x}(t), \dot{\bar{x}}(t)) - \frac{1}{2} \beta(t)^T H^i \beta(t) \right) dt$$

i.e. the objective values of the problem are equal. By Theorem 1 the efficiency of $\bar{x}(t)$ for (NWVEP) follows.

5 Nondifferentiable Multiobjective Variational Problems with Natural Boundary Values

The following is a pair of Wolfe type second-order nondifferentiable multiobjective variational problems with natural boundary values:

(NVEP)₁: Minimize

$$\left(\int_I \left(f^1(t, x(t), \dot{x}(t)) + (x(t)^T B^1(t)x(t))^{1/2} \right) dt, \dots, \right.$$

$$\left. \int_I \left(f^p(t, x(t), \dot{x}(t)) + (x(t)^T B^p(t)x(t))^{1/2} \right) dt \right)$$

subject to $g(t, x, \dot{x}) \leq 0, h(t, x, \dot{x}) = 0, \quad t \in I$

(WNVED)₂: Maximize

$$\left[\int_I \left(f^1(t, u(t), \dot{u}(t)) + u(t)^T B^1(t)w^1(t) \right. \right. \\ \left. \left. + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) \right) dt, \right. \\ \dots, \int_I \left(f^p(t, u(t), \dot{u}(t)) + u(t)^T B^p(t)w^p(t) \right. \\ \left. \left. + y(t)^T g(t, u(t), \dot{u}(t)) + z(t)^T h(t, u(t), \dot{u}(t)) \right) dt \right]$$

subject to

$$\sum_{i=1}^p \lambda^i \left(f_u^i(t, u(t), \dot{u}(t)) + B^i(t)w^i(t) - Df_u^i(t, u(t), \dot{u}(t)) + y(t)^T g_u(t, u(t), \dot{u}(t)) \right. \\ \left. + z(t)^T h_u(t, u(t), \dot{u}(t)) - D(y(t)^T g_u(t, u(t), \dot{u}(t)) + z(t)^T h_u(t, u(t), \dot{u}(t))) \right) = 0, \quad t \in I$$

$$w^i(t)^T B^i(t)w^i(t) \leq 1, \quad i \in K$$

$$y(t) \geq 0, t \in I, \lambda > 0, \lambda^T e = 1.$$

$$\lambda^T f_{\dot{u}}(t, u(t), \dot{u}(t)) + y(t)^T g_{\dot{u}}(t, u(t), \dot{u}(t)) + z(t)^T h_{\dot{u}}(t, u(t), \dot{u}(t)) = 0, \quad t \in I$$

The duality theorems validated in the preceding section can easily be established with slight modifications.

6 Wolfe Type Second-Order Nondifferentiable Multiobjective Mathematical Programming Problems

If f, g and h variational problems (NWVEP) and (NWVED) are independent of t i.e. functions do not depend explicitly on t, these problems reduce to the following second-order nondifferentiable multiobjective mathematical programming problems:

$$\begin{aligned}
 \text{(NVEP)}_0: \quad & \text{Minimize} \quad \left(f^1(x) + (x^T B^1 x)^{1/2}, \dots, f^p(x) + (x^T B^p x)^{1/2} \right) \\
 & \text{subject to} \quad g(x) \leq 0 \quad h(x) = 0
 \end{aligned}$$

NWVED: Maximize

$$\begin{aligned}
 & \left(f^1(u) + u^T B^1 w^1 + y^T g(u) + z^T h(u) - \frac{1}{2} \beta^T \nabla^2 (\lambda^T f^1(u) + y^T g(u) + z^T h(u)) \beta, \dots, \right. \\
 & \left. f^p(u) + u^T B^p w^p + y^T g(u) + z^T h(u) - \frac{1}{2} \beta^T \nabla^2 (\lambda^T f^p(u) + y^T g(u) + z^T h(u)) \beta \right) \\
 & \text{subject to} \quad \sum \lambda^T (f_u^i(u) + B^i w^i) + y^T g_u(u) + z^T h_u(u) = 0 \\
 & \qquad \qquad \qquad w^{iT} B^i w^i \leq 1 \\
 & \qquad \qquad \qquad y \geq 0, \lambda > 0, \lambda^T e = 1, e = (1, \dots, 1).
 \end{aligned}$$

7 Conclusion

Both Fritz John and Karush-Kuhn-Tucker type optimality conditions for a class of nondifferentiable multiobjective variational problems with equality and inequality constraints are obtained. Here the nondifferentiability occurs due to appearance of a square root function in each component of the objective functional. As an application of Karush-Kuhn-Tucker type optimality conditions Wolfe and Mond-Weir type duals to the problem treated in the research are formulated. Lastly, the relationship between our results and those of their static counterparts has been indicated. The research exposition in this paper has scope to revisit in the context of mixed type second-order duality which has some computational advantage over either of Wolfe or Mond-Weir type duality.

References

1. Mangasarian, O.L.: Second- and higher-order duality in nonlinear programming. *J. Math. Anal. Appl.* **51**, 607–620 (1979)
2. Chen, X.: Second order duality for the variational problems. *J. Math. Anal. Appl.* **286**, 261–270 (2003)
3. Husain, I., Ahmed, A., Masoodi, M.: Second-order duality for variational problems. *Eur. J. Pure Appl. Math.* **2(2)**, 278–295 (2009)

4. Husain, I., Jain, V.K.: On multiobjective variational problems with equality and inequality constraints. *J. Appl. Math. Bioinform.* **3**(2), 15–34 (2013)
5. Chandra, S., Craven, B.D., Husain, I.: A class of nondifferentiable continuous programming problem. *J. Math. Anal. Appl.* **107**, 122–131 (1985)
6. Mangasarian, O.L.: *Nonlinear Programming*. McGraw-Hill, New York (1969)

Three Echelon Supply Chain Design with Supplier Evaluation

Kanika Gandhi, P. C. Jha, Kannan Govindan and Diego Galar

Abstract An effective supply chain management (SCM) facilitates companies to react to changing demand by swiftly communicating their needs to the supplier. Optimizing a supply chain (SC) performance is a key factor for success in long term SC relationships. Substantial information such as price, delivery time percentage and acceptance percentage are discussed in the process. Imprecise demand as one of the factors is added in the same process that fuzzifies coordination between buyer and supplier. The paper considers non-deterministic conditions in the environment of business, coordination in procurement and distribution in a supplier selection problem and a fuzzy model with two objectives is defined. The proposed model is a “fuzzy bi-objective mixed integer nonlinear” problem. To process the solution the fuzzy model is converted into crisp and further fuzzy goal programming approach is employed. The model is validated with the help of a real case problem.

K. Gandhi (✉)

Bharatiya Vidya Bhavan’s Usha & Lakshmi Mittal Institute of Mangement, Copernicus Lane, K. G. Marg, New Delhi 110001, India
e-mail: gandhi.kanika@gmail.com

P. C. Jha

Department of Operational Research, Faculty of Mathematical Sciences, University of Delhi, Delhi 110007, India
e-mail: jhafc@yahoo.com

K. Govindan

Department of Business and Economics, University of Southern Denmark, Odense M, Denmark
e-mail: gov@sam.sdu.dk

D. Galar

Division of Operation and Maintenance Engineering, Lulea University of Technology, Lulea, Sweden
e-mail: diego.galar@ltu.se

Keywords Fuzzy logics · Supplier selection · Supply chain coordination · Truckload and less than truckload

1 Introduction

A key issue in SCM is developing mechanisms to align objectives of independent SC members and coordinating decisions to optimize system performance. Hence to reduce global cost, companies are forced to coordinate all SC activities. Manufacturers have to cooperate with suppliers to maximize productivity in least cost while satisfying customer requirements, as discussed by [1]. A well integrated SC requires coordinated flow of materials and information between suppliers, manufacturers, retailers and other SC components, shown by [2]. An integration of SC activities with discounts and freight policies in two stages is discussed by [3].

For effective coordination, some models on supplier selection are discussed in the literature. The vendor selection problem under quality, delivery, and capacity constraints as well as price-break regimes is considered by [4]. Linear and mixed integer models were formulated for a single objective (cost) problem. The problem of supplier selection and determination of supply quotas from selected vendors when quantity discounts are granted, are dealt by [5]. A mixed integer programming model to select right suppliers to maximize revenue while satisfying customer needs is developed by [6]. Changes in suppliers' capabilities are considered along with customer requirements. In the model, suppliers meeting many parts of the ideal procurement condition are selected more often than other suppliers.

In continuation of the supplier selection problem, the issue of coordination between the suppliers and buyer is also addressed. A minimization model with single objective for the total cost of SC is proposed by [7]. The chain includes a buyer and some potential suppliers. The aim of their model is minimizing total cost of chain (a buyer and suppliers) by selecting suppliers and determining orders from each of them. The study shows that minimizing total SC cost (including cost of buyer and suppliers) ensures better results for SC rather than individual decision-making. Further, some more literature on supplier selection and fuzzy optimization was discussed. A fuzzy goal programming approach applied to the problem of selecting vendors in a SC is proposed by [8]. This problem is a mixed-integer and fuzzy goal programming problem with three objectives; to minimize the net cost of vendor network, rejects within the network and delays in deliveries. With this approach, the author uses triangular fuzzy number for each fuzzy objective. The solution applies a minimum operator on the intersection of the membership function of the fuzzy objectives. A similar problem is solved by [9], using the multi-objective fuzzy programming approach proposed by [10]. Further, the problem of adequately selecting suppliers within a SC is discussed by [11]. The literature discussed above has revealed models for supplier selection or to show coordination between supplier and buyer by minimizing the cost.

To manage the minimum coordination cost and simultaneously measuring suppliers' performances in the conditions of uncertainty, the paper explains two objectives of the mixed integer non-linear model. The first objective of the model shows an integration of procurement and distribution of a three echelon SC model with multi sources (suppliers), one warehouse & one destination (buyer), incorporating transportation policies. The second objective concentrates on suppliers' performance and their selection on the basis of delivery time percentage and acceptance percentage of the ordered lot. The model integrates inventory, procurement and transportation mechanism to minimize all costs discussed above and also chooses the best supplier. The total cost of the model becomes fuzzy due to fuzzy holding cost and consumption. On the other hand, performance level is also fuzzy as percentage of on-time delivery and acceptances are fuzzy. So the model discussed above is fuzzy bi-objective mixed integer non-linear model.

In the solution process, the fuzzy model is converted into crisp and further fuzzy goal programming approach is employed where each objective is assigned with a different weight. The imprecise objectives and imprecise demands considered are related to cutting costs, increasing quality and improved service to selected suppliers.

2 Model Assumptions, Sets and Symbols

2.1 Assumptions

The assumptions of this research are essentially the same as an EOQ model, except for the transportation cost. The section considers a finite planning horizon. The consumption at destination is uncertain, and no shortages are allowed. The initial inventory of each product is positive at the start of planning. Inventory deteriorates with constant rate as a percentage of deterioration of stored units and inspection cost is also assumed to be constant. No transportation is considered from source to warehouse as it is taken care of by supplier.

2.2 Sets

Products set with Cardinality P and indexed by i , whereas periods set with Cardinality T and indexed by t . Supplier set with cardinality J and indexed by j .

2.3 Parameters

\tilde{C} is fuzzy total cost, C_0 & C_0^* are aspiration & tolerance level of fuzzy total cost, \widetilde{PR} & \overline{PR} are fuzzy & de-fuzzified performance of supplier, PR_0^* & PR_0 are tolerance & aspiration level of fuzzy performance of supplier, \widetilde{HS}_{it} & \overline{HS}_{it} & are fuzzy & de-fuzzified holding cost per unit of product i in period t at warehouse, ϕ_{it} is unit purchase cost for product i in period t , s is cost per weight of transportation in LTL policy, β_t is fixed freight cost for each truck load in period t , \widetilde{HD}_{it} & \overline{HD}_{it} are fuzzy & defuzzified holding cost per unit of product i for t th period at destination, λ_{it} is inspection cost per unit of product i in period t , \widetilde{CR}_{it} & \overline{CR}_{it} are fuzzy & defuzzified consumption at destination for product i in period t , ISN_{i1} is inventory level at warehouse in beginning of planning horizon for product i , IDN_{i1} is inventory level at destination in beginning of planning horizon for product i , η is deterioration percentage of i th product at destination, w_i is per unit weight of product i , ω is weight transported in each full truck, \widetilde{DT}_{ijt} & \overline{DT}_{ijt} are fuzzy & defuzzified percentage of on-time delivery time for product i in period t for supplier j , \widetilde{AC}_{ijt} & \overline{AC}_{ijt} are fuzzy & defuzzified percentage of acceptance for product i in period t for supplier j , CP_{ij} is capacity at supplier j for product i .

2.4 Decision Variables

IS_{it} is inventory level at warehouse at the end of period t for product i , ID_{it} is inventory levels at warehouse at the end of period t for product i , X_{ijt} is optimum ordered quantity of product i ordered in period t from supplier j , X_{it} is total optimum quantity ordered from all the suppliers, L_t is total weighted quantity transported in stage I & II respectively of period t , J_t is total number of truck loads in period t , y_t is weighted quantity in excess of truckload capacity, u_t is usage of modes, either TL & LTL mode (where value is 1) or only TL mode (where value is 0), V_{ijt} is defined as, if ordered quantity is transported by supplier j for product i in period t then the variable takes value 1, otherwise zero, D_{it} is demand for product i in period t .

3 Model Formulation

3.1 Formulation of Objectives

Fuzzy optimization is a flexible approach that permits adequate solutions of real problems when vague information is available, providing well defined mechanisms

to quantify uncertainties directly. Therefore, we formulate fuzzy optimization model for vague aspiration levels on total cost, consumption, on-time delivery percentage and acceptance percentage where the decision maker may decide his aspiration levels on the basis of past experience and knowledge possessed by him.

Initially a bi-objective fuzzy model is formulated which discusses about fuzzy total cost and performance of the suppliers. The first objective of the model minimizes the total cost, including purchasing cost of goods from supplier, holding cost at warehouse for ordered quantity, transportation cost from warehouse to destination, cost of holding at destination and finally inspection cost of the reached quantity at destination.

$$\begin{aligned} \text{Minimize } \tilde{C} = & \sum_{t=1}^T \sum_{j=1}^J \sum_{i=1}^P \phi_{ijt} X_{ijt} V_{ijt} + \sum_{t=1}^T \sum_{i=1}^P \widetilde{HS}_{it} X_{it} \\ & + \sum_{t=1}^T [(sy_t + j_t \beta_t) u_t + (j_t + 1) \beta_t (1 - u_t)] \\ & + \sum_{t=1}^T \widetilde{HD}_{it} ID_{it} + \sum_{i=1}^P \sum_{t=1}^T \lambda_{it} X_{it} \end{aligned}$$

The second objective discusses the performance of suppliers and maximizes the performance percentage of supplier as per delivery time percentage and acceptance percentage of ordered lot.

$$\text{Maximize } \widetilde{PR} = \sum_{t=1}^T \sum_{j=1}^J \sum_{i=1}^P \left(\widetilde{DT}_{ijt} + \widetilde{AC}_{ijt} \right) V_{ijt}$$

3.2 Constraints' Formulation

The constraints in the model handle the capacity restrictions, shortages restrictions. The following constraint ensures that an activated supplier cannot supply more that capacity.

$$X_{ijt} \leq CP_{ij} V_{ijt} \quad \forall i, j, t$$

X_{it} calculates total quantity to be supplied from all active suppliers.

$$X_{it} = \sum_{j=1}^J X_{ijt} \quad \forall i, t$$

In a period, for a particular product, only one supplier will be allowed to supply goods is assured by the constraint $\sum_{j=1}^J V_{ijt} = 1 \quad \forall i, t$

Following three equations calculate inventory in period t at warehouse and ensure no shortage at destination.

$$\begin{aligned}
 IS_{it} &= IS_{it-1} + X_{it} - D_{it} \quad \forall i, t > 1 \\
 IS_{it} &= ISN_{it} + X_{it} - D_{it} \quad \forall i, t = 1 \\
 \sum_{t=1}^T IS_{it} + \sum_{t=1}^T X_{ij} &\geq \sum_{t=1}^T D_{it} \quad \forall i
 \end{aligned}$$

Next equation is an integrator and calculates the total weighted quantity to be transported from warehouse to destinations.

$$L_t = \sum_{i=1}^P \omega_i X_{it} \quad \forall t$$

The constraint mentioned below checks the transportation policy as per the weighted quantity. It clearly specifies that, if the total weighted quantity is above the capacity of truck then LTL policy will get activated and otherwise only TL policy will be used.

$$L_t \leq (y_t + j_t w) u_t + (j_t + 1) w (1 - u_t) \quad \forall t$$

The equation measures the overhead weights from truckload capacity.

$$L_t = y_t + j_t w \quad \forall t$$

Next three constraints calculate inventory and optimum demand size at the destination, while considering deterioration percentage and ensuring no shortages.

$$\begin{aligned}
 ID_{it} &= ID_{it-1} + D_{it} - \widetilde{CR}_{it} - \eta ID_{it} \quad \forall i, t > 1 \\
 ID_{it} &= IDN_{it} + D_{it} - \widetilde{CR}_{it} - \eta ID_{it} \quad \forall i, t = 1 \\
 (1 - \eta) \sum_{t=1}^T ID_{it} + \sum_{t=1}^T D_{it} &\geq \sum_{t=1}^T \widetilde{CR}_{it} \quad \forall i
 \end{aligned}$$

Finally constraint mentioned below enforces the binary and non-negativity restrictions on decision variables.

$$X_{ijt}, X_{it}, L_t, D_{it} \geq 0; V_{ijt}, u_t \in [0, 1]; IS_{it}, ID_{it}, y_t, j_t \text{ are integer.}$$

3.3 Formulated Fuzzy Model

$$\begin{aligned}
 \text{Minimize } \tilde{C} &= \sum_{t=1}^T \sum_{j=1}^J \sum_{i=1}^P \phi_{ijt} X_{ijt} V_{ijt} + \sum_{t=1}^T \sum_{i=1}^P \widetilde{HS}_{it} X_{it} + \sum_{t=1}^T [(s y_t + j_t \beta_t) u_t + (j_t + 1) \beta_t (1 - u_t)] \\
 &+ \sum_{t=1}^T \widetilde{HD}_{it} ID_{it} + \sum_{i=1}^P \sum_{t=1}^T \lambda_{it} X_{it}
 \end{aligned} \tag{1}$$

$$\text{Maximize } \widetilde{PR} = \sum_{t=1}^T \sum_{j=1}^J \sum_{i=1}^P (\widetilde{DT}_{ijt} + \widetilde{AC}_{ijt})V_{ijt} \tag{2}$$

$$\text{Subject to } X_{ijt} \leq CP_{ij}V_{ijt} \quad \forall i, j, t \tag{3}$$

$$X_{it} = \sum_{j=1}^J X_{ijt} \quad \forall i, t \tag{4}$$

$$\sum_{j=1}^J V_{ijt} = 1 \quad \forall i, t \tag{5}$$

$$IS_{it} = IS_{it-1} + X_{it} - D_{it} \quad \forall i, t > 1 \tag{6}$$

$$IS_{it} = ISN_{it} + X_{it} - D_{it} \quad \forall i, t = 1 \tag{7}$$

$$\sum_{t=1}^T IS_{it} + \sum_{t=1}^T X_{ij} \geq \sum_{t=1}^T D_{it} \quad \forall i \tag{8}$$

$$L_t = \sum_{i=1}^P \omega_i X_{it} \quad \forall t \tag{9}$$

$$L_t \leq (y_t + j_t w)u_t + (j_t + 1)w(1 - u_t) \quad \forall t \tag{10}$$

$$L_t = y_t + j_t w \quad \forall t \tag{11}$$

$$ID_{it} = ID_{it-1} + D_{it} - \widetilde{CR}_{it} - \eta ID_{it} \quad \forall i, t > 1 \tag{12}$$

$$ID_{it} = IDN_{it} + D_{it} - \widetilde{CR}_{it} - \eta ID_{it} \quad \forall i, t = 1 \tag{13}$$

$$(1 - \eta) \sum_{t=1}^T ID_{it} + \sum_{t=1}^T D_{it} \geq \sum_{t=1}^T \widetilde{CR}_{it} \quad \forall i \tag{14}$$

$$X_{ijt}, X_{it}, L_t, D_{it} \geq 0; V_{ijt}, u_t \in [0, 1]; IS_{it}, ID_{it}, y_t, j_t \text{ are integer.} \tag{15}$$

4 Solution Algorithm

4.1 Fuzzy Solution Algorithm

The following algorithm specifies the sequential steps to solve the fuzzy mathematical programming problems, discussed by [12].

Step 1. Compute the crisp equivalent of the fuzzy parameters using a defuzzification function. Here, ranking technique is employed to defuzzify the parameters as $F_2(A) = (a_l + 2a_m + a_u)/4$, where a_l, a_m, a_u are the Triangular Fuzzy Numbers (TFN).

Let \overline{CR}_i be the defuzzified value of \widetilde{CR}_i and $(CR_{it}^1, CR_{it}^2, CR_{it}^3)$ for each i & t be triangular fuzzy numbers then, $\overline{CR}_i = (CR_{it}^1 + 2CR_{it}^2 + CR_{it}^3)/4$. Similarly, \overline{HS}_i and \overline{HD}_i are defuzzified aspired holding cost at warehouse and destination.

Step 2. Since industry is highly volatile and customer demand changes in every short span, a precise estimation of cost and performance aspirations is a major area of discussion. Hence the better way to come out of such situation is to incorporate tolerance and aspiration level with the main objectives. So the model discussed in Sect. 3.3 can be re-written as follows:

$$\begin{aligned}
 & \text{Find } X \\
 & X \in S \\
 & (1 - \eta) \sum_{i=0}^T ID_{it} + \sum_{i=0}^T D_{it} \underset{\sim}{\geq} \sum_{i=0}^T \overline{CR}_{it} \quad \forall i \\
 & C(X) \underset{\sim}{\leq} C_0 \\
 & PR \underset{\sim}{\geq} PR_0 \\
 & X_{ijt}, X_{it}, L_t, D_{it} \geq 0; V_{ijt}, u_t \in [0, 1]; IS_{it}, ID_{it}, y_t, j_t \text{ are integer}
 \end{aligned}$$

Step 3. Define appropriate membership functions for each fuzzy inequalities as well as constraint corresponding to the objective functions.

$$\begin{aligned}
 \mu_C(X) &= \begin{cases} 1 & ; C(X) \leq C_0 \\ \frac{C_0^* - C(X)}{C_0^* - C_0} & ; C_0 \leq C(X) < C_0^* \\ 0 & ; C(X) > C_0^* \end{cases} \\
 \mu_{PR}(X) &= \begin{cases} 1 & ; PR \geq PR_0 \\ \frac{PR - PR_0^*}{PR_0 - PR_0^*} & ; PR_0^* \leq PR < PR_0 \\ 0 & ; PR < PR_0^* \end{cases} \\
 \mu_{ID_{it}}(X) &= \begin{cases} 1 & ; ID_{it}(X) \geq \overline{CR}_0 \\ \frac{ID_{it}(X) - \overline{CR}_0^*}{\overline{CR}_0 - \overline{CR}_0^*} & ; \overline{CR}_0^* \leq ID_{it}(X) < \overline{CR}_0 \\ 0 & ; ID_{it}(X) > \overline{CR}_0^* \end{cases}
 \end{aligned}$$

where $\overline{CR}_0 = \sum_{i=1}^T \overline{CR}_{it}$ is the aspiration and \overline{CR}_0^* is the tolerance level for inventory constraints.

Step 4. Employ extension principles to identify the fuzzy decision, which results in a crisp mathematical programming problem given by

$$\begin{aligned}
 & \text{Maximize } \alpha \\
 & \text{Subject to } \mu_c(X) \geq w_1\alpha \\
 & \mu_{PR}(X) \geq w_2\alpha \\
 & \mu_{ID_i}(X) \geq \alpha \\
 & X \in S \\
 & w_1 \geq 0, w_2 \geq 0, w_1 + w_2 = 1, \alpha \in [0, 1]
 \end{aligned}$$

where, α represents the degree to which the aspiration of the decision-maker is met. The above problem can be solved by the standard mathematical programming algorithms.

Step 5. Following [13], while solving the problem by following steps 1–4, the objective of the problem is also treated as a constraint. Each constraint is considered to be an objective for the decision-maker and the problem is looked at as a fuzzy bi objective mathematical programming problem. Further, each objective can have a different level of importance and can be assigned weights to measure relative importance. The resulting problem can be solved by the weighted min max approach. On substituting values for $\mu_{PR}(x)$ and $\mu_c(x)$ the problem becomes

$$\begin{aligned}
 & \text{Maximize } \alpha \\
 & \text{subject to } PR(X) \geq PR_0 - (1 - w_1\alpha)(PR_0 - PR_0^*) \\
 & C(X) \leq C_0 + (1 - w_2\alpha)(C_0^* - C_0) \\
 & \mu_{ID_i}(X) \geq \alpha \\
 & X \in S \\
 & w_1 \geq 0, w_2 \geq 0, w_1 + w_2 = 1, \alpha \in [0, 1]
 \end{aligned} \tag{P1}$$

Step 6. If a feasible solution is not possible in Step 5, then fuzzy goal programming approach is resorted to obtain a compromised solution given by [14]. The method is discussed in detail in the next section.

4.2 Fuzzy Goal Programming Approach

On solving the problem, we found that the problem (P1) is not feasible; hence the management goal cannot be achieved for a feasible value of $\alpha \in [0,1]$. We then apply the fuzzy goal programming technique to obtain a compromised solution. The approach is based on the goal programming technique to solve the crisp goal programming problem given by [14]. The maximum value of any membership

function can be 1; maximization of $\alpha \in [0,1]$. This can be achieved by minimizing the negative deviational variables of goal programming (i.e., η) from 1. The fuzzy goal programming formulation for the given problem (P1) introducing the negative and positive deviational variables η_j, ρ_j is given as

$$\begin{aligned}
 & \text{Minimize } u \\
 & \text{subject to } \mu_{PR}(X) + \eta_1 - \rho_2 = 1 \\
 & \quad \mu_C(X) + \eta_2 - \rho_2 = 1 \\
 & \quad u \geq w_j * \eta_j \quad j = 1, 2 \\
 & \quad \eta_j * \rho_j = 0 \quad j = 1, 2 \\
 & \quad w_1 + w_2 = 1 \\
 & \quad \alpha = 1 - u \\
 & \eta_j, \rho_j \geq 0, X \in S, u \in [0, 1], w_1, w_2 \geq 0
 \end{aligned}$$

The above described model is coded into Lingo 11.0 to find the optimal solution.

5 Case Study

The modern food industry has developed and expanded due to its ability to deliver a wide variety of high quality food products to consumers on a nationwide/worldwide basis. This feat has been accomplished by building stability into the product through processing, packaging, and additives that enable foods to remain fresh and wholesome throughout the distribution process. The reasons for processed food deterioration may be microbiological spoilage, sometimes accompanied by pathogen development or chemical and enzymatic activity.

The same is the problem of BTR Co. (named changed), who purchases fruits & vegetables from wholesale suppliers, processes them at a processing point (warehouse) to increase their shelf life so that deterioration is reduced at retail stores; then they are transported to different retail stores. In this case, we are discussing a tiny problem of three suppliers (source point) and one retail outlet (destination) with one processing point (warehouse). Four products including Spinach (P1), broccoli (P2), peas (P3) and corn (P4) are considered for managing procurement and distribution for 3 weeks in summer. The major objectives of the company are to manage optimal order quantity from suppliers, as consumption at the retail end is uncertain, reducing fuzziness in environment to ensure that the company can calculate precise consumption, cost for future and select the best supplier, who can deliver the required quantity on-time with an acceptable lot. The company desires to reduce procurement cost, holding cost at warehouse, cost of transportation from warehouse to retail outlet, inspection cost of reached quantity at outlet and ending inventory carrying cost at outlet. From the above discussion it is clear that for the company, both the objectives i.e. minimizing the cost and maximizing the performance are equally important. That helps the solution process to consider both the objectives equal. A fuzzy optimization model is developed, converted into a crisp form and a fuzzy goal programming approach is employed

Table 1 Holding cost at warehouse (in INR) per packet

PDT	Period 1		Period 2		Period 3	
	Fuzzy	DF	Fuzzy	DF	Fuzzy	DF
P1	3, 2, 3	2.5	2.5, 3, 2.7	2.8	2.5, 2, 2.7	2.3
P2	2, 2.6, 2	2.3	2.6, 3.5, 2.4	3	2.3, 2.6, 2.5	2.5
P3	3, 3.4, 3.8	3.4	2.5, 3, 2.3	2.7	2.5, 2, 1.5	2
P4	3, 3.5, 2.8	3.2	2, 2.5, 1.8	2.2	2, 2.4, 1.6	2.1

Where *PDT* Product type; *DF* Defuzzified

Table 2 Holding cost at retail outlet (in INR) per packet

PDT	Period 1		Period 2		Period 3	
	Fuzzy	DF	Fuzzy	DF	Fuzzy	DF
P1	3, 3.5, 2.8	3.2	3.3, 3, 3.1	3.1	2.5, 3.2, 3.1	3.0
P2	3, 2.6, 3	2.8	2.6, 3, 3	2.9	2.3, 2.9, 2.3	2.6
P3	3, 3.8, 3.4	3.5	3.4, 2.3, 3.6	2.9	2.3, 2.8, 2.1	2.5
P4	3.9, 3.5, 3.5	3.6	2.7, 2.5, 2.7	2.6	2, 2.4, 2.4	2.3

to reach at a feasible and optimal solution for company’s current situation. The company’s data is as follows:

The purchase cost of each product discussed as P1 costs ₹45, ₹48, ₹50 per 1.03 kg, P2 costs ₹64, ₹62, ₹67 per 0.5 kg, P3 costs ₹53, ₹56, ₹54 per 1 kg and cost of P4 is ₹59, ₹54, ₹57 per 0.5 kg. Initial inventory at warehouse in the beginning of the planning horizon is 70, 80, 40, and 59 packets. Similarly initial inventory at destination is 90, 85, 79, and 83 packets. The inspection cost per unit is ₹1 for P1 & P3 and ₹1.5 for P2 & P4. While transporting weighted quantity from warehouse to destination, TL<L policies may be used. In such a case, cost per full truck is ₹950, ₹1000, and ₹1200, and capacity per truck is 250 kg. In the case of LTL policy, per extra weight costs ₹4. The deterioration percentage at destination is 6 % (Tables 1, 2, 3, 4, 5, 6).

5.1 Results and Managerial Implications

The total cost incurred in coordinating all entities is ₹92,734.72, keeping the fuzzy aspiration levels as ₹81,000, ₹89,000, ₹93,000 and defuzzified aspirations cost as ₹88,000 with tolerance level of ₹115,000. On the other side the suppliers’ performance is 10.6 with respect to level of aspiration 11 keeping three fuzzy aspirations as 10.3, 11.1, 11.5 and tolerance level as 10 of the supplier. Nearby 78 % of the aspiration level of cost and performance has been attained which makes the environment more certain and crisp for future discussions. The model tries to

Table 3 Consumption at retail outlet per packet

PDT	Period 1		Period 2		Period 3	
	Fuzzy	DF	Fuzzy	DF	Fuzzy	DF
P1	160, 169, 178	169	134, 139, 140	138	189, 184, 183	185
P2	137, 133, 137	135	179, 174, 177	176	186, 179, 184	182
P3	170, 165, 168	167	187, 183, 183	184	168, 163, 182	169
P4	176, 172, 176	174	180, 182, 172	179	160, 164, 172	165

Table 4 On-time delivery percentage

PDT	Period 1		Period 2		Period 3	
	Fuzzy	DF	Fuzzy	DF	Fuzzy	DF
P1	0.91, 0.95, 0.91	0.93	0.95, 0.93, 0.99	0.95	0.93, 0.91, 0.89	0.91
P2	0.95, 0.93, 0.95	0.94	0.89, 0.96, 0.91	0.93	0.89, 0.96, 0.91	0.93
P3	0.92, 9.96, 9.96	0.95	0.86, 0.91, 0.92	0.90	0.91, 0.95, 0.99	0.95
P4	0.88, 0.93, 0.90	0.91	0.90, 0.95, 0.92	0.93	0.89, 0.93, 0.89	0.91

Table 5 Acceptance percentage

PDT	Period 1		Period 2		Period 3	
	Fuzzy	DF	Fuzzy	DF	Fuzzy	DF
P1	0.90, 0.95, 0.92	0.93	0.91, 0.95, 0.95	0.94	0.93, 0.97, 0.89	0.94
P2	0.89, 0.93, 0.89	0.91	0.89, 0.96, 0.99	0.95	0.88, 0.92, 0.88	0.90
P3	0.89, 0.91, 0.89	0.90	0.86, 0.91, 0.96	0.91	0.91, 0.95, 0.95	0.94
P4	0.88, 0.93, 0.94	0.92	0.90, 0.95, 0.88	0.92	0.89, 0.93, 0.93	0.92

Table 6 Supplier’s capacity (in no. of packets)

PDT	S1	S2	S3
P1	250	300	130
P2	120	220	170
P3	350	150	340
P4	230	120	240

Where S Supplier

employ high performers to procure ordered quantity and ensure that only the best supplier shall fulfill the demand. Table 7 shows the exact quantity procured from suppliers per product per period as per the performance.

The positive quantity in the table indicates that only the corresponding supplier is activated to supply goods as they have highest performance among the three. Some exceptions also exist, as in period 2 for product 1 and 2, supplier 2 is activated but no supply takes place.

Table 7 Optimum ordered quantity from supplier (S1–S3)

PDT	Period 1			Period 2			Period 3		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
P1	0	269	0	0	0	0	0	70	0
P2	0	216	0	0	0	0	0	118	0
P3	0	0	114	0	0	190	0	0	102
P4	230	0	0	0	120	0	0	32	0

Table 8 Inventory (in no. of packets)

PDT	Warehouse inventory			Destination inventory		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
P1	176	92	0	79	23	0
P2	153	54	0	87	10	0
P3	0	41	0	62	25	0
P4	156	75	0	39	58	48

Table 9 Demand at destination (in no. of packets)

PDT	Period 1	Period 2	Period 3
P1	163	84	162
P2	143	99	172
P3	154	149	144
P4	133	201	107

Table 10 Transported weights, no. of trucks, transportation mode, overhead weight

	Period 1	Period 2	Period 3
Transported quantity	613	250	250
No. of trucks	2	1	1
Transportation mode	TL<L	TL	TL
Overhead quantity	113	0	0

Table 8 shows the ending inventory figures. It is observed that in almost all cases in the last periods, in-hand inventory is consumed and reaches zero or a small positive value.

The demand at destinations for source is shown in Table 9 and depends on consumption at the destination. The demand of the product gives the basic idea of the quantity to be ordered.

While transporting weighted quantity to the demand point, policy type, number of trucks and overhead weights are to be checked as each of them incurs cost to the company. In Table 10, it is observed that in period 1 TL<L policy is employed, as there is a positive overhead quantity. And in the other two periods only TL

policy as there is no overhead quantity. In the case of TL<L policy, if overhead weighted quantity is transported through full TL, the cost of transportation will become much higher than using LTL policy.

6 Conclusion

To handle uncertainty in SCM, a fuzzy two-objective mixed integer nonlinear model is proposed in this paper. The considered model includes the supplier selection and procurement, distribution and coordination. Since many parameters are imprecise in nature, a fuzzy goal programming is employed in the solution process. The current study proposes a two objective fuzzy mathematical model for the literature gap identified and mentioned in this study. The first objective is to provide optimum cost by selecting right transportation policies for optimum quantity. The first objective is not completed without the help of the second, as the second objective selects the best supplier for each with respect to delivery and acceptance percentage. Finally, the current study finds a balance between cost and suppliers.

References

1. Jain, V., Wadhwa, S., Deshmukh, S.G.: Select supplier-related issues in modelling a dynamic supply chain: potential, challenges and direction for future research. *Int. J. Prod. Res.* **47**, 3013–3039 (2009)
2. Narasimhan, R., Carter, J.R.: Linking business unit and material sourcing strategies. *J. Bus. Logist.* **19**, 155–171 (1998)
3. Gandhi, K., Jha, P.C., Ali, S.S.: A two stage EOQ model for deteriorating products incorporating quantity & freight discounts, under fuzzy environment. In: Bansal, J.C., Singh, P.K., Deep, K., Pant, M., Nagar, A.K. (eds.) *Advances in Intelligent Systems and Computing*, 202nd edn, pp. 367–380. Springer, India (2012)
4. Chaudhry, S.S., Frost, F.G., Zydiak, J.L.: Vendor selection with price breaks. *Eur. J. Oper. Res.* **70**, 52–66 (1993)
5. Xu, J., Lu, L.L., Glover, F.: The deterministic multi-item dynamic lot size problem with joint business volume discount. *Ann. Oper. Res.* **96**, 317–337 (2000)
6. Hong, G.H., Park, S.C., Janga, D.S., Rho, H.M.: An effective supplier selection method for constructing a competitive supply-relationship. *Expert Syst. Appl.* **28**, 629–639 (2005)
7. Kheljani, J.G., Ghodspour, S.H., Ghomi, S.M.T.F.: Supply chain optimization policy for a supplier selection problem: a mathematical programming approach. *Iran. J. Oper. Res.* **2**, 17–31 (2010)
8. Kumar, M., Vart, P., Shankar, R.: A fuzzy goal programming approach for vender selection problem in a supply chain. *Comput. Ind. Eng.* **46**, 69–85 (2004)
9. Kumar, M., Vart, P., Shankar, R.: A fuzzy programming approach for vender selection problem in a supply chain. *Int. J. Prod. Econ.* **101**, 273–285 (2006)
10. Zimmermann, H.J.: Fuzzy linear programming with several objective functions. *Fuzzy Sets Syst.* **1**, 46–55 (1978)
11. Amid, A., Ghodspour, S., O'Brien, C.: Fuzzy multi objective linear model for vender selection problem in a supply chain. *Int. J. Prod. Econ.* **104**, 394–407 (2006)

12. Zimmermann, H.J.: Description and optimization of fuzzy systems. *Int. J. Gen. Syst.* **2**, 209–215 (1976)
13. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. *Manag. Sci.* **17**, 141–164 (1970)
14. Mohamed, R.H.: The relationship between goal programming and fuzzy programming. *Fuzzy Sets Syst.* **89**, 215–222 (1997)

A Carbon Sensitive Multi Echelon Reverse Logistics Network Design for Product Value Recovery

Jyoti Dhingra Darbari, Vernika Agarwal and P. C. Jha

Abstract Increase in environmental concerns and consumer awareness together with imposition of government regulations are forcing electronic industries to set up their own reverse logistics network. An evaluation of the impact of their reverse logistics activities is also imperative for the design of an effective and efficient reverse supply chain. This paper presents a bi objective mixed integer linear programming model for a reverse logistics network design that considers value added recovery of return of end of life (EOL) products and also focuses on controlling the transportation activities involved in the reverse logistics system, a major contributor to the increase in carbon emission. The objectives are to maximise the product's value recovery by determining the optimum flow of products and components across facilities in the network and minimise the carbon emission by determining the optimum routes to be taken by vehicles and by appropriate selection of vehicles. The reverse logistics model developed is goal programming model and it takes into account two objectives with almost equal weightage. The model captures the trade-offs between total profit and emission of CO₂. The model is justified by a case study in the context of the reverse logistics network design of an Indian company manufacturing air conditioners and refrigerators.

Keywords Reverse logistics · Fabrication · Component fabrication · Carbon emission

J. D. Darbari (✉) · V. Agarwal · P. C. Jha
Department of Operational Research, University of Delhi, Delhi, India
e-mail: jydbdr@hotmail.com

V. Agarwal
e-mail: vernika.agarwal@gmail.com

P. C. Jha
e-mail: jhpc@yahoo.com

1 Introduction

The Electronics industry is one of the most dynamic industries in the world. The continuous upgradation of the electronic products requires large inputs of material and resources and also results in large surplus of unwanted electronics products. An effective reverse supply chain should be designed to handle the tremendous increase in the return flows of these products by allowing the product to reenter the network at a facility where some value can be recovered. Typically, the product return process is more complicated than forward logistics operations. Recognizing this inherent complexity of the product return process many researchers have worked in this field. Reverse Logistics (RL) design models have traditionally focused on minimizing costs such as those of [1] and [2] or maximising profit proposed by [3] and [4]. Recently, the focus has also been on reducing the carbon emission generated by the reverse logistics activities. Waste and pollutant gases released due to transportation activities are biggest issue in the supply chain as 35 % ratio of transportation greenhouse gas emission is actualized via heavy and light trucks used in supply chain [5]. Elhedhli and Merrick [6] Present a problem formulation that minimizes the combined expenses associated with the fixed costs, transportation and cost of emissions generated on the shipping lanes. Sundarakani et al. [7] Examine the carbon footprint across supply chains. [8] Proposed three echelons, supply chain design model that derives nonlinear concave expressions relating vehicle weight to CO₂ emissions. Vehicle routing has also been an important factor in RL as an efficient routing helps in minimising both the cost and carbon emission associated with the network and has been studied by [9] and [10].

In this paper, a mixed integer linear programming model has been designed for a RL network which maximizes profit while managing the returns of products and also accounts for reducing the carbon emission. The bi-objective model is single period, multi echelon and multi product and it coordinates the collection, inspection, fabrication, dismantling, recycling and component fabrication of the returned products. A vehicle routing approach for the transport of the returned products is used to minimize the distance of transportation between the collection, fabrication and dismantling centres. For the flow of fabricated components, selection of vehicle is done based on optimum vehicle capacity. The goal of the model is to simultaneously maximise profit by optimally determining the flow of products and components across various facilities and minimize the CO₂ emission by strategically routing the vehicles and selecting the vehicles optimally.

Following the introduction, the rest of the paper is organised as follows: [Sect. 2](#) which provides a discussion of research problem, the mathematical model and the formulation of the model; in [Sect. 3](#), we present a case study and [Sect. 4](#) concludes the paper.

2 Model Description and Formulation

This study aims at designing a RL network for value recovery of end of use, defective and broken electronic products returned by customers as well as the obsolete products discarded by the manufacturer. Managing huge number of discarded products is one of the major concerns of the manufacturer. Through product fabrication, the life span of the product can be extended and it creates an opportunity for reselling these at the secondary market. Old and defective products can be disassembled and key components can be recovered and reused after fabrication. Increase in profitability can be attained through resale of the fabricated components at the spare market. The amount of carbon emissions resulting from the transportation activities is also a growing concern for the manufacturer. The RL operations are therefore aimed at avoiding unnecessary transportation moves thereby helping to reduce carbon emissions.

The network includes customer zones, one collection zone consisting of collection/inspection centers (CIC), a fabrication center (FC), an integrated disassembly (DC) and component fabrication center (CFC), a recycling center, a disposal center, secondary markets, spare market and a company owned service center as shown in Fig. 1. The customers are grouped in zones and all customer zones are allocated one collection zone consisting of CICs. A CIC can cater to more than one customer zone depending on its capacity and its approximately so the transportation between them is negligible. Manufacturer has a reliable estimate of the amount of returns from each customer zone. The returned products are transported from customer zones to the allocated CIC by a collection agent. After inspection at CIC, the returned products are bifurcated into products that can be fabricated and the rest that are to be dismantled. A fraction of the returned products are sent for fabrication depending upon the condition λ of the products and also on their demand in the secondary markets. The remaining volumes of products are transported to DC for disassembling into components. The components are further inspected at the integrated facility and are classified according to the best component recovery decision which includes component fabrication, recycling or disposal. A fraction of the suitable disassembled components are fabricated at CFC unit of the same center. The fabricated components are further sent to either FC (to be used as new components) or the service centre after realising the demand at FC and service center and the rest are sold in the spare market. The components which are not fit to be reused are either collected by a recycling agent for recycling or sent further for disposal. The old components which are replaced at FC are either send for disposal or are sold to the recycling agent. The recycling agent pays per unit revenue for the components collected but the cost of disposal for the remaining components which includes the transportation cost is to be borne by the manufacturer. Currently, there is one collection zone (a cluster of collection center), one DC and one FC. The manufacturer has a contract with a 3PL which provides a fleet of vehicles of substantial capacities (assumed to be identical) for the product flow and smaller trucks of different

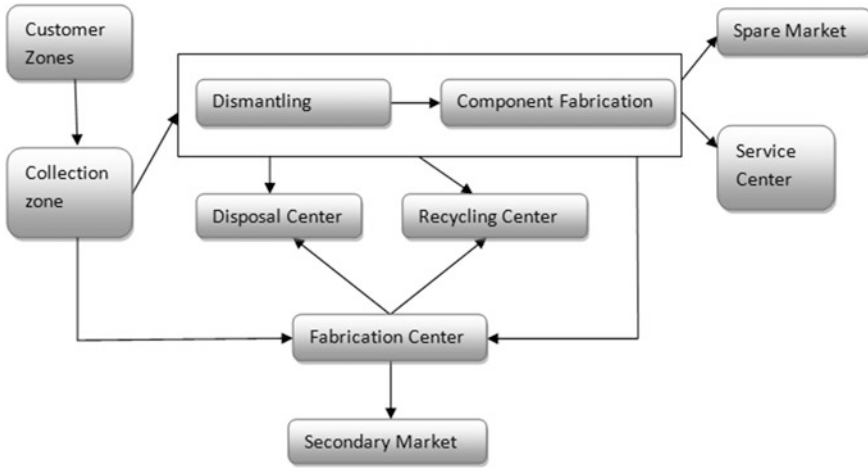


Fig. 1 Reverse logistics network design

capacities for the component flow. We use vehicle routing problem to develop the logistics for transportation of the returned products. The vehicles are stationed at FC and DC which act as central depots. The vehicles start from FC/DC and collect the returned products which need to be fabricated/dismantled from the CICs and return to the FC/DC after having finished their tour. The components are transported via two routes (1) from CFC to FC and service center and (2) from CFC to spare market. The routes for transportation of components are fixed and components are carried by smaller trucks stationed at CFC chosen as per capacity and requirement.

The company gains revenue from secondary markets, service center, spare market and from the recycling agent. The costs borne by the company includes the cost of inspection, dismantling, fabrication, component fabrication, disposal and the various transportation costs. The first objective is to maximize the total profit. Second objective is to minimize the CO₂ emission by determining routes for the vehicles such that the total distance travelled is minimized and also choosing a vehicle of suitable capacity as per requirement.

2.1 Assumptions and Limitations

- Locations and capacities of CICs are known and fixed.
- The locations of DC, FC and disposal center are known and fixed.
- Demand of the fabricated products and components are known.
- There is no holding of inventory at any facility.
- The estimated emission rates of CO₂ for vehicles available are known.

2.2 Sets

Customer Zones with cardinality C indexed by c , Collection/Inspection centres with cardinality I indexed by i , Product Set with cardinality N indexed by n , Component Set with cardinality A indexed by a , Secondary markets with cardinality M indexed by m , Vehicle set for product flow with cardinality K indexed by k , Small truck set with cardinality T indexed by t , $V = I \cup \{0\}$ where 0 represents fabrication center node and $V' = I \cup \{0'\}$ where $0'$ represents dismantling center node.

2.3 Parameters

CIC_n is per unit cost of inspection of n th product at CIC, CDM_n is per unit cost of dismantling of n th product at DC, CFC_n is per unit cost of fabrication of n th product at FC, CCF_a is per unit cost of component fabrication of a th component at CFC, CDC_a is per unit cost of disposal of a th component at disposal center, TC is transportation cost (per km) of the returned products transported from CIC to DC, TS is unit transportation cost of the disassembled components transported from DC to disposal center, TI and TK are unit transportation costs of the fabricated components transported from CFC to FC and service center respectively and finally TCF_t is hiring cost of the t th truck for transporting components from CFC.

X_{cn} are units of n th product returned to customer zone c , Cap_i is capacity of the i th CIC, $SDEM_a$ is the demand of the a th component at the service center, $FDEM_a$ is demand of the a th component at FC, Y_{ci} is defined as, if customer zone c returns to i th CIC then it is 1, otherwise zero, Q_{na} is defined as, if n th product consists of a th component then it is 1, otherwise zero, α_n is the fraction of total unit of n th products returned by the customers ρ_n is fraction of total units of n th product to be transported for fabrication from i th CIC, μ_a is fraction of total units of a th component transported to recycling center from FC, γ_a and δ_a are fractions of total units of a th component transported from DC to CFC and recycling center respectively, φ_a and ω_a are fractions of total units of a th component transported from CFC to FC and service centre respectively, DEM_{mn} and $PREV_{mn}$ are the demand of per unit revenue generated by the n th product at m th secondary market respectively, $CREV_a$, $RREV_a$ and $SREV_a$ are per unit revenue generated by a th component at spare market, recycling center and service center respectively, d_{ij} and d'_{ij} are distances between nodes i and j of V and V' respectively, e_j and e'_j are real numbers used to avoid subtouring in the route from DC to CIC s to DC and FC to CICs to FC respectively, $Cmax$ is the maximum capacity of vehicle k (in kg), $Dmax$ is the maximum allowable route length (in km) for transportation of returned products, W_n is weight of n th product and Wt_a is weight of a th component, CO_2^k and CO_2^t are carbon emission per km of the k th and t th vehicle respectively, Route1 is the distance travelled from CFC to FC to service centre and Route2 is the distance travelled from CFC to spare market.

2.4 Decision Variables

RX_{ni} are units of n th product collected at i th CIC, Z_n and O_n are units of n th product transported from CICs to DC and FC respectively, B_{mn} is number of returned units of n th product transported from FC to m th secondary market, G_a , V_a and DIS_a are units of a th component transported to CFC, recycling center and disposal center, RK_a , RI_a and RF_a are units of a th component transported from CFC to service center, FC and spare market respectively, α_a and β_a take value 1 if the a th component transported to the service center and FC do not exceed the respective demands, else take value 0, x_{ijk} and x'_{ijk} are defined as, if vehicle k travels from node i to node j of set V and V' respectively then it is 1, otherwise zero aw_i and aw'_i are returned amounts at CIC to be sent to FC and DC respectively, H_t and L_t are defined as, if vehicle t is selected for Route 1 and Route 2 respectively then it is 1, otherwise zero, R and RS are total weight of the components transported via Route 1 and Route 2 respectively.

2.5 Constraints

$$\begin{aligned}
 RX_{ni} &= \alpha_n \sum_c X_{cn} \cdot Y_{ci} \quad \forall n, i \\
 \sum_i Y_{ci} &= 1 \quad \forall c \\
 \sum_n RX_{ni} &\leq Cap_i \quad \forall i
 \end{aligned}$$

These constraints determine the number of returned products collected at CICs.

$$\begin{aligned}
 O_n &= \sum_i \rho_n \cdot RX_{ni} \quad \forall n \\
 \sum_m B_{mn} &\leq O_n \quad \forall n \\
 B_{mn} &\leq DEM_{mn} \quad \forall m, n
 \end{aligned}$$

These constraints determine the total number of units of each product transported to fabrication center after initial inspection and after realizing the total demand at the secondary markets.

$$\begin{aligned}
 Z_n &= \sum_i (1 - \rho_n) \cdot RX_{ni} \quad \forall n \\
 G_a &= \sum_n \gamma_a \cdot Q_{na} \cdot Z_n \quad \forall a \\
 V_a &= \sum_n \delta_a \cdot Q_{na} \cdot Z_n + \mu_a RI_a \quad \forall a \\
 DIS_a &= \sum_n (1 - \gamma_a - \delta_a) \cdot Q_{na} \cdot Z_n + (1 - \mu_a) \cdot RI_a \quad \forall a
 \end{aligned}$$

These constraints determine the volume flow of products/components at various facilities.

$$\begin{aligned}
 RK_a &\leq \omega_a \cdot G_a \quad \forall a \\
 RI_a &\leq \phi_a \cdot G_a \quad \forall a \\
 RK_a &\geq SDEM_a \quad \forall a \\
 RI_a &\geq FDEM_a \quad \forall a \\
 RF_a &= (G_a - RK_a - RI_a) \quad \forall a
 \end{aligned}$$

These constraints determine the volume flow of components at service center, fabrication center and spare market while satisfying the demand of service and fabrication center.

$$\begin{aligned}
 aw_i &= \sum_n O_n \cdot W_n \quad \forall i = 1..I \\
 \sum_{j=0}^I \sum_{k=1}^K x_{ijk} &= 1 \quad \forall i \neq j; i \in V \\
 \sum_{i=0}^I \sum_{k=1}^K x_{ijk} &= 1 \quad \forall i \neq j; j \in V \\
 \sum_{i \in V} x_{ijk} &= \sum_{i \in V} x_{jik} \quad \forall k \in K; j = 1..I; j \neq i \\
 \sum_{j \in I} x_{0jk} &\leq 1 \quad \forall k \in K \\
 \sum_{i \in I} x_{i0k} &\leq 1 \quad \forall k \in K \\
 \sum_{i \in V} \sum_{j \in V} d_{ij} x_{ijk} &\leq Dmax \quad \forall k \in K; i \neq j \\
 \sum_{i \in I} a_i \sum_{j \in V} x_{ijk} &\leq Cmax \quad \forall k \in K; i \neq j \\
 e_j - e_i &\geq (I + 1) \sum_{k \in K} x_{ijk} - I \quad \forall j \neq i; i, j \in I \\
 aw_i' &= \sum_n Z_n W_n \quad \forall i = 1..I \\
 \sum_{j=0}^I \sum_{k=1}^K x_{ijk}' &= 1 \quad \forall i \neq j; i \in V' \\
 \sum_{i=0}^I \sum_{k=1}^K x_{ijk}' &= 1 \quad \forall i \neq j; j \in V' \\
 \sum_{i \in V'} x_{ijk}' &= \sum_{i \in V'} x_{jik}' \quad \forall k \in K; j = 1..I; j \neq i
 \end{aligned}$$

$$\begin{aligned}
 \sum_{j \in I} x_{0jk}' &\leq 1 \quad \forall k \in K \\
 \sum_{i \in I} x_{i0k}' &\leq 1 \quad \forall k \in K \\
 \sum_{i \in V'} \sum_{j \in V'} d_{ij}' x_{ijk}' &\leq D' \max \quad \forall k \in K; i \neq j \\
 \sum_{i \in I} a_i' \sum_{j \in V'} x_{ijk}' &\leq C' \max \quad \forall k \in K; i \neq j \\
 e_j' - e_i' &\geq (I + 1) \sum_{k \in K} x_{ijk}' - I \quad \forall j \neq i; i, j \in I
 \end{aligned}$$

These constraints ensure that minimum distance is covered between the depot (DM) and CICs while not exceeding the capacity of the vehicle and the maximum allowable distance.

$$\begin{aligned}
 R &= \sum_a (RK_a + RI_a) wt_a \\
 RS &= \sum_a RF_a \cdot wt_a \\
 R &\leq \sum_t cap_t \cdot H_t \\
 \sum_t H_t &= 1 \\
 RS &\leq \sum_t cap \cdot L_t \\
 \sum_t L_t &= 1
 \end{aligned}$$

These constraints determine the selection of vehicle for component flow. $B_{mn}, RX_{ni}, X_{cn}, Z_n, O_n, V_a, G_a, DIS_a, RK_a, RI_a, RF_a, aw_i, aw_i' \geq 0 \quad \forall m, n, a, i, c$ and integers $Y_{ci}, H_t, L_t, x_{ijk}, x_{ijk}', \alpha_a, \beta_a \in \{0, 1\} \quad \forall c, i, a, j, k, t$

These are the integer and binary restrictions.

2.6 Objectives

- **Total revenue generated:**

$$\begin{aligned}
 TRG &= \sum_m \sum_n PREV_{mn} \cdot B_{mn} \\
 &+ \sum_a (SREV_a \cdot RK_a + RREV_a \cdot V_a + FREV_a \cdot RI_a + CREV_a \cdot RF_a)
 \end{aligned}$$

The equation represents the revenue generated from (a) secondary markets; (b) the company’s service center; (c) recycling of components; (d) fabrication center and (e) spare markets.

• **Total transportation costs:**

$$TTC = \sum_a (TS \cdot DIS_a + TK \cdot RK_a + TI \cdot RI_a) + \sum_t TCF_t (H_t \cdot Route1 + L_t Route2) + TC \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d_{ij}' \cdot x_{ijk}') \right)$$

The equation represents the transportation cost incurred by transporting (f) transporting components from disassemble center to disposal; (g) transporting fabricated components from component fabrication to company’s service center; (h) transporting fabricated components from component fabrication to fabrication center; (i) transporting cost for fabricated components to service center, fabrication center and spare market; (j) transportation cost from FC to CIC to FC and DC to CIC to DC.

• **Total facility/processing cost:**

$$TFC = \sum_n \left(CIC_n \sum_i RX_{ni} + CDM_n Z_n + CFC_n \sum_m B_{mn} \right) + \sum_a (CCF_a \cdot G_a + DIS_a \cdot CDC_a)$$

The equation represents the cost parameters which include (k) cost of inspecting units; (l) cost of dismantling units; (m) cost of fabricating units; (n) cost of component fabricating and (n) cost of disposal.

$$\text{Max Profit} = \text{TRG} - \text{TTC} - \text{TFC}$$

Finally, the equation represents the objective which intends to maximize profit which is the difference between the total revenue generated and the various costs namely transportation cost and cost incurred at various facilities.

$$\text{Min Carbon Emission} = CO_2^k \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d_{ij}' \cdot x_{ijk}') \right) + \sum_t CO_2^t (H_t \cdot Route1 + L_t Route2)$$

The equation represents the carbon emission by the trucks used in transportation of product and components across facilities.

We use goal programming approach to find the solution of the problem.

Problem P1:

$$\begin{aligned}
 \text{Objectives: Maximize Profit} = & \sum_m \sum_n \text{PREV}_{mn} \cdot B_{mn} + \sum_a (\text{SREV}_a \cdot \text{RK}_a + \text{RREV}_a \cdot V_a \\
 & + \text{FREV}_a \cdot \text{RI}_a + \text{CREV}_a \cdot \text{RF}_a) - \sum_a (\text{TS} \cdot \text{DIS}_a + \text{TK} \cdot \text{RK}_a + \text{TI} \cdot \text{RI}_a) \\
 & - \sum_t \text{TCF}_t (H_t \cdot \text{Route1} + L_t \cdot \text{Route2}) - \text{TC} \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d'_{ij} \cdot x'_{ijk}) \right) \\
 & - \sum_n \left(\text{CIC}_n \sum_i \text{RX}_{ni} + \text{CDM}_n Z_n + \text{CFC}_n \sum_m B_{mn} \right) \\
 & - \sum_a (\text{CCF}_a \cdot G_a + \text{DIS}_a \cdot \text{CDC}_a) \tag{1}
 \end{aligned}$$

$$\begin{aligned}
 \text{Min Carbon Emission} = & \text{CO}_2^k \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d'_{ij} \cdot x'_{ijk}) \right) \\
 & + \sum_t \text{CO}_2^t (H_t \cdot \text{Route1} + L_t \cdot \text{Route2}) \tag{2}
 \end{aligned}$$

Subject to:

$$\text{RX}_{ni} = \alpha_n \sum_c X_{cn} \cdot Y_{ci} \quad \forall n, i \tag{3}$$

$$\sum_i Y_{ci} = 1 \quad \forall c \tag{4}$$

$$\sum_n \text{RX}_{ni} \leq \text{Cap}_i \quad \forall i \tag{5}$$

$$O_n = \sum_i \rho_n \cdot \text{RX}_{ni} \quad \forall n \tag{6}$$

$$\sum_m B_{mn} \leq O_n \quad \forall n \tag{7}$$

$$B_{mn} \leq \text{DEM}_{mn} \quad \forall m, n \tag{8}$$

$$Z_n = \sum_i (1 - \rho_n) \cdot \text{RX}_{ni} \quad \forall n \tag{9}$$

$$G_a = \sum_n \gamma_a \cdot Q_{na} \cdot Z_n \quad \forall a \tag{10}$$

$$V_a = \sum_n \delta_a \cdot Q_{na} \cdot Z_n + \mu_a \text{RI}_a \quad \forall a \tag{11}$$

$$DIS_a = \sum_n (1 - \gamma_a - \delta_a) \cdot Q_{na} \cdot Z_n + (1 - \mu_a) \cdot RI_a \quad \forall a \tag{12}$$

$$RK_a \leq \omega_a \cdot G_a \quad \forall a \tag{13}$$

$$RI_a \leq \phi_a \cdot G_a \quad \forall a \tag{14}$$

$$RK_a \geq SDEM_a \quad \forall a \tag{15}$$

$$RI_a \geq FDEM_a \quad \forall a \tag{16}$$

$$RF_a = (G_a - RK_a - RI_a) \quad \forall a \tag{17}$$

$$aw_i = \sum_n O_n \cdot W_n \quad \forall i = 1 \dots I \tag{18}$$

$$\sum_{j=0}^I \sum_{k=1}^K x_{ijk} = 1 \quad \forall i \neq j; i \in V \tag{19}$$

$$\sum_{i=0}^I \sum_{k=1}^K x_{ijk} = 1 \quad \forall i \neq j; j \in V \tag{20}$$

$$\sum_{i \in V} x_{ijk} = \sum_{i \in V} x_{jik} \quad \forall k \in K; j = 1 \dots I; j \neq i \tag{21}$$

$$\sum_{j \in I} x_{0jk} \leq 1 \quad \forall k \in K \tag{22}$$

$$\sum_{i \in I} x_{i0k} \leq 1 \quad \forall k \in K \tag{23}$$

$$\sum_{i \in V} \sum_{j \in V} d_{ij} x_{ijk} \leq D \max \quad \forall k \in K; i \neq j \tag{24}$$

$$\sum_{i \in I} a_i \sum_{j \in V} x_{ijk} \leq C \max \quad \forall k \in K; i \neq j \tag{25}$$

$$e_j - e_i \geq (I + 1) \sum_{k \in K} x_{ijk} - I \quad \forall j \neq i; i, j \in I \tag{26}$$

$$aw_i' = \sum_n Z_n W_n \quad \forall i = 1, \dots, I \tag{27}$$

$$\sum_{j=0}^I \sum_{k=1}^K x_{ijk}' = 1 \quad \forall i \neq j; i \in V' \tag{28}$$

$$\sum_{i=0}^I \sum_{k=1}^K x_{ijk}' = 1 \quad \forall i \neq j; j \in V' \tag{29}$$

$$\sum_{i \in V'} x_{ijk}' = \sum_{i \in V'} x_{jik}' \quad \forall k \in K; j = 1 \dots I; j \neq i \tag{30}$$

$$\sum_{j \in I} x_{0jk}' \leq 1 \quad \forall k \in K \tag{31}$$

$$\sum_{i \in I} x_{i0k}' \leq 1 \quad \forall k \in K \tag{32}$$

$$\sum_{i \in V'} \sum_{j \in V'} d_{ij}' x_{ijk}' \leq D' \max \quad \forall k \in K; i \neq j \tag{33}$$

$$\sum_{i \in I} a_i' \sum_{j \in V'} x_{ijk}' \leq C' \max \quad \forall k \in K; i \neq j \tag{34}$$

$$e_j' - e_i' \geq (I + 1) \sum_{k \in K} x_{ijk}' - I \quad \forall j \neq i; i, j \in I \tag{35}$$

$$R = \sum_a (RK_a + RI_a) wt_a \tag{36}$$

$$RS = \sum_a RF_a \cdot wt_a \tag{37}$$

$$R \leq \sum_t cap_t \cdot H_t \tag{38}$$

$$\sum_t H_t = 1 \tag{39}$$

$$RS \leq \sum_t cap_t \cdot L_t \tag{40}$$

$$\sum_t L_t = 1 \tag{41}$$

$$B_{mn}, RX_{ni}, X_{cn}, Z_n, O_n, V_a, G_a, DIS_a, RK_a, RI_a, RF_a, aw_i, aw_i' \geq 0 \quad \forall m, n, a, i, c \text{ and integers} \tag{42}$$

$$Y_{ci}, H_t, L_t, x_{ijk}, x'_{ijk}, \alpha_a, \beta_a \in \{0, 1\} \quad \forall c, i, a, j, k, t \tag{43}$$

2.7 Solution Methodology: Goal Programming

In goal programming, management sets goals and relative importance (weights) for different objectives. An optimal solution is defined as one that minimizes both positive and negative deviations from set goals simultaneously or minimizes the amount by which each goal can be violated. First the problem is solved using only rigid constraints and then goals of objectives are incorporated depending upon the priorities or relative importance of different objectives being well defined or not.

The problem can be solved in as follows:

Step 1: Solve two problems considering one objective at a time to obtain the target (aspiration) for each objective as follows:

Problem P2:

$$\begin{aligned} \text{Maximize Profit} = & \sum_m \sum_n PREV_{mn} \cdot B_{mn} \\ & + \sum_a (SREV_a \cdot RK_a + RREV_a \cdot V_a + FREV_a \cdot RI_a + CREV_a \cdot RF_a) \\ & - \sum_a (TS \cdot DIS_a + TK \cdot RK_a + TI \cdot RI_a) \\ & - \sum_t TCF_t (H_t \cdot Route1 + L_t Route2) \\ & - TC \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d'_{ij} \cdot x'_{ijk}) \right) \\ & - \sum_n \left(CIC_n \sum_i RX_{ni} + CDM_n Z_n + CFC_n \sum_m B_{mn} \right) \\ & - \sum_a (CCF_a \cdot G_a + DIS_a \cdot CDC_a) \end{aligned}$$

Subject to

Constraints (3)–(43)

Problem P3:

$$\begin{aligned} \text{Min Carbon Emission} = & CO_2^k \left(\sum_i \sum_j \sum_k (d_{ij} \cdot x_{ijk} + d'_{ij} \cdot x'_{ijk}) \right) \\ & + \sum_t CO_2^t (H_t \cdot Route1 + L_t Route2) \end{aligned}$$

Subject to

Constraints (3)–(43).

f_1^* and f_2^* are the optimal values obtained on solving (P2) and (P3) respectively.

Step 2A: Now convert the problem (P1) as vector optimization goal programming problem as follows:

Let $f(X)$ and b be the function and its goal respectively and η and ρ be the over and under achievement (negative and positive deviational) variables then the

choice of deviational variable in the goal objective functions which has to be minimized depend upon the following rule:

if $f(X) \leq b$, ρ is minimize under the constraints $f(X) + \eta - \rho = b$
 if $f(X) \geq b$, η is minimized under the constraints $f(X) + \eta - \rho = b$ and
 if $f(X) = b$, $\eta + \rho$ is minimized under the constraints $f(X) + \eta - \rho = b$.

Step 2B: The problem (P1) can be reformulated as vector goal programming problem incorporating optimal objective function values of the problem (P2) and (P3) as aspirations of their respective objectives using Step 2A.

Problem P4:

$$VMin g(\eta, \rho, X) = \begin{bmatrix} g_1(\eta, \rho, X) \\ g_2(\eta, \rho, X) \end{bmatrix}$$

Subject to:

$$RX_{ni} + \eta 1_{ni} - \rho 1_{ni} = \alpha_n \sum_c X_{cn} \cdot Y_{ci} \quad \forall n, i \tag{C1}$$

$$\sum_n RX_{ni} + \eta 2_{ni} - \rho 2_{ni} = Cap_i \quad \forall i \tag{C2}$$

$$O_n + \eta 3_n - \rho 3_n = \sum_i \rho_n \cdot RX_{ni} \quad \forall n \tag{C3}$$

$$\sum_m B_{mn} + \eta 4_n - \rho 4_n = O_n \quad \forall n \tag{C4}$$

$$B_{mn} + \eta 5_{mn} - \rho 5_{mn} = DEM_{mn} \quad \forall m, n \tag{C5}$$

$$Z_n + \eta 6_n - \rho 6_n = \sum_i (1 - \rho_n) \cdot RX_{ni} \quad \forall n \tag{C6}$$

$$G_a + \eta 7_a - \rho 7_a = \sum_n \gamma_a \cdot Q_{na} \cdot Z_n \quad \forall a \tag{C7}$$

$$V_a + \eta 8_a - \rho 8_a = \sum_n \delta_a \cdot Q_{na} \cdot Z_n + \mu_a RI_a \quad \forall a \tag{C8}$$

$$DIS_a + \eta 9_a - \rho 9_a = \sum_n (1 - \gamma_a - \delta_a) \cdot Q_{na} \cdot Z_n + (1 - \mu_a) \cdot RI_a \quad \forall a \tag{C9}$$

$$RK_a + \eta 10_a - \rho 10_a = \omega_a \cdot G_a \quad \forall a \tag{C10}$$

$$RI_a + \eta 11_a - \rho 11_a = \phi_a \cdot G_a \quad \forall a \tag{C11}$$

$$RK_a + \eta 12_a - \rho 12_a = SDEM_a \quad \forall a \tag{C12}$$

$$RI_a + \eta 13_a - \rho 13_a = FDEM_a \quad \forall a \tag{C13}$$

$$RF_a + \eta 14_a - \rho 14_a = (G_a - RK_a - RI_a) \quad \forall a \tag{C14}$$

$$aw_i + \eta 15_i - \rho 15_i = \sum_n O_n \cdot W_n \quad \forall i = 1 \dots I \tag{C15}$$

$$\sum_{i \in V} \sum_{j \in V} d_{ij} X_{ijk} + \eta 16_k - \rho 16_k = D \max \quad \forall k \in K; i \neq j \tag{C16}$$

$$\sum_{i \in I} a_i \sum_{j \in V} x_{ijk} + \eta 17_k - \rho 17_k = C \max \quad \forall k \in K; i \neq j \tag{C17}$$

$$aw'_i + \eta 18_i - \rho 18_i = \sum_n z_n \cdot W_n \quad \forall i = 1 \dots I \tag{C18}$$

$$\sum_{i \in V'} \sum_{j \in V'} d'_{ij} X'_{ijk} + \eta 19_k - \rho 19_k = D' \max \quad \forall k \in K; i \neq j \tag{C19}$$

$$\sum_{i \in I} a'_i \sum_{j \in V} x'_{ijk} + \eta 20_k - \rho 20_k = C' \max \quad \forall k \in K; i \neq j \tag{C20}$$

$$\sum_a ((RK_a + RI_a)wt_a + \eta 21_a - \rho 21_a) = R \tag{C21}$$

$$\sum_a (RF_a wt_a + \eta 22_a - \rho 22_a) = RS \tag{C22}$$

$$\sum_t cap_t \cdot H_t + \eta 23 - \rho 23 = R \tag{C23}$$

$$\sum_t cap_t \cdot L_t + \eta 24 - \rho 24 = RS \tag{C24}$$

$$(f1/f^*) + \eta 1 - \rho 1 = 1 \tag{C25}$$

$$(f2/f^*) + \eta 2 - \rho 2 = 1 \tag{C26}$$

$$\eta 1_{ni}, \rho 1_{ni}, \eta 2_{ni}, \rho 2_{ni}, \eta 3_n, \rho 3_n, \eta 4_n, \rho 4_n, \eta 5_{mn}, \rho 5_{mn}, \eta 6_n, \rho 6_n, \eta 7_a, \rho 7_a, \eta 8_a, \rho 8_a, \eta 9_a, \rho 9_a, \eta 10_a, \rho 10_a, \eta 11_a, \rho 11_a, \eta 12_a, \rho 12_a, \eta 13_a, \rho 13_a, \eta 14_a, \rho 14_a, \eta 15_i, \rho 15_i, \eta 16_k, \rho 16_k, \eta 17_k, \rho 17_k, \eta 18_i, \rho 18_i, \eta 19_k, \rho 19_k, \eta 20_k, \rho 20_k, \eta 21_a, \rho 21_a, \eta 22_t, \rho 22_t \geq 0 \tag{C27}$$

where $g_i(\eta, \rho, X)$ is goal deviation of the i th objective and $(f1/f^*)$ and $(f2/f^*)$ are used as normalized objectives functions to make these scale free.

We first solve the problem using rigid constraints only with the objective to minimize the non-achievement of those constraints by minimizing specific deviational variables using Step 2A.

Problem P5:

$$\begin{aligned} \text{MIN} = & \sum_n \sum_i (\eta_{1ni} + \rho_{1ni} + \eta_{2ni} + \rho_{2ni}) + \sum_n (\eta_{3n} + \rho_{3n} + \rho_{4n} + \eta_{6n} + \rho_{6n}) \\ & + \sum_n \sum_m \rho_{5mn} + \sum_a \left(\eta_{7a} + \rho_{7a} + \eta_{8a} + \rho_{8a} + \eta_{9a} + \rho_{9a} + \rho_{10a} + \rho_{11a} \right. \\ & \left. + \eta_{12a} + \eta_{13a} + \eta_{14a} + \rho_{14a} + \eta_{21a} + \rho_{21a} + \eta_{22a} + \rho_{22a} \right) \\ & + \sum_i (\eta_{15i} + \rho_{15i} + \eta_{18i} + \rho_{18i}) + \sum_k (\rho_{16k} + \rho_{17k} + \rho_{19k} + \rho_{20k}) + \eta_{23} + \eta_{24} \end{aligned}$$

Subject to:

(C1)–(C24), (C27).

Problem P6: The problem (P4) can be reformulated incorporating the optimal solution of the problem (P5) and assigning the weights λ_i ($i = 1, 2$) to the i th goal as follows:

$$\min = \lambda_1 \eta_1 + \lambda_2 \rho_2$$

Subject to:

$$\begin{aligned} (f1/f^*) + \eta_1 - \rho_1 &= 1 \\ (f2/f^*) + \eta_2 - \rho_2 &= 1 \\ \sum_n RX_{ni} + \eta_{2ni} &= Cap_i & \forall i \\ \sum_m B_{mn} + \eta_{4n} &= O_n & \forall n \\ B_{mn} + \eta_{5mn} &= DEM_{mn} & \forall m, n \\ RK_a + \eta_{10a} &= \omega_a \cdot G_a & \forall a \\ RI_a + \eta_{11a} &= \phi_a \cdot G_a & \forall a \\ RK_a - \rho_{12a} &= SDEM_a & \forall a \\ RI_a - \rho_{13a} &= FDEM_a & \forall a \\ \sum_{i \in V} \sum_{j \in V} d_{ij} X_{ijk} + \eta_{16k} &= Dmax & \forall k \in K; i \neq j \\ \sum_{i \in I} a_i \sum_{j \in V} x_{ijk} + \eta_{17k} &= Cmax & \forall k \in K; i \neq j \\ \sum_{i \in V'} \sum_{j \in V'} d'_{ij} X'_{ijk} + \eta_{19k} &= D'max & \forall k \in K; i \neq j \\ \sum_{i \in I} a'_i \sum_{j \in V} x'_{ijk} + \eta_{20k} &= C'max & \forall k \in K; i \neq j \\ \sum_t cap_t \cdot H_t - \rho_{23} &= R \\ \sum_t cap_t \cdot L_t - \rho_{24} &= RS \end{aligned}$$

and (3), (4), (9)–(12), (17)–(23), (26)–(32), (35)–(37), (39), (41)–(43).

Since both objectives are equally important, for the organization, equal weights can be assigned to both the objectives to obtain the solution of the problem (P6). However a trade-off between goals can be obtained to satisfy the management, a numerical illustration through case has been discussed in the following section giving equal weight to both the goals.

3 Case Study

The company ABC under study is an established manufacturing company which manufactures air conditioners (AC) and refrigerators. It is already handling all its operations in the reverse flow but needs to redesign the network to make it more profitable and environmentally sustainable. Hence while the main focus of the company has always been to maximise profit but now it also wants to minimise the carbon emission generated due to the transportation activities in their reverse logistic activities. The company aims to capitalise on the high demand of fabricated products in the secondary markets and of fabricated components at the service centre and spare market and wants to minimise the carbon emission by controlling the transportation activities. Since both objectives are equally important, for any supply chain system, equal weights are attached to each of the objectives. The resulting solution may provide a lead to the management for further analysis/improvement over the present scenario if required. A suitable tradeoff can be obtained based on various relative importances of the objectives provided by the management. The particular case of equal relative importance is presented in the following case illustration. In this study, there are 6 customer zones, 4 collection centers (capacity 400) and 2 secondary markets. 4 vehicles (of capacity 80,000 kg) for distribution of returned products, 4 vehicles of varying capacities for distribution of fabricated components.

The major components are: Compressor (A1), Condenser (A2), Orifice Tube/Expansion Valve (A3), Evaporator (A4), Accumulator/Drier (A5), Refrigerant (A6), Valve (A7), Compressor Clutch (A8), Refrigerant Oil (A9), Hose Assembly (A10), Switch (A11), Control Panel (A12), Metal and plastics (A13). Components A5, A7, A8, A9, A10 and A12 are only in AC and the rest are common to both.

Transportation cost (per km) of the returned products from FC/DC to CICs and back is Rs. 10, Unit transportation cost(per component) from DC to disposal center is Rs. 7, from CFC to FC is Rs. 8 and from CFC to service center is Rs. 7. Hiring + cost of the t th truck for transporting components from the CFC is Rs. 20, 30, 40, 25 respectively. The value of CIC is Rs. 70, Rs. 60, CDM is Rs. 70, Rs. 95, CFC is Rs. 110, Rs. 115 for AC and refrigerator and their weights are 35 kg and 50 kg respectively. We assume α as 0.4, 0.4 while ρ as 0.32 and 0.3 for AC and refrigerator respectively. The value of returned AC at different CIC is 60, 80, 75, 88 and 40 while for refrigerator it is 100, 72, 108, 110 and 64. The revenue, demand of AC from each secondary market is Rs. 5,000, 50 and Rs. 5,400, 44 while for refrigerator is Rs. 5,000, 60 and Rs. 5,800, 51

The maximum distance allowed for the route between FC/DC and CICs per trip is 100 km/120 km per trip. The CO₂ emission by the truck is 515.2 g/km. [11] The maximum distance allowed for route1 is 60 km and route2 is 40 km. The CO₂ emission for four trucks is 390.2 g/km, 400.2 g/km, 412.2 g/km and 450.2 g/km and their capacities are 1,000, 1,900, 3,000, 5,000 kg (Tables 1, 2).

The revenues from recycling agent for different components is Rs. 10, 20, 10, 20, 10, 20, 10, 10, 10, 30, 10, 20, 30 respectively. The revenues from fabrication

Table 1 Distance FC/DC and collection centers (CIC)

Node FC/DC	FC/DC	CC1	CC2	CC3	CC4
FC/DC	0/0	10/20	15/15	20/10	15/25
CC 1	10/20	0/0	20/10	25/15	10/20
CC2	15/15	20/10	0/0	15/25	20/20
CC3	20/10	25/15	15/25	0/0	25/15
CC4	15/25	10/20	20/20	25/15	0/0

Table 2 Costs (in Rs), demands, fractions and weights (in kg) for components

	CCF	CDC	FDEM	SDEM	γ	δ	μ	Φ	Ω	WT
A1	45	1	5	20	0.98	0	0.43	0.03	0.43	3
A2	45	2	8	30	0.94	0	0.54	0.04	0.54	1.94
A3	20	2	8	20	0.97	0	0.54	0.04	0.54	2.97
A4	30	3	12	25	0.94	0	0.66	0.06	0.66	1.94
A5	35	1	6	26	0.97	0	0.67	0.07	0.67	1.97
A6	0	4	0	0	0	0	0	0	0	2
A7	20	2	4	25	0.98	0	0.45	0.05	0.45	2.98
A8	25	2	7	26	0.98	0	0.38	0.08	0.38	2.98
A9	0	3	0	0	0	0	0	0	0	2
A10	0	1	0	0	0	1	0	0	0	1
A11	35	2	4	31	0.98	0	0.62	0.02	0.62	0.98
A12	35	1	6	35	0.99	0	0.47	0.07	0.47	1.99
A13	30	2	0	13	0.1	0	.54	0.04	0.54	1

center and service center for different components is Rs. 400, 200, 200, 190, 150, 0, 200, 250, 0, 0, 190, 100,100 respectively. The revenues from spare market for different components is Rs. 400, 200, 200, 190, 150, 0, 200, 250, 0, 0, 190, 100,100 respectively.

3.1 Result

Above data is entered to validate the proposed model. A LINGO code for generating the proposed mathematical models of the given data was developed and solved using LINGO11.0 for the given set of data, the optimal profit or cost of the network is obtained as Rs. 8,53,341 and minimum carbon emission is 91,192 g/km.

Allocation to Customer zones 1, 2, 3-CIC 1, Customer Zone 4, 5-CIC 2 and Customer zone 6-CIC 3 (Table 3).

Number of ACs sold at secondary markets-5, 44 and refrigerators-11, 51.

49/104 units of ACs and 62/145 refrigerators are transported to FC/DC.

Figure 2 show the vehicle routing between FC and CICs, Truck 2 is selected. Figure 3 show the vehicle routing between DC and CICs, Truck 4 is selected. In our study, since there is only one collection zone and therefore the quantity of the

Table 3 Optimal values of returned products at collection centres

	CIC1	CIC2	CIC3	CIC4
RX_{11}	0	153	0	0
RX_{21}	0	207	0	0
aw_i	4,823	4,823	4,823	4,823
aw_i'	10,898	10,898	10,898	10,898

Fig. 2 Route from fabrication center

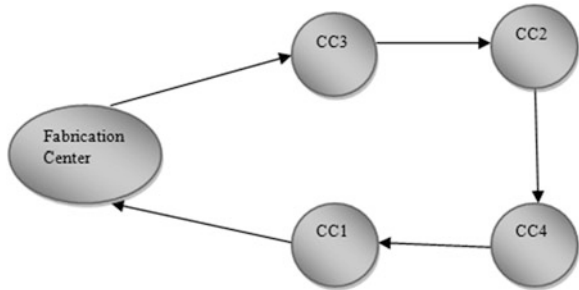
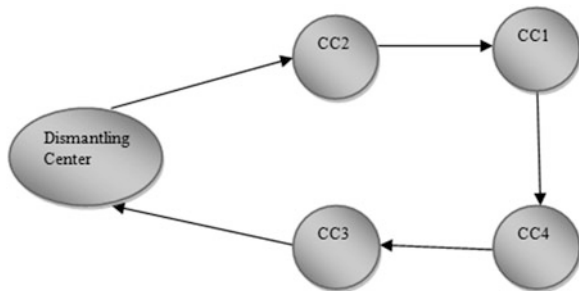


Fig. 3 Route from dismantling center



returned products does not exceed the available capacity of the trucks and hence products that are transported for fabrication/dismantling are carried by a single truck.

For Route 1, truck 3 is selected (weight of components 685 kg) and for Route 2, truck 1 is selected (weight of components 2,951 kg) (Table 4).

4 Conclusion

Due to Extended producer Responsibility, manufactures are facing the challenge of developing a reverse logistics network which meets environmental and economic goals. However, achieving desired optimality from both perspectives is

Table 4 Determines the optimal value of components at various facilities

	RC	CFC	Disposal	Spare market	Service center	Fabrication center
A1	2	244	7	219	20	5
A2	4	234	18	196	30	8
A3	4	241	11	213	20	8
A4	9	234	19	195	25	14
A5	4	101	5	69	26	6
A6	0	0	249	0	0	0
A7	1	102	4	73	25	4
A8	2	102	6	69	26	7
A9	0	0	104	0	0	0
A10	104	0	0	0	0	0
A11	2	244	6	209	31	4
A12	2	103	4	62	35	6
A13	1	24	224	10	13	1

unlikely in most situations and trade-offs are required. In this paper, a bi objective optimisation model is developed which provides the manufacturer an optimal strategy to maximise his profit while reducing the environmental impact by taking carbon emissions into account. Profit is generated by selling of the fabricated products and components. The mixed-integer linear programming model determines the amount of returned products to be fabricated and disassembled and the number of components to be fabricated, recycled or disposed. Further, the paper determines a strategic routing for distribution of the returned product and appropriate selection of vehicle for transporting the components to reduce CO₂ emission. The above model can be extended for multi period and multi collection zones.

References

1. Mutha, A., Pokharel, S.: Strategic network design for reverse logistics and remanufacturing using new and old product modules. *Comput. Ind. Eng.* **56**(1), 334–346 (2009)
2. Pishvae, M.S., Kianfar, K., Karimi, B.: Reverse logistics network design using simulated annealing. *Int. J. Adv. Manuf. Technol.* **47**(1–4), 269–281 (2010)
3. Kara, S.S., Onut, S.: A two step stochastic and robust programming approach to strategic level planning of a reverse supply network: the case of paper recycling. *Expert Syst. Appl.* **37**, 6129–6137 (2010)
4. Alumur, S.A., Nickel, S., Saldanha-da-Gama, F., Verter, V.: Multi-period reverse logistics network design. *Eur. J. Oper. Res.* **220**(1), 67–78 (2012)
5. Paksoy, T., Bektas, T., Özceylan, E.: Operational and environmental performance measures in a multi-product closed-loop supply chain. *Transp. Res. Part E* **47**(4), 532–546 (2011)
6. Elhedhli, S., Merrick, R.: Green supply chain network design to reduce carbon emissions. *Transp. Res. Part D Transp. Environ.* **17**(5), 370–379 (2012)
7. Sundarakani, B., De Souza, R., Goh, M., Wagner, S.M., Manikandan, S.: Modeling carbon footprints across the supply chain. *Int. J. Prod. Econ.* **128**(1), 43–50 (2010)

8. Kwon, Y.J.: Green supply chain network design to reduce carbon emissions. *Transp. Res. Part D Transp. Environ.* **17**, 370–379 (2012)
9. Kim, H., Yang, J., Lee, K.D.: Vehicle routing in reverse logistics for recycling end-of-life consumer electronic goods in South Korea. *Transp. Res. Part D Transp. Environ.* **14**(5), 291–299 (2009)
10. Aras, N., Aksen, D.: Tuğrul Tekin, M: Selective multi-depot vehicle routing problem with pricing. *Transp. Res. Part C Emerg. Technol.* **19**(5), 866–884 (2011)
11. Ramachandra, T.V.: Emissions from India's transport sector: statewise synthesis. *Atmos. Environ.* **43**(34), 5510–5517 (2009)

Multi Period Advertising Media Selection in a Segmented Market

Sugandha Aggarwal, Arshia Kaul, Anshu Gupta and P. C. Jha

Abstract A product passes through different life cycle phases once it comes in the market. During the launch phase it is promoted on mass level, second phase stresses on capturing maximum potential, and in later phases company emphasizes on retention of the product in the minds of the customers until its continuation in the market. The advertising budget and media mix used to advertise the product must be planned according to the current phase of the product. Thus it becomes imperative for a company to look at media planning problems in a dynamic manner over a planning period as the market conditions change and the product moves through its cycle with time. In this paper we have formulated a multi-period mathematical programming problem that dynamically computes the optimal number of insertions and allocates advertising budget in various media channels of a segmented market. The aim is to maximize the reach obtained through each media from all the segments in each time period under budgetary constraints and bounds on the decision variables. The reach in any period for a media is taken as a sum of current period reach and retention of the previous periods reach. Goal programming technique is used to solve the problem. A case study is presented to show the real life application of the model.

Keywords Media planning · Segmentation · Optimization · Multiple periods

S. Aggarwal (✉) · A. Kaul · P. C. Jha
Department of Operational Research, University of Delhi, Delhi, India
e-mail: sugandha_or@yahoo.com

A. Kaul
e-mail: arshia.kaul@gmail.com

P. C. Jha
e-mail: jhpc@yahoo.com

A. Gupta
SBPPSE, Dr B.R. Ambedkar University, Delhi, India
e-mail: anshu@aud.ac.in

1 Introduction

Companies allocate a huge percentage of their overall promotional budget on advertising. To efficiently use this budget a well thought media plan is developed. Different media exist in the market from television to websites and their usage changes with time. The selection of media mix depends on a number of factors like, specific objectives of the advertisement campaign, customer profiles of different market segments, market and competitive situation, and the product's stage in its life cycle. In the introductory stage, a large portion of advertising budget is allocated to induce trial. In the growth stage, however, advertising efforts are used primarily to capture maximum market potential. And when a product moves to the maturity stage, advertising is primarily a reminder to keep consumers aware of the product. In the decline stage of its life cycle, most of the advertising support is removed. The selection of media mix in a planning period varies according to the product's life cycle stage in the market. This calls for dynamic media planning.

A successful integrated marketing campaign also looks at to which media their target customers get exposed to, as it affect their decisions to buy the product; accordingly the media mix is selected. Advertising in a media is effective only if a significant number of targeted customers are exposed to that media. The impact of media can also vary in different segments and so is the choice of media mix. Distinct segments of the target population respond differently to different media and each medium plays a distinguishing role in advertising the product. Consider a shoe firm that comes up with a promotional campaign targeting female athletes for its new line of women's athletic shoes. In this case advertising would be targeted to the female athletes. The advertisements will be aired in those slots of television in which majority female athletes are expected to get exposed, or placed in that section of the newspaper which is read by most of the female athletes based on previous research. On the other hand if the marketer is advertising for an energy drink meant for school going children then they will target through children's most preferred channels on television. With an aim to capture maximum market potential, firms often look for maximizing the reach of their advertisements from each media in all the segments of its potential market.

Often marketers build media plans based on prior-period advertising response. Making suitable changes in the media mix according to present market scenario can yield significant increase in product performance. In the light of above considerations it becomes important for companies to study its present target market, consumer behavior towards different media, the stage of product cycle and accordingly select and place its advertising in media categories in different time periods. In this paper we propose a media planning optimization model for a segmented market to determine the optimal number of insertions to be given in different media categories maximizing the reach obtained from each media over the planning horizon. The dynamic nature of the target market is captured by dividing the planning horizon into multiple time periods. At the beginning of each period, the market is analyzed to understand the consumer behavior towards

different media in each segment, the changing advertisement cost, media popularity and its circulation. On the basis of this information, the multi-objective optimization problem is formulated with an objective of maximizing reach for all chosen media from all the segments with constraints on total budget allocated to the time period and lower and upper bound constraints on the number of insertions. The reach objective for each media is taken as a sum of current period reach and the retention factor on previous periods reach. The resultant integer programming problem is solved using goal programming approach with aspirations on reach to be attained from different media. A case study of a health insurance company is provided to illustrate and validate the proposed model.

The paper is organized as follows: [Sect. 2](#) gives a brief review of literature on media planning problems. [Section 3](#) discusses the model development and solution methodology. In [Sect. 4](#), we present a case study to illustrate applicability of the proposed model. [Section 5](#) concludes the paper.

2 Literature Review

In the literature lot of research work has been carried on media planning problems [9]. Charnes et al. [3] introduced a Goal Programming model for media selection to address problems associated with the critical advertising measurement of frequency and reach. A media model to get maximum reach in the press media while minimizing the cost was proposed by Kumar [6]. Moynihan et al. [8] contended that the mathematical requirements of the multi criteria decision making model for media selection force the media planner to create an artificial structuring of the media selection criteria. Balakrishnan and Hall [1] developed a computational procedure to calculate the time between the insertions within a pulse in order to maximize the minimum effectiveness level at any point in time with budget constraint. A mathematical programming model that gives the way a given product should be advertised on television in order to achieve the highest rating under specific restrictions was given by Mihiotis and Tsakiris [7]. Bhattacharya [2] developed a model to decide the number of advertisements in different advertising media and the optimal allocation of the budget assigned to the different media. Jha et al. [5] formulated a media planning problem for allocating the budget in multiple media that are found suitable for the advertising of multiple products considering marketing segmentation aspect of advertising with an objective of maximizing the reach. Royo et al. [10] proposed and analyzed a model to optimize the advertising investment for several products on a planning horizon that enables to compute the optimal advertising budget and its optimal allocation. However, the existing research ignores the dynamism involved in media planning. With the product growth in the market and change in the lifecycle stage of the products, impact of media vehicles and consumer behavior towards them also varies. To yield maximum reach from all chosen media it is important to base the selection of advertisements in various media on the current market conditions and

media performance. In this paper we have proposed a multi-objective programming problem to maximize the reach obtained in each media from all the segments and find the number of insertions to be given in different media by dividing the planning horizon in multiple periods. In the beginning of each time period the market is analyzed on the basis of which media planning is done with an aim to maximize reach from each media in all the segments under budgetary and system constraints.

3 Model Development

3.1 Notations

i	Index representing segments ($i = 1, 2, \dots, N$)
j	Index representing medium of advertisement ($j = 1, 2, \dots, M$)
k_j	Media option of j th media ($k_j = 1, 2, \dots, K$)
l_j	Slot in j th media ($l_j = 1, 2, \dots, L$)
r	Customer profile criteria under consideration for the potential population
T_q	Starting point of q th time period ($q = 1, 2, \dots, D$)
T	Length of the total planning horizon
$a_{ijk_l, l, q}$	Reach to the target audience for one advertisement in i th segment, j th medium, k_j th media option, l_j th slot/insertion in q th time period
$C_{ijk_l, l, q}$	Average number of readers/viewers/listeners in i th segment, j th medium, k_j th media option, l_j th slot/insertion in q th time period
$c_{ijk_l, l, q}$	Cost of inserting one advertisement in i th segment, j th medium, k_j th media option, l_j th slot/insertion in q th time period
$v_{ijk_l, l, q}$	Minimum number of advertisements in different positions in i th segment, j th medium, k_j th media option, l_j th slot/insertion in q th time period
$u_{ijk_l, l, q}$	Maximum number of advertisements in different positions in i th segment, j th medium, k_j th media option, l_j th slot/insertion in q th time period
$x_{ijk_l, l, q}$	Decision variable corresponding to number of advertisements in i th segment's j th medium, k_j th media option, l_j th slot/insertion in q th time period
$p_{irjk_l, l, q}$	Percentage of people in i th segment who fall under r th criteria and read/view j th medium, k_j th media option, l_j th slot/insertion in q th time period
$\alpha_{ijk_l, l, q}$	Retention factor of $(q-1)$ th period advertisement given in j th medium, k_j th media option, l_j th slot of in q th time period
W_{rj}	Weight corresponding to r th criteria and j th medium
A_q	Total advertisement budget to be allotted to different media in q th time period.

3.2 Problem Formulation

Firms market their products by means of advertising in various media such as newspaper, television, radio, websites, etc. The advertising media mix varies in different segments of the market in different time periods. Here we divide the total planning horizon T for which advertisement planning is to be carried into D time periods. The model proposed here will enable firms to allocate the total advertising budget A_q , in q th time period to M different media catering to the N market segments such that the reach is maximized from each media. The problem is formulated under the budgetary constraints. Upper bound and lower bound are imposed on number of insertions in each media. In the beginning of each time period the market is analyzed to determine the changing impact of media on segments, advertisement cost, its circulation, retention factor of previous period advertisements, etc. on the basis of which the media planning is done for the sequential time periods of the planning horizon. The model is based on the following assumptions

1. The market for the product is divided into N distinct segments.
2. The planning horizon is divided into multiple periods (say D periods, where q th time period starts at time T_q and ends at T_{q+1}).
3. Medias are independent of each other, thus one media has no impact on others.
4. Consumer behavior of segments is independent of each other and advertising in one segment has no impact on others.
5. At the beginning of the planning horizon, retention factor is zero.
6. For a particular time period, the retention factor is taken as a proportion of reach from previous periods advertising.

The multi-objective optimization model for the q th time period for determining the optimal number of insertions in different media that maximizes the reach obtained from each media from all the segments is formulated as follows

$$\text{Maximize } Z_q = \begin{bmatrix} Z_{1q} = \sum_{i=1}^N \sum_{k_1=1}^K \sum_{l_1=1}^L a_{i1k_1l_1q} * x_{i1k_1l_1q} \\ Z_{2q} = \sum_{i=1}^N \sum_{k_2=1}^K \sum_{l_2=1}^L a_{i2k_2l_2q} * x_{i2k_2l_2q} \\ \vdots \\ Z_{Mq} = \sum_{i=1}^N \sum_{k_M=1}^K \sum_{l_M=1}^L a_{iMk_Ml_Mq} * x_{iMk_Ml_Mq} \end{bmatrix}$$

$$\begin{aligned}
 \text{subject to } & \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L c_{ijk_lj,q} x_{ijk_lj,q} \leq A_q \\
 & x_{ijk_lj,q} \leq u_{ijk_lj,q} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad \quad \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & x_{ijk_lj,q} \geq v_{ijk_lj,q} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad \quad \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & \text{and } x_{ijk_lj,q} \text{ takes integral values}
 \end{aligned} \tag{P1}$$

where $a_{ijk_lj,q} = a_{ijk_lj,q} + \alpha_{ijk_lj,q} a_{ijk_lj,q-1}$ and $a_{ijk_lj,q} = \left\{ \sum_{r=1}^R w_{rj} P_{irjk_lj,q} \right\} C_{ijk_lj,q}$

Each objective in the multi-objective function maximizes the total reach in current time period q for a media in all the segments. First constraint ensures that the total amount of budget allocated to the medias in different segments in q th time period does not exceed the total amount of advertising budget available for this period i.e. A_q . Second and third constraint ensures that the number of advertisements in each media does not exceed its upper and lower bounds set by the firm. Last constraint ensures that the decision variable takes integral value for the allocated number of advertisements (can be taken greater than equal to zero when lower bound is not specified).

The multi-objective problem formulated here can be solved either by weighted sum approach where weights are assigned according to the relative importance of media or by assigning priorities to the different objectives, i.e. prioritising the goal for more preferable media and using lexicographic approach. Here we have used weighted sum approach to solve the problem.

Management might also wish to target each media so as to obtain maximum reach from all medias collectively under the limited resources. In such a situation one can fix high aspiration goals for those media. To fix goals/aspirations for the media, optimization model (P1) can be solved considering single objective corresponding to each media under the model constraints at a particular time period, which determines the maximum achievable reach for each media. These reach aspiration are expected to vary in different time periods and will depend on the current marketing conditions. For a particular time period q , once we solve and fix the goal or aspirations $Z_{1q}^*, Z_{2q}^*, \dots, Z_{Mq}^*$ to be obtained from j th media, the resulting problem can be formulated as follows

$$\begin{aligned}
 \text{Maximize } Z_q = & \begin{bmatrix} Z_{1q} = \sum_{i=1}^N \sum_{k_1=1}^K \sum_{l_1=1}^L a_{i1k_1l_1q} * x_{i1k_1l_1q} \\ Z_{2q} = \sum_{i=1}^N \sum_{k_2=1}^K \sum_{l_2=1}^L a_{i2k_2l_2q} * x_{i2k_2l_2q} \\ \vdots \\ Z_{Mq} = \sum_{i=1}^N \sum_{k_M=1}^K \sum_{l_M=1}^L a_{iMk_Ml_Mq} * x_{iMk_Ml_Mq} \end{bmatrix} \\
 \text{subject to } & \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L c_{ijk_jl_jq} x_{ijk_jl_jq} \leq A_q \tag{P2} \\
 & x_{ijk_jl_jq} \leq u_{ijk_jl_jq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad \quad \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & x_{ijk_jl_jq} \geq v_{ijk_jl_jq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad \quad \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & Z_{jq} \geq Z_{jq}^* \quad \forall j = 1, 2, \dots, M \\
 & \text{and } x_{ijk_jl_jq} \text{ takes integral values}
 \end{aligned}$$

Since high goal/aspiration of each objective has been incorporated into the constraint set of the original problem (P1), the problem (P2) becomes infeasible due to limitation on budget. The problem (P2) is solved to obtain compromised solution using goal programming approach.

3.3 Solution Methodology: Goal Programming

The goal programming formulation of the optimization model (P2) can be formulated using Appendix A as follows

$$\begin{aligned}
 \text{Vector Minimize } g(\eta, \rho, X) = & \begin{bmatrix} g_1(\eta, \rho, X) \\ g_2(\eta, \rho, X) \\ \cdot \\ \cdot \\ \cdot \\ g_m(\eta, \rho, X) \end{bmatrix} \\
 \text{subject to } & \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L c_{ijk_jl_jq} x_{ijk_jl_jq} + \eta_1 - \rho_1 = A_q \\
 & x_{ijk_jl_jq} + \eta_{ijk_jl_jq} - \rho_{ijk_jl_jq} = u_{ijk_jl_jq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad \quad \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L
 \end{aligned}$$

$$\begin{aligned}
 x_{ijk_ljq} + \eta'_{ijk_ljq} - \rho'_{ijk_ljq} &= v_{ijk_ljq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 Z_{jq} + \eta_{jq} - \rho_{jq} &= Z_{jq}^* \quad \forall j = 1, 2, \dots, M \\
 \text{and } x_{ijk_ljq} &\text{ takes integral values} \\
 \eta_1, \rho_1 &\geq 0 \\
 \eta_{ijk_ljq}, \rho_{ijk_ljq}, \eta'_{ijk_ljq}, \rho'_{ijk_ljq} &\geq 0 \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 \eta_{jq}, \rho_{jq} &\geq 0 \quad \forall j = 1, 2, \dots, M
 \end{aligned} \tag{P3}$$

where η and ρ 's are the positive & negative deviation variable of the goals for their respective objective/constraint function.

In weighted sum approach, the rigid constraints are first satisfied by minimizing the deviational variables corresponding to them and then, goals on objective function are incorporated in the second stage. In first stage all rigid constraints are weighted equally and in stage II, the objective is expressed as the weighted sum of the deviational variables to be minimized corresponding to the objective function's goal. When all goals have equal relative importance, weight of each objective can be taken to as one. Here we have assumed differential weights for the reach objectives for different media. The problem (P3) can be solved in two stages as follows

Stage 1

$$\begin{aligned}
 \text{Minimize} \quad & \rho_1 + \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L \rho_{ijk_ljq} + \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L \eta'_{ijk_ljq} \\
 \text{subject to} \quad & \sum_{i=1}^N \sum_{j=1}^M \sum_{k_j=1}^K \sum_{l_j=1}^L c_{ijk_ljq} x_{ijk_ljq} + \eta_1 - \rho_1 = A_q \\
 & x_{ijk_ljq} + \eta_{ijk_ljq} - \rho_{ijk_ljq} = u_{ijk_ljq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & x_{ijk_ljq} + \eta'_{ijk_ljq} - \rho'_{ijk_ljq} = v_{ijk_ljq} \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & \eta_1, \rho_1 \geq 0 \\
 & \eta_{ijk_ljq}, \rho_{ijk_ljq}, \eta'_{ijk_ljq}, \rho'_{ijk_ljq} \geq 0 \quad \forall i = 1, 2, \dots, N; j = 1, 2, \dots, M; \\
 & \quad k_j = 1, 2, \dots, K; l_j = 1, 2, \dots, L \\
 & \text{and } x_{ijk_ljq} \text{ takes integral values}
 \end{aligned} \tag{P4}$$

Table 1.1 Solution for Newspaper

City	Channel 1		Channel 2	
	PT	OT	PT	OT
Time period 1				
City 1	20	36	12	24
City 2	18	28	15	34
City 3	15	25	12	28
City 4	18	36	15	30
Time period 2				
City 1	3	33	4	26
City 2	20	30	12	32
City 3	15	28	10	30
City 4	15	34	2	32
Time period 3				
City 1	2	36	2	26
City 2	12	32	12	34
City 3	14	28	2	26
City 4	1	32	2	32
Time period 4				
City 1	1	38	1	26
City 2	20	30	15	36
City 3	18	28	8	24
City 4	1	26	1	32

advertisement on front page (FP), last page and other pages (OP) etc. Similarly, two different time slots (prime time (PT) and other time (OT)) for each of television and radio channels has been decided based on the importance and number of viewers/listeners. The target market is divided in accordance with the income levels of population in three groups: low (R1), middle (R2) and high (R3) level income groups. The advertisement planning horizon (one year) is divided into four sub-periods. In the beginning of each quarter, market is studied to analyze the behavior of population towards different media in each city. Data for circulation of the media options, advertisement cost, percentage profile matrix, retention factor in different media is gathered. The aim is to advertise insurance company’s product dynamically over a planning horizon, in different media so as to maximize reach in each media in all segments of target market within its specified constraints.

The total advertising budget is taken to be ₹20 crore which is divided into four time periods as ₹7 crore, ₹5 crore, ₹4 crore, and ₹4 crore respectively. Circulation figures for media are given in Tables 3.1, 4.1 and 5.1 in Appendix C. It is the average number of readers (for newspapers)/viewers (for television)/listeners (for radio). Data for per insertion advertisement cost in different media is given in Tables 3.2, 4.2, and 5.2 of Appendix C. In case of newspaper, advertisement cost

Table 1.2 Solution for Radio

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time period 1					
65	4	50	10	24	10
68	6	20	6	14	5
72	6	34	10	16	10
48	8	19	8	14	8
Time period 2					
18	6	17	12	22	12
72	4	22	8	35	3
14	5	16	10	19	12
26	6	14	6	16	7
Time period 3					
16	5	18	10	18	8
13	7	22	10	13	4
15	8	12	6	19	12
20	6	14	8	14	7
Time period 4					
12	8	20	15	12	7
9	5	18	12	12	4
19	12	10	5	18	10
16	6	13	7	13	6

Table 1.3 Solution for Television

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time period 1					
120	68	110	20	90	32
118	54	90	42	80	30
124	52	78	24	90	18
130	65	95	20	75	12
Time period 2					
118	64	112	45	86	28
24	30	94	12	75	14
120	10	38	10	32	15
128	8	100	16	78	10
Time period 3					
124	61	110	40	90	33
108	47	98	12	75	18
127	12	90	8	90	19
118	16	98	21	90	8
Time period 4					
130	65	115	38	88	35
89	47	95	18	80	17
129	16	34	11	40	12
22	8	110	25	85	18

Table 2 Reach obtained from media at different time periods

Media	Time period 1		Time period 2		Time period 3		Time period 4	
	Aspired	Achieved	Aspired	Achieved	Aspired	Achieved	Aspired	Achieved
Newspaper	22501770	22501773	24804830	22556566	27776480	22746647	26964420	23281555
Television	55678550	44786987	43191340	30166631	32056380	18480446	28219430	15792821
Radio	23006180	22917331	14034790	12407175	24127410	23659174	15930720	14097089
Total	101186500	90206091	82030960	65130372	83960270	64886267	71114570	53171465

is per column centimetre for a 100 column cm advertisement, for television and radio it is for 10 second spot. Upper and lower bound on the number of advertisements in various media is given in Tables 3.4, 4.3, and 5.3 of Appendix C. The data for customer profile given in Tables 3.5, 4.5, and 5.5 of Appendix C represents the proportion of people in different criteria (i.e. lower, middle and higher income group) who are reading, viewing or listening the advertisement in the media. These profile matrices are based on a random sample of size 200 across all the cities and time periods. The data for retention factor has been given in Tables 3.3, 4.4, and 5.4 of Appendix C. The assignment of weights for a particular customer profile criteria in a media is based on the expert's judgment which is given in Table 6 of Appendix C. The optimization model (P1) is programmed as a single objective problem to obtain reach aspirations for each media. The optimal values so obtained are then set as aspiration level to be achieved for reach corresponding to each media and problem (P2) is solved. This gives an infeasible solution; Goal Programming technique is used to obtain a compromised solution to the problem. First we solve for stage 1 (P4) with rigid constraints and compute stage one solution and then in stage 2 (P5), the deviational variables that become zero in stage 1 are removed and the deviations corresponding to goals are minimized. Weights are assigned (Table 7, Appendix C) to media on the basis of management decisions considering different tradeoff solutions, consumer response to different media and media effectiveness in the market. This process is repeated for time period 2, 3, and 4. All optimizations models are programmed on LINGO software [11] to obtain optimal solutions. Optimal solution is guaranteed as the proposed optimization model is a convex programming problem. The solution obtained for number of advertisements to be given in different media at different time periods is given in Tables 1.1, 1.2 and 1.3 below.

The reach aspired and achieved from different media in different time periods on solving the problem is displayed in Table 2 below.

In time period 1, it can be seen that the total reach obtained is higher than any other time periods, this is because the highest budget allocated to this period. It matches with the fact that it is the initial stage of the product and the firm wants to advertise at large scale and reach maximum people in this period. In every period, there is a compromise from the aspired reach, except for newspaper media in period 1 and total compromise is increasing from period 1 to 4. Reach obtained through television is highest in all the periods possibly due to the visual impact of television. However compromise percentage for television is also highest in every period and from period 1 to period 4, the percentage by which it is compromised keeps on increasing. Higher cost of insertion is the most possible reason for higher compromise. Lot of further analysis of the case is done by the author but is not discussed here due to limitation on pages of the manuscript.

5 Conclusion

In this paper, we have formulated and solved an optimization model which can be applied for planning and scheduling of appropriate media for the advertisement of a product. The model allocates the advertising efforts to each of the media dynamically in the market segments by dividing the planning horizon into small planning periods. The optimal number of advertisements to be given in media in different time periods changes based on the behaviour of consumer towards the media and the position of the media in the market. The objective is to maximize the total reach which comes from the current period reach and the proportional retention of the previous periods reach in each media. This allows us to capture clearer picture of the market by planning according to the current behaviour of the market. The model is formulated under budget and lower and upper bounds on frequency constraint in each media. The problem is solved using goal programming approach. A case is given to illustrate the applicability of the proposed optimization model. The paper offers a lot of scope for further research such as explicit consideration for multiple products, interaction between the advertising of different products of the firm, and fuzzy environment.

Appendix A

Let $f(X)$ and b be the function and its goal respectively and η and ρ be the over and under achievement (negative and positive deviational) variables then the choice of deviational variable in the goal objective functions which has to be minimized depend upon the following rule:

if $f(X) \leq b$, ρ is minimize under the constraints $f(X) + \eta - \rho = b$
 if $f(X) \geq b$, η is minimized under the constraints $f(X) + \eta - \rho = b$ and
 if $f(X) = b$, $\eta + \rho$ is minimized under the constraints $f(X) + \eta - \rho = b$

Appendix B

Definition 1[4]: A solution x^0 is said to be efficient solution if $x^0 \in S$ and there exist no other feasible point x such that $F(x) \geq F(x^0)$ and $F(x) \neq F(x^0)$.

Definition 2[4]: A solution x^0 is said to be a properly efficient solution if it is efficient and if there exist a scalar $M > 0$ such that, for each i , we have $\frac{f_i(x) - f_i(x^0)}{f_j(x^0) - f_j(x)} \leq M$. For some j such that $f_j(x) < f_j(x^0)$ whenever $x \in X$ and $f_i(x) > f_i(x^0)$

Lemma 1:Through definition 1 and 2, optimal solution of the problem (P5) will be properly efficient solution to the problem (P3).

Appendix C

Table 3.1 Circulation in Newspapers ('0000)

City	Newspaper 1	Newspaper 2
Time Period 1		
City 1	56	35
City 2	86	67
City 3	75	49
City 4	89	72
Time Period 2		
City 1	54	36
City 2	86	65
City 3	76	49
City 4	88	74
Time Period 3		
City 1	55	38
City 2	90	65
City 3	76	48
City 4	87	74
Time Period 4		
City 1	57	38
City 2	89	67
City 3	77	45
City 4	84	76

Table 3.2 Advertisement Cost in Newspapers('000)

Newspaper 1		Newspaper 2	
FP	OP	FP	OP
Time Period 1			
55	25	37	20
60	29	52	22
55	28	42	21
62	30	53	29
Time Period 2			
53	22	37	24
62	34	48	18
55	35	42	22
58	25	56	32
Time Period 3			
55	24	40	25
65	35	48	20
55	29	41	21
60	32	57	32
Time Period 4			
56	22	44	28
66	36	45	23
56	30	40	23
58	32	60	34

Table 3.3 Retention Factor in Newspapers (%)

Newspaper 1		Newspaper 2	
FP	OP	FP	OP
Time Period 1			
–	–	–	–
–	–	–	–
–	–	–	–
–	–	–	–
Time Period 2			
8	5	5	4
6	5	3	2
7	6	5	2
8	5	6	3
Time Period 3			
5	7	6	8
7	6	4	4
7	6	6	4
8	7	4	5
Time Period 4			
5	7	4	6
4	8	9	5
9	8	4	4
6	7	3	8

Table 3.4 Upper and Lower Bounds on Advertisement in Newspapers

City	Newspaper 1		Newspaper 2	
	FP	OP	FP	OP
Time Period 1				
City 1	[3,20]	[6,36]	[3,12]	[10,24]
City 2	[3,18]	[8,28]	[3,15]	[10,34]
City 3	[3,15]	[8,25]	[3,12]	[12,28]
City 4	[3,18]	[12,36]	[3,15]	[10,30]
Time Period 2				
City 1	[3,18]	[7,33]	[4,12]	[10,26]
City 2	[2,20]	[6,30]	[2,12]	[12,32]
City 3	[4,15]	[8,28]	[2,10]	[13,30]
City 4	[2,15]	[14,34]	[2,18]	[8,32]
Time Period 3				
City 1	[2,18]	[7,36]	[2,12]	[8,26]
City 2	[1,20]	[5,32]	[3,12]	[12,34]
City 3	[2,15]	[10,28]	[2,10]	[12,26]
City 4	[1,130]	[12,32]	[2,16]	[7,32]
Time Period 4				
City 1	[1,20]	[8,38]	[1,12]	[6,26]
City 2	[1,20]	[4,30]	[1,15]	[10,36]
City 3	[1,18]	[8,28]	[1,8]	[10,24]
City 4	[1,10]	[10,26]	[1,19]	[7,32]

Table 3.5 Readership Profile Matrix for Newspapers

Newspaper 1			Newspaper 2		
R1	R2	R3	R1	R2	R3
Time Period 1					
.14	.36	.21	.19	.36	.27
.17	.42	.25	.23	.42	.32
.16	.40	.24	.22	.40	.30
.13	.33	.20	.18	.33	.25
Time Period 2					
.17	.40	.25	.23	.43	.27
.16	.38	.24	.22	.42	.26
.20	.48	.30	.27	.52	.32
.14	.34	.21	.19	.37	.23
Time Period 3					
.21	.49	.33	.28	.51	.39
.19	.44	.29	.25	.45	.35
.19	.44	.30	.25	.46	.35
.16	.36	.24	.20	.38	.29
Time Period 4					
.25	.45	.32	.27	.5	.39
.21	.38	.27	.23	.42	.33
.26	.46	.33	.28	.51	.40
.17	.30	.21	.18	.33	.26

Table 4.1 Average Viewership in Television ('0000)

City	Channel 1		Channel 2		Channel 3	
	PT	OT	PT	OT	PT	OT
Time Period 1						
City 1	160	90	120	92	60	25
City 2	175	85	90	32	134	37
City 3	178	85	87	69	125	79
City 4	98	56	150	78	165	84
Time Period 2						
City 1	140	78	160	67	68	41
City 2	180	89	80	32	120	52
City 3	170	90	101	69	108	73
City 4	88	60	160	79	145	88
Time Period 3						
City 1	142	84	156	78	95	45
City 2	167	75	67	27	125	47
City 3	164	94	120	72	98	48
City 4	100	64	154	72	162	90
Time Period 4						
City 1	134	80	86	56	120	75
City 2	170	90	98	46	140	45
City 3	78	42	145	57	97	54
City 4	85	45	128	67	178	96

Table 4.2 Advertisement Cost in Television(‘000)

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time Period 1					
105	50	101	43	65	34
106	48	82	45	100	56
120	56	78	38	110	49
69	34	110	52	115	67
Time Period 2					
90	40	120	30	80	40
115	55	75	40	80	65
120	60	80	38	105	30
56	35	120	55	110	70
Time Period 3					
92	40	118	28	88	41
110	40	70	35	81	52
115	60	100	40	85	30
56	35	115	55	120	70
Time Period 4					
124	60	76	25	98	40
136	67	80	30	110	56
89	52	128	70	100	59
86	45	120	70	142	78

Table 4.3 Upper and Lower bounds on advertisement in Television

City	Channel 1		Channel 2		Channel 3	
	PT	OT	PT	OT	PT	OT
Time Period 1						
City 1	[16,65]	[4,36]	[20,50]	[10,24]	[24,32]	[10,20]
City 2	[12,68]	[6,40]	[20,40]	[6,30]	[14,52]	[5,25]
City 3	[14,72]	[6,50]	[18,48]	[10,25]	[16,60]	[10,30]
City 4	[23,48]	[8,48]	[19,62]	[8,35]	[14,75]	[8,40]
Time Period 2						
City 1	[18,60]	[6,33]	[17,54]	[12,20]	[22,35]	[12,24]
City 2	[12,72]	[4,38]	[22,35]	[8,30]	[15,48]	[3,30]
City 3	[14,68]	[5,52]	[16,50]	[10,25]	[19,58]	[12,26]
City 4	[26,40]	[6,28]	[14,65]	[6,40]	[16,70]	[7,42]
Time Period 3						
City 1	[16,54]	[5,34]	[18,58]	[10,22]	[18,39]	[8,26]
City 2	[13,70]	[7,35]	[22,32]	[10,26]	[13,50]	[4,24]
City 3	[15,65]	[8,54]	[12,58]	[6,30]	[19,50]	[12,24]
City 4	[20,45]	[6,30]	[14,63]	[8,38]	[14,72]	[7,45]
Time Period 4						
City 1	[12,50]	[8,30]	[20,38]	[15,20]	[12,69]	[7,35]
City 2	[9,65]	[5,38]	[18,35]	[12,28]	[12,55]	[4,25]
City 3	[19,35]	[12,24]	[10,66]	[5,35]	[18,50]	[10,25]
City 4	[16,49]	[6,25]	[13,55]	[7,30]	[13,75]	[6,45]

Table 4.4 Retention Factor in Television (%)

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time Period 1					
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
Time Period 2					
5	5	8	8	3	1
7	4	5	5	5	3
6	4	5	3	6	8
5	3	8	5	7	5
Time Period 3					
4	3	7	4	7	5
9	5	6	3	8	2
5	3	1	4	3	3
5	2	7	3	3	4
Time Period 4					
6	4	4	6	3	4
4	7	5	3	6	8
4	4	6	6	5	6
6	4	5	3	8	5

Table 4.5 Viewership Profile Matrix for Television

City	Channel 1						Channel 2						Channel 3					
	R1		R2		R3		R1		R2		R3		R1		R2		R3	
	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT
Time period 1																		
City 1	0.23	0.05	0.25	0.04	0.30	0.05	0.23	0.08	0.28	0.05	0.20	0.05	0.19	0.07	0.23	0.07	0.21	0.09
City 2	0.18	0.11	0.23	0.09	0.25	0.09	0.22	0.10	0.20	0.09	0.18	0.08	0.20	0.08	0.15	0.11	0.18	0.07
City 3	0.19	0.08	0.27	0.11	0.28	0.03	0.19	0.07	0.25	0.06	0.21	0.06	0.18	0.06	0.16	0.07	0.16	0.06
City 4	0.14	0.06	0.19	0.08	0.26	0.09	0.14	0.05	0.18	0.11	0.18	0.09	0.13	0.04	0.09	0.09	0.18	0.05
Time period 2																		
City 1	0.19	0.09	0.25	0.12	0.18	0.09	0.23	0.03	0.26	0.15	0.16	0.12	0.18	0.06	0.30	0.14	0.19	0.12
City 2	0.23	0.10	0.21	0.14	0.24	0.15	0.23	0.08	0.25	0.07	0.21	0.14	0.22	0.07	0.23	0.15	0.24	0.09
City 3	0.17	0.18	0.19	0.12	0.18	0.13	0.20	0.07	0.27	0.19	0.20	0.12	0.19	0.06	0.20	0.17	0.18	0.13
City 4	0.16	0.14	0.23	0.10	0.21	0.11	0.17	0.06	0.24	0.16	0.23	0.10	0.16	0.05	0.19	0.12	0.22	0.10
Time period 3																		
City 1	0.20	0.09	0.23	0.12	0.24	0.09	0.21	0.07	0.22	0.08	0.26	0.11	0.19	0.06	0.28	0.05	0.27	0.06
City 2	0.26	0.11	0.19	0.05	0.21	0.11	0.27	0.09	0.19	0.06	0.19	0.06	0.25	0.08	0.19	0.09	0.15	0.09
City 3	0.18	0.08	0.25	0.11	0.16	0.08	0.19	0.06	0.25	0.12	0.24	0.05	0.18	0.05	0.15	0.04	0.23	0.11
City 4	0.16	0.07	0.27	0.10	0.11	0.05	0.17	0.03	0.24	0.11	0.21	0.08	0.16	0.03	0.16	0.08	0.22	0.08
Time period 4																		
City 1	0.20	0.09	0.23	0.12	0.27	0.14	0.21	0.07	0.19	0.08	0.26	0.12	0.19	0.06	0.18	0.15	0.18	0.13
City 2	0.26	0.12	0.27	0.16	0.23	0.11	0.27	0.09	0.23	0.09	0.30	0.12	0.25	0.08	0.19	0.12	0.26	0.09
City 3	0.21	0.09	0.21	0.13	0.17	0.14	0.21	0.07	0.15	0.05	0.27	0.13	0.20	0.06	0.14	0.09	0.19	0.08
City 4	0.17	0.08	0.15	0.11	0.25	0.12	0.18	0.06	0.21	0.03	0.23	0.11	0.17	0.05	0.24	0.07	0.23	0.05

Table 5.1 Average no. of listeners in Radio ('0000)

City	Channel 1		Channel 2		Channel 3	
	PT	OT	PT	OT	PT	OT
Time period 1						
City 1	44.7	18.67	28.4	13.7	19.78	9.2
City 2	24	1.8	18	7.8	13.5	7.8
City 3	34.5	15	9.7	5.6	14.5	7.6
City 4	27.8	14.35	11	4.5	13.5	3.4
Time period 2						
City 1	3.5	17.8	30	16	14.5	8
City 2	19	9	26	10	9.5	5.7
City 3	25	10	13.5	6.7	22.1	12
City 4	23.45	13.8	16.58	8.9	15.9	7.8
Time period 3						
City 1	38.5	14	19	9.8	29	11
City 2	17	7.2	24	9	14	6.7
City 3	28	1.5	21	8.5	13	5.8
City 4	17	8.79	19	8.7	22	11
Time period 4						
City 1	37.5	14	25	9	28	12
City 2	15	7.2	21	8	17	8.5
City 3	29	11	23	8.5	12	5.6
City 4	24.85	9	21	8.6	19	7.6

Table 5.2 Advertisement cost in Radio('00)

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time period 1					
95	32	57	25	43	21
78	21	50	19	41	20
97	35	45	15.5	68	26.7
80	28.9	55	24.5	65	26
Time period 2					
87	27.6	63	32	38	18.9
64.2	19	64	22.5	35.5	18.6
86	31	54	18	75	30
74	22	64	32.4	62	28
Time period 3					
89	22	35	12	65	19.8
54.2	13	71	30	42	20
96	33	72	22	58	18
58.9	17.8	69	31	74	34
Time period 4					
87	25	54	12.5	60	22
32.5	1.5	67	28.9	50	22
92	30	72.5	25.8	42	17.5
86	25.6	62	20	48	15.5

Table 5.3 Upper and Lower Bounds on Advertisement in Radio

City	Channel 1		Channel 2		Channel 3	
	PT	OT	PT	OT	PT	OT
Time period 1						
City 1	[15,120]	[6,68]	[34,110]	[13,40]	[48,90]	[25,32]
City 2	[18,118]	[15,54]	[40,90]	[17,42]	[48,80]	[12,30]
City 3	[22,124]	[8,52]	[45,78]	[12,24]	[36,90]	[18,32]
City 4	[23,130]	[6,65]	[40,95]	[20,34]	[30,75]	[12,28]
Time period 2						
City 1	[12,118]	[8,64]	[32,112]	[10,45]	[52,86]	[28,28]
City 2	[22,114]	[18,52]	[36,94]	[12,46]	[52,75]	[14,26]
City 3	[26,120]	[10,48]	[38,82]	[10,26]	[32,94]	[15,34]
City 4	[24,128]	[8,68]	[38,100]	[16,36]	[32,78]	[10,30]
Time period 3						
City 1	[10,124]	[12,61]	[36,110]	[14,40]	[50,90]	[25,33]
City 2	[24,108]	[20,47]	[34,98]	[12,42]	[51,75]	[18,30]
City 3	[25,127]	[12,52]	[34,90]	[8,32]	[36,90]	[19,30]
City 4	[28,118]	[13,62]	[36,98]	[20,42]	[30,90]	[8,36]
Time period 4						
City 1	[12,130]	[8,65]	[30,115]	[12,38]	[52,88]	[23,35]
City 2	[32,89]	[24,47]	[36,95]	[18,40]	[42,80]	[17,34]
City 3	[28,129]	[16,55]	[34,95]	[11,32]	[36,85]	[22,28]
City 4	[22,120]	[8,59]	[37,110]	[25,40]	[32,85]	[18,45]

Table 5.4 Retention Factor in Radio (%)

Channel 1		Channel 2		Channel 3	
PT	OT	PT	OT	PT	OT
Time period 1					
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
Time period 2					
7	6	5	4	3	3
7	2	8	5	4	2
8	4	4	3	6	8
8	4	5	2	4	5
Time period 3					
9	6	6	4	7	6
7	5	8	6	4	2
8	5	5	3	4	3
9	6	8	4	8	4
Time period 4					
8	5	4	4	7	6
6	6	8	4	5	2
8	4	6	2	6	1
7	6	6	3	5	3

Table 5.5 Listenership Profile Matrix for Radio

City	Channel 1						Channel 2						Channel 3					
	R1		R2		R3		R1		R2		R3		R1		R2		R3	
	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT	PT	OT
Time period 1																		
City 1	.24	.08	.1	.07	.18	.03	.29	.07	.27	.02	.18	.03	.23	.05	.19	.14	.25	.10
City 2	.14	.07	.15	.04	.17	.05	.14	.11	.18	.10	.14	.04	.22	.14	.29	.11	.17	.09
City 3	.17	.10	.3	.06	.24	.07	.21	.10	.25	.09	.27	.12	.14	.10	.09	.02	.18	.07
City 4	.22	.05	.26	.10	.16	.10	.28	.06	.11	.06	.12	.03	.19	.05	.16	.12	.22	.10
Time period 2																		
City 1	.1	.08	.13	.07	.13	.09	.16	.06	.09	.08	.12	.07	.09	.06	.08	.07	.12	.09
City 2	.06	.05	.08	.07	.1	.05	.11	.06	.13	.05	.16	.06	.07	.05	.13	.03	.09	.04
City 3	.07	.045	.1	.035	.1	.05	.07	.04	.11	.03	.1	.05	.045	.03	.05	.035	.06	.04
City 4	.15	.04	.11	.036	.18	.04	.09	.03	.15	.028	.056	.032	.14	.028	.11	.02	.11	.04
Time period 3																		
City 1	.25	.03	.29	.10	.17	.04	.25	.04	.26	.10	.25	.03	.22	.03	.17	.12	.19	.08
City 2	.15	.12	.21	.13	.18	.08	.26	.04	.16	.04	.3	.06	.24	.03	.27	.08	.26	.05
City 3	.18	.05	.16	.05	.23	.09	.18	.12	.28	.07	.22	.02	.28	.10	.15	.05	.28	.13
City 4	.23	.02	.12	.08	.13	.06	.25	.10	.14	.12	.14	.03	.11	.08	.11	.02	.29	.05
Time period 4																		
City 1	.2	.07	.14	.06	.12	.07	.08	.07	.07	.06	.2	.08	.1	.07	.11	.06	.2	.1
City 2	.09	.06	.09	.02	.2	.06	.15	.08	.09	.08	.16	.06	.07	.07	.16	.04	.09	.04
City 3	.08	.05	.13	.04	.12	.02	.08	.05	.05	.04	.09	.06	.16	.03	.05	.05	.11	.04
City 4	.06	.03	.09	.04	.09	.04	.16	.04	.14	.028	.19	.03	.04	.032	.15	.03	.15	.02

Table 6 Weights assigned for customer profile criteria

	R1	R2	R3
Newspaper	0.03	0.15	0.12
Television	0.04	0.16	0.11
Radio	0.07	0.2	0.12

Table 7 Weights assigned to media in stage 2 of Goal Programming in different time periods

	Time period 1	Time period 2	Time period 3	Time period 4
Newspaper	0.38	0.30	0.30	0.32
Television	0.32	0.35	0.4	0.31
Radio	0.30	0.35	0.30	0.37

References

1. Balakrishnan, P.V., Hall, N.G.: A maximin procedure for the optimal insertion timing of ad executions. *Eur. J. Oper. Res.* **85**, 368–382 (1995)
2. Bhattacharya, U.K.: A chance constraints goal programming model for the advertising planning problem. *Eur. J. Oper. Res.* **192**, 382–395 (2009)
3. Charnes, A., Cooper, W.W., DeVoe, J.K., Learner, D.B., Reinecke, W.: A goal programming model for media planning. *Manag. Sci.* **14**, 422–430 (1968)
4. Geoffrion, M.A.: Proper efficiency and vector maximization. *J. Math. Anal. Appl.* **22**(3), 618–630 (1968)
5. Jha, P.C., Aggarwal, R., Gupta, A.: Optimal media planning for multi-product in segmented market. *Appl. Math. Comput.* **217**(16), 6802–6818 (2011)
6. Kumar, A.J.: Quantitative model for press media planning in India. *Vikalpa.* **1**(1), 39–50 (1976)
7. Mihiotis, A., Tsakiris, I.: A mathematical programming study of advertising allocation problem. *Appl. Math. Comput.* **148**(2), 373–379 (2004)
8. Moynihan, G.P., Kumar, A., D'Souza, G., Nockols, W.G.: A decision support system for media planning. *Comput. Ind. Eng.* **29**(1), 383–386 (1995)
9. Pasadeos, Y., Barban, A., Huiuk, Y., Kim, B.-H.: A 30-year assessment of the media planning literature. *J. Curr. Issues Res. Advert.* **19**(1), 23–36 (1997). doi:[10.1080/10641734.1997.10505055](https://doi.org/10.1080/10641734.1997.10505055)
10. Royo, C.B.: Zhang, Blanco, L.A., Almagro, J.: Multistage multiproduct advertising budgeting. *Eur. J. Oper. Res.* **225**(1), 179–188 (2013)
11. Thiriez, H.: OR software LINGO. *Eur. J. Oper. Res.* **12**, 655–656 (2000)

Fuzzy Multi-criteria Approach for Component Based Software System Under Build-or-Buy Scheme

P. C. Jha, Ramandeep Kaur, Sonam Narula and Sushila Madan

Abstract With the rising awareness of advancements in Information technology amongst various industries, the predilection to selection of commercial-off the shelf (COTS) components have increased invariably. It provides the ability to reuse the software components, thereby, maximizing the reliability while reducing the developmental cost. The decision of whether to buy the component or build from scratch, is known as build-or-buy decision. In order to prevent the software from failure, redundant components have to be incorporated which can be ascertained using fault tolerant schemes. Further, the innovation in the field of Application Package Software (APS) has supplemented the industry with highly configurable, sophisticated applications. Through this paper, we shall discuss a framework concentrating upon whether to build or buy the software components while designing a fault-tolerant modular software system. The objective of the paper is to maximize the reliability of the software while minimizing the overall cost. Further, the components with comparatively less execution time are chosen over the ones which require more time for executing the software. Hence the objective of the paper shall further be elaborated upon minimizing the execution time with the aid of a case study on supplementing an APS for Airline Industry.

Keywords Application package software (APS) · Build-or-buy · COTS · Fault tolerant system · Software reliability · Execution time

P. C. Jha (✉) · S. Narula · S. Madan
University of Delhi, Delhi, India
e-mail: jhpc@yahoo.com

S. Narula
e-mail: sonam.narula88@gmail.com

S. Madan
e-mail: sushila_lsr@yahoo.com

R. Kaur
Institute of Information Technology and Management, Delhi, India
e-mail: rrdk_07@yahoo.com

1 Introduction

Today's need for ready-made application software has generated a great interest in employing Component Based Software Engineering (CBSE) approach for software development. These software are built by independent software development companies by way of integrating various software components that are designed to fulfill a whole range of business functions, and these offerings are popularly known as *Application Package Software (APS)*. Most of the large organizations have already adopted these software and more of small- and medium-sized enterprises (SMEs) too are finding it cost effective and a competitive necessity for sustainability. Packaged software, although ready-made, rarely comes ready-to-run. It typically requires weeks or months of integration effort to set up for the specific needs of the client's or business. Development of APS using CBSE approach has a great potential for reducing development time, cost and effort. It follows a well-defined process, which starts with the identification of functional requirements by the client organization. Each functional requirement can be performed by invoking at least one module. Furthermore, the number of modules in the software system are decided by software development team. Each module is then built by integrating at least one component, also termed as alternative. These alternative components can be obtained either as Commercial off-the Shelf (COTS) components or as in-house built components. The schematic representation of the software system developed using CBSE approach is given in Fig. 1.

Most of the software companies have now adopted build-or-buy strategy [1] for software development. COTS are the readymade software components that are available in the software market. The software vendor provides information on cost, reliability, execution time and delivery time of the COTS components. If COTS components are not available in the market for a particular functional requirement then it becomes a mandatory decision to build it which can be cost effective also. The parameters of cost, reliability, delivery time and execution time of in-house built components can be estimated by software development team. In this paper, we are throwing light on problem of optimal selection of components during the design phase of the software development. During this phase the software development team ends up with n number of components for selection. Each component is then evaluated on the basis of different parameters such as cost, reliability, delivery time, execution time etc. It is beyond human tendency to evaluate all the alternatives simultaneously on the basis of different parameters and then select the best component. Therefore, the technique of mathematical optimization will help in taking the best decision.

Since software has become a life line of the business organization, therefore, it would expect failure free software system to be developed for their organization. In reality none of the software can be made failure free but there are techniques available in the literature that can minimize the probability of failure. One such technique is called fault tolerance. A system is known as a Fault tolerant system if it has an inbuilt ability to tolerate faults and prevents the system from failure. Two

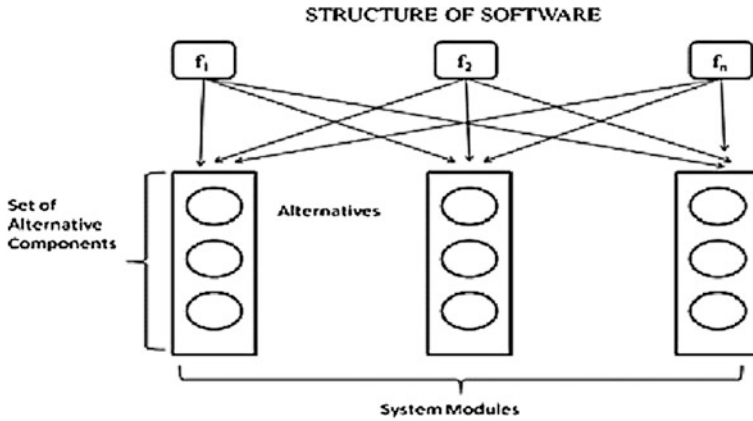


Fig. 1 Structure of software

widely known fault-tolerant techniques exist in the literature are Recovery Block (RB) and N-version Programming (NVP). As already mentioned, software components are available for each module. So, we allow redundancy, i.e., at least one component is selected for a module. If the first component fails then the other component will take over and hence reliability is improved.

The objective of this paper shall be to focus on a framework that will aid the developers to resolve upon the issue of whether to build or buy the software components on the basis of various factors, viz., cost, reliability, execution time and delivery time. Since human judgments about cost, reliability, delivery time and execution time are often vague and imprecise. Therefore, we shall formulate a fuzzy multi-criteria optimization model which maximizes reliability and minimizes cost and execution time of the software system with the threshold on reliability, execution time and delivery time. The optimization model is based on Recovery Block scheme of fault tolerance.

2 Literature Review

Vast research has been carried out on CBSE approach for software development. Jung and Choi [2] introduced two optimization models for COTS selection in the development of modular software system considering cost-reliability trade-off. Ashrafi and Berman [3] presented optimization models which addresses the trade-off between reliability and cost in development of large software packages that consist of large software programs. Another important criteria that need to be considered while developing a software system is fault tolerance. It can be realized through various techniques of fault tolerance such as NVP, RBS and Consensus Recovery Block (CRB). Tremendous efforts have been spent to study fault tolerant

architecture in software systems. Scott et al. [4–7] conducted incredible research and introduced data domain reliability models of several fault-tolerant software schemes, including NVP, RB, and CRB. Levitin et al. [8] proposed a simple straightforward algorithm for evaluating reliability and expected execution time for software systems consisting of fault-tolerant components. He formulated a universal model both for NVP and RBS where the components are built from functionally equivalent but independently developed versions characterized by different reliability and performance. Tang et al. [9] proposed an optimization model to solve component selection problem by considering the concept of reusability and compatibility matrix. The model can be used to assist software developers in selecting software components when multi-applications are undertaken concurrently. Jadhav et al. [10] proposed hybrid knowledge based system approach which can be used by decision makers as a tool for software selection as it supports various software evaluation activities such as: choosing criteria for software evaluation; specifying and changing user requirements of the software package; determining the fit between software package and user needs of that package; and for reusing knowledge/experience. Daneshgar et al. [11] proposed that an understanding of the deeply rooted reasons for adoption of a particular method for acquiring software systems by the SMEs, among other things, will enhance effectiveness of these decisions. The most relevant factors that are universally acceptable and important to both SMEs and large organizations include ‘Cost’, ‘Scale & Complexity’, ‘Support Structure’ and ‘Operational factors’.

3 Decision Framework for Component Selection Using Build-or-Buy Strategy

An optimization model has been formulated for a software system under recovery block fault tolerant scheme. The model is formulated for the structure given in Fig. 1. Recovery Block requires n alternatives of a program and a testing segment called acceptance test (AT). Whenever an alternative fails, the testing segment activates the succeeding alternative. The function of the testing segment is to ensure that the operation performed by an alternative is correct. If the output of the alternative is incorrect, then the testing segment recovers the initial state and activates the next alternative.

3.1 Notations

<i>RB</i>	System reliability measure
<i>CT</i>	Overall system cost
<i>ET</i>	Overall execution time

F_l	Frequency of use, of function l
S_l	Set of modules required for function l
R_i	Reliability of i th module
L	Number of functions the software is required to perform
n	Number of modules in the software
m_i	Number of alternatives available for i th module
N_{ij}^{tot}	Total number of test cases performed on j th in-house component of i th module
N_{ij}^{suc}	Total number of successful test cases performed on j th in-house component of i th module
t_1	Probability that next alternative is not invoked upon failure of the current alternative
t_2	Probability that a correct result is judged wrong
t_3	Probability that an incorrect result is accepted as correct
X_{ij}	Event that output of j th component of i th module is rejected
Y_{ij}	Event that correct result of j th component of i th module is accepted
r_{ij}	Reliability of j th component of i th module
\widetilde{C}_{ij}	Fuzzy cost of j th COTS component of i th module
\overline{C}_{ij}	Defuzzified cost of j th COTS component of i th module
\widetilde{s}_{ij}	Fuzzy reliability of j th COTS component of i th module
\overline{s}_{ij}	Defuzzified reliability of j th COTS component of i th module
\widetilde{d}_{ij}	Fuzzy delivery time of j th COTS component of i th module
\overline{d}_{ij}	Defuzzified delivery time of j th COTS component of i th module
\widetilde{c}_{ij}	Fuzzy unitary development cost of j th in-house component of i th module
\overline{c}_{ij}	Defuzzified unitary development cost of j th in-house component of i th module
\widetilde{t}_{ij}	Fuzzy estimated development time of j th in-house component of i th module
\overline{t}_{ij}	Defuzzified estimated development time of j th in-house component of i th module
τ_{ij}	Average time required to perform a test case on j th in-house component of i th module
π_{ij}	Probability that a single execution of software fails on a test case chosen from a certain input distribution of j th in-house component of i th module
ρ_{ij}	the probability that the j th in-house developed component of i th module is failure free during a single run given that N_{ij}^{suc} test cases have been successfully performed
R_o	Threshold on module reliability
D_o	Threshold on delivery time
E_l	Threshold on execution time of the l th function
\widetilde{T}_{ij}	Fuzzy execution time of j th COTS component of i th module
\overline{T}_{ij}	Defuzzified execution time of j th COTS component of i th module
$\widetilde{T}e_{ij}$	Fuzzy execution time of j th in-house component of i th module

$\bar{T}e_{ij}$	Defuzzified execution time of j th in-house component of i th module
y_{ij}	$\begin{cases} 1 & \text{if } j\text{th component of } i\text{th module is in-house developed} \\ 0 & \text{otherwise} \end{cases}$
x_{ij}	$\begin{cases} 1, & \text{if } j\text{th component of } i\text{th module is COTS} \\ 0, & \text{otherwise} \end{cases}$
z_{ij}	$\begin{cases} 1, & \text{if } j\text{th component of } i\text{th module is selected} \\ 0, & \text{otherwise} \end{cases}$

3.2 Model Assumptions

The optimization model is based on the following assumptions:

1. Software is developed using modular approach where the number of modules considered are finite.
2. Each module is a logical collection of several independent developed components. The components available are also finite in number.
3. A software system is required to perform a known number of functions. The program written for a function can call a series of modules. A failure occurs if a module fails to carry out an intended operation.
4. Codes written for integration of modules don't contain any bug.
5. Several alternatives are available for each module.
6. Fault tolerant architecture is desired in the modules (it has to be within the specified budget). Independently developed alternatives (COTS or in-house build components) are attached in the modules and work similar to the recovery block scheme discussed in [12, 13]
7. Redundancy is allowed in the modules. For each module more than one component can be selected.
8. The cost of an alternative is the development cost, if developed in house; otherwise it is the buying price for the COTS product. Reliability for all the COTS components are known and no separate testing is done.
9. Cost and reliability of an in-house component can be specified by using basic parameters of the development process, e.g. a component cost may depend on a measure of developer skills, or the component reliability depends on the amount of testing.
10. Different COTS alternatives, with respect to cost, reliability, delivery time and execution time of a module are available.
11. Different In-house alternatives with respect to unitary development cost, estimated development time, average time and testability of a module are available.

3.3 Model Formulation

Optimization model formulated in this paper aims at optimal selection of components so as to maximize system reliability and minimize cost and execution time of the software system. Let S be a software architecture made of n modules, with a maximum number of m_i alternatives (i.e. COTS or in-house) available for each module.

In order to select components for the modular software system, the following criteria may be used.

3.3.1 Build Versus Buy Decision

If j th alternative of i th module is bought (i.e. some $x_{ij} = 1$), then no in-house development of the module will take place (i.e. $y_{ij} = 0$) and vice versa.

Hence

$$y_{ij} + x_{ij} = z_{ij}$$

$$z_{ij} \leq 1; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

Also, to incorporate redundancy at modular level for both build-or-buy components we have

$$\sum_{j=1}^{m_i} z_{ij} \geq 1; \quad i = 1, 2, \dots, n$$

3.3.2 Test Cases Performed

The number of successful test cases N_{ij}^{suc} performed on the j th alternative of i th module can be given as:

$$N_{ij}^{suc} = (1 - \pi_{ij})N_{ij}^{tot}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

3.3.3 Reliability of In-house Components

Consider an event A characterized by accomplishment of N_{ij}^{suc} failure free test cases while event B as the failure-free execution of alternative during a single run.

Then, the probability ρ_{ij} that the in-house developed alternative of a module is failure free during a single run may be written as below from the Baye's theorem:

$$\rho_{ij} = P(B/A) = \frac{P(A/B)P(B)}{P(A/B)P(B) + P(A/\bar{B})P(\bar{B})}$$

The following equalities can be ascertained as:

- $P(A/B) = 1$
- $P(B) = 1 - \pi_{ij}$
- $P(A/\bar{B}) = (1 - \pi_{ij})^{N_{ij}^{suc}}$
- $P(\bar{B}) = \pi_{ij}$

Therefore, we have

$$\rho_{ij} = \frac{1 - \pi_{ij}}{(1 - \pi_{ij}) + \pi_{ij}(1 - \pi_{ij})^{N_{ij}^{suc}}}; \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m_i$$

3.3.4 Reliability for Both In-house and COTS Components

The probability of failure free operation of j th alternative of i th module is given as:

$$r_{ij} = \rho_{ij}y_{ij} + \tilde{s}_{ij}x_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

3.3.5 Reliability of Recovery Block Scheme

Reliability of i th module under recovery block scheme with m_i alternatives is given by:

$$R_i = \sum_{j=1}^{m_i} z_{ij} \left[\prod_{p=1}^{j-1} P(X_{ij})^{z_{ip}} \right] P(Y_{ij})^{z_{ij}}; \quad i = 1, 2, \dots, n$$

where

$$P(X_{ij}) = (1 - t_1) [(1 - r_{ij})(1 - t_3) + r_{ij}t_2]$$

$$P(Y_{ij}) = r_{ij}(1 - t_2)$$

3.3.6 Threshold on Reliability

The threshold on reliability of each module is gives as:

$$R_i \geq R_o; \quad i = 1, 2, \dots, n$$

3.3.7 Threshold on Delivery Time

The total delivery includes development, integration and system testing time. The maximum threshold D_o has been given on the delivery time of the whole system. If the component is COTS then the delivery time is simply given by d_{ij} , whereas for in-house developed component the delivery time is given as $(t_{ij} + \tau_{ij}N_{ij}^{tot})$. Hence delivery time of j th alternative of i th module is given by D_{ij} which is expressed as:

$$(\tilde{t}_{ij} + \tau_{ij}N_{ij}^{tot})y_{ij} + \tilde{d}_{ij}x_{ij} = D_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$\max(D_{ij}) \leq D_o$$

3.3.8 Threshold on Execution Time

Threshold is set on the total execution time of the l th function and can be written as:

$$\sum_{i \in S_l} \sum_{j=1}^{m_i} [\tilde{T}e_{ij}y_{ij} + \tilde{T}_{ij}x_{ij}] \leq E_l; \quad l = 1, 2, \dots, L$$

3.3.9 Reliability

Reliability objective function maximizes the system quality (in terms of reliability) through a weighted function of module reliabilities. Reliability of modules that are invoked more frequently during use is given higher weights. Analytic Hierarchy Process can be effectively used to calculate these weights.

$$\text{Maximize } RB \cong \sum_{l=1}^L f_l \prod_{i \in S_l} R_i$$

where R_i is the reliability of the i th module under recovery block scheme and is calculated as discussed in [Sect. 3.3.7](#).

3.3.10 Cost

The overall cost of all the modules calculated from the selection of components obtained as a result of build-or buy strategy is minimized using the cost objective. The development cost of in-house components of j th alternative of i th module can be expressed as $\tilde{c}_{ij}(t_{ij} + \tau_{ij}N_{ij}^{tot})$

$$\text{Minimize } CT \cong \sum_{i=1}^n \sum_{j=1}^{m_i} \left[\tilde{c}_{ij} \left(\tilde{t}_{ij} + \tau_{ij} N_{ij}^{tot} \right) y_{ij} + \tilde{C}_{ij} x_{ij} \right]$$

3.3.11 Execution Time

Long execution time for performing a function may causes dissatisfaction to the user and hence can result in low productivity. The following objective function minimizes execution time of the software system.

$$\text{Minimize } ET \cong \sum_{l=1}^L f_l \sum_{i \in S_l} \sum_{j=1}^{m_i} \left[\tilde{T} e_{ij} y_{ij} + \tilde{T}_{ij} x_{ij} \right]$$

4 Optimization Model

The optimization model incorporates fault tolerance using Recovery Block Scheme. The modules are bought from the vendor in the form of COTS components. More than one alternative may exists for a given module. When a module gets selected for a particular function, its corresponding first alternative is submitted to the recovery block for execution and the result is presented for acceptance testing. If the first alternative is rejected, the second module is executed. This process follows till all the alternatives are either rejected i.e. the entire module fails or any one is accepted. This fault tolerance is achieved by incorporating redundancies in the components.

Problem (P1)

$$\text{Maximize } RB \cong \sum_{l=1}^L f_l \prod_{i \in S_l} R_i \quad (1)$$

$$\text{Minimize } CT \cong \sum_{i=1}^n \sum_{j=1}^{m_i} \left[\tilde{T} c_{ij} \left(\tilde{T} t_{ij} + \tau_{ij} N_{ij}^{tot} \right) y_{ij} + \tilde{T} C_{ij} x_{ij} \right] \quad (2)$$

$$\text{Minimize } ET \cong \sum_{l=1}^L f_l \sum_{i \in S_l} \sum_{j=1}^{m_i} \left[\tilde{T} T e_{ij} y_{ij} + \tilde{T}_{ij} x_{ij} \right] \quad (3)$$

Subject to $X \in S = \{x_{ij}, y_{ij}, z_{ij} \text{ are decision variables}$

$$\sum_{i \in S_l} \sum_{j=1}^{m_i} \left[\tilde{T} e_{ij} y_{ij} + \tilde{T}_{ij} x_{ij} \right] \leq E_l; \quad l = 1, 2, \dots, L \quad (4)$$

$$(\tilde{t}_{ij} + \tau_{ij}N_{ij}^{tot})y_{ij} + \tilde{d}_{ij}x_{ij} = D_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{5}$$

$$\max(D_{ij}) \leq D_o \tag{6}$$

$$\rho_{ij} = \frac{1 - \pi_{ij}}{(1 - \pi_{ij}) + \pi_{ij}(1 - \pi_{ij})^{N_{ij}^{suc}}}; \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m_i \tag{7}$$

$$r_{ij} = \rho_{ij}y_{ij} + \tilde{s}_{ij}x_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{8}$$

$$R_i = \sum_{j=1}^{m_i} z_{ij} \left[\prod_{p=1}^{j-1} P(X_{ij})^{z_{ip}} \right] P(Y_{ij})^{z_{ij}} \geq R_0; \quad i = 1, 2, \dots, n \tag{9}$$

$$P(X_{ij}) = (1 - t_1) [(1 - r_{ij})(1 - t_3) + r_{ij}t_2] \tag{10}$$

$$P(Y_{ij}) = r_{ij}(1 - t_2) \tag{11}$$

$$N_{ij}^{suc} = (1 - \pi_{ij})N_{ij}^{tot}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{12}$$

$$y_{ij} + x_{ij} = z_{ij} \tag{13}$$

$$z_{ij} \leq 1; \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m_i \tag{14}$$

$$\sum_{j=1}^{m_i} z_{ij} \geq 1; \quad i = 1, 2, \dots, n \tag{15}$$

$$x_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{16}$$

$$y_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{17}$$

$$z_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i \tag{18}$$

5 Fuzzy Approach for Finding Solution

Under many situations, crisp data are inadequate to model real life scenarios. Since human judgments about cost, reliability, delivery time and execution time etc. are often vague and cannot be estimated with an exact numerical value. Therefore, fuzzy optimization is a flexible approach that permits more adequate solutions of real problems in the presence of vague information.

Following algorithm specifies the sequential steps to solve the fuzzy mathematical programming problems.

- Step 1. Compute the crisp equivalent of the fuzzy parameters using a defuzzification function. Ranking technique $F_2(A) = (a^1 + 2a^2 + a^3)/4$ is used for defuzzification of parameters. Same defuzzification function is to be used

for each of the parameters where a^1, a^2, a^3 are the triangular fuzzy numbers.

Step 2. Since software industry is volatile and development team changes time to time frequently, a precise estimation of cost, reliability and execution time aspirations are a big question mark. Hence the better way to come out in such situation is that the problem should be allowed to have some tolerance with respect to aspiration of each objective. The resulting problem can be formulated in fuzzy environment as follows:

Problem (P2)

Find X

Subject to $RB \underset{\sim}{\geq} R_0$

$$CT \underset{\sim}{\leq} C_0$$

$$ET \underset{\sim}{\leq} E_0$$

$X \in \bar{S} = \{x_{ij}, y_{ij}, z_{ij} \text{ are decision variables}$

$$\sum_{i \in S_l, j=1}^{m_i} [\bar{T}e_{ij}y_{ij} + \bar{T}_{ij}x_{ij}] \leq E_l; \quad l = 1, 2, \dots, L$$

$$(\bar{t}_{ij} + \tau_{ij}N_{ij}^{tot})y_{ij} + \bar{d}_{ij}x_{ij} = D_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$\max(D_{ij}) \leq D_o$$

$$\rho_{ij} = \frac{1 - \pi_{ij}}{(1 - \pi_{ij}) + \pi_{ij}(1 - \pi_{ij})^{N_{ij}^{suc}}}; \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m_i$$

$$r_{ij} = \rho_{ij}y_{ij} + \bar{s}_{ij}x_{ij}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$R_i = \sum_{j=1}^{m_i} z_{ij} \left[\prod_{p=1}^{j-1} P(X_{ij})^{z_{ip}} \right] P(Y_{ij})^{z_{ij}} \geq R_0; \quad i = 1, 2, \dots, n$$

$$P(X_{ij}) = (1 - t_1) [(1 - r_{ij})(1 - t_3) + r_{ij}t_2]$$

$$P(Y_{ij}) = r_{ij}(1 - t_2)$$

$$N_{ij}^{suc} = (1 - \pi_{ij})N_{ij}^{tot}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$y_{ij} + x_{ij} = z_{ij}$$

$$z_{ij} \leq 1; \quad i = 1, 2, \dots, n \text{ and } j = 1, 2, \dots, m_i$$

$$\sum_{j=1}^{m_i} z_{ij} \geq 1; \quad i = 1, 2, \dots, n$$

$$x_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$y_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i$$

$$z_{ij} \in \{0, 1\}; \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m_i\}$$

where R_0, C_0 and E_0 are defuzzified aspiration levels of system reliability, cost and execution time.

Step 3. Define appropriate membership functions for each fuzzy inequalities as well as constraint corresponding to the objective function. The membership function for the fuzzy less than or equal to and greater than or equal to type are given as

$$\mu_{R(X)} = \begin{cases} 1; & RB \geq R_0 \\ \frac{RB - R_0^*}{R_0 - R_0^*}; & R_0^* \leq RB < R_0 \\ 0; & CT \leq C_0 \end{cases}$$

$$\mu_{C(X)} = \begin{cases} 1; & RB \geq R_0 \\ \frac{C_0^* - CT}{C_0^* - C_0}; & C_0 \leq CT < C_0^* \\ 0; & CT > C_0^* \end{cases}$$

$$\mu_{E(X)} = \begin{cases} 1; & ET \leq E_0 \\ \frac{E_0^* - ET}{E_0^* - E_0}; & E_0 \leq ET < E_0^* \\ 0; & ET > E_0^* \end{cases}$$

where R_0 is aspiration level and R_0^* is the tolerance levels to the fuzzy reliability objective function constraint. C_0^* is aspiration level and C_0 is the tolerance levels to the fuzzy cost objective function constraint. E_0^* is aspiration level and E_0 is the tolerance levels to the fuzzy cost objective function constraint.

Step 4. Employ extension principles to identify the fuzzy decision, which results in a crisp mathematical programming problem (P3) given by

Problem (P3)

$$\begin{aligned} & \text{Maximize} && \alpha \\ & \text{Subject to} && \lambda_1 \alpha \leq \mu_{R(x)} \\ & && \lambda_2 \alpha \leq \mu_{C(x)} \\ & && \lambda_3 \alpha \leq \mu_{E(x)} \\ & && 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^3 \lambda_i = 1 \\ & && X \in \bar{S} \end{aligned}$$

Step 5. Following Bellman-Zadeh’s maximization principle [14] and using the above defined fuzzy membership functions; the fuzzy multi-objective optimization model for the problem (P1) is formulated as follows.

Problem (P4)

$$\begin{aligned} & \text{Maximize} && \alpha \\ & && R(X) \geq R_0 - (1 - \lambda_1 \alpha)(R_0 - R_0^*) \\ & && C(X) \leq C_0 + (1 - \lambda_2 \alpha)(C_0^* - C_0) \\ & && E(X) \leq E_0 + (1 - \lambda_3 \alpha)(E_0^* - E_0) \\ & \text{s.t} && 0 \leq \lambda_i \leq 1, \quad \sum_{i=1}^3 \lambda_i = 1 \\ & && X \in \bar{S} \end{aligned}$$

The optimal value of α represents the best compromise solution between three objective functions. In software industry, priorities and relative importance can be obtained by interacting with the management. The optimal solution to the problem (P4) will also be the solution of problem (P1). The solution will provide the optimal mix of components (COTS or In-house) that gives the trade-off between Reliability, Cost and Execution time.

6 Case Study

A case study of 'Airlines Industry' software system is presented to illustrate the given methodology. The methodology is applied to select right mix of components for development of fault-tolerant modular software system. The optimal solution of the optimization model will give a set of components, either in-house built or COTS, so as to have a system which is highly reliable, delivered on time, within a budget and also takes less execution time. A real world case problem is selected to illustrate the application of the given approach. The selected organization is an airline company. The company was facing a problem in integration and functional performance of various departments as well as in booking and cancellation of air tickets, estimation of airfares, calculation of discounts, estimation of the food that is to be served to the customers during travel, allocation of flights to pilots, scheduling of flights, optimal procurement of fuel, simulation and training, airport management and more. The present system which they have is not efficient and up to date. Therefore, the members in the top management decided to procure a new system for their company. The Airline Company invited proposals from various software companies for development of a customized software package for them so as to bring out smoothness in functioning and managing of the airport and airplanes. The top management evaluated numerous proposals on the basis of various parameters such as delivery time, execution time, reliability, budget, functionality etc. and then they selected and signed a contract with one of the software companies.

Software Company adopts CBSE approach for development of software system for this airline company. On the basis of functional requirements given by Airlines Company, software development team has identified nine software modules which can perform almost all the functions given by the airline company and is given in Table 1.

Since, each module is build-up by integrating components, they can be either COTS or in-house build components. COTS vendor provides information on cost, reliability, delivery time and execution time of each component. The software development team estimates values of cost, reliability, development time, execution time of in-house build components. The objective of the software development team is to take a decision on selection of right mix of components which will help in developing software which is highly reliable, within a budget, delivered on

Table 1 Functional requirements for airline company software

Modules	Functional requirements
Reservation system	<ul style="list-style-type: none"> ✓ Provides online booking and cancellation, ✓ flight fare search, ✓ seat availability and ✓ schedule etc.
Flight management	<ul style="list-style-type: none"> ✓ Flight data monitoring ✓ Flight tracking ✓ Flight crew scheduling
Marketing and planning	<ul style="list-style-type: none"> ✓ Cargo management ✓ On board catering ✓ Schedule management
Pricing and revenue management	<ul style="list-style-type: none"> ✓ Fares management ✓ Passenger revenue management ✓ Revenue analysis
CRM	<ul style="list-style-type: none"> ✓ Helps improve the interaction and involvement of customer
Accounts/finance	<ul style="list-style-type: none"> ✓ Periodic and as on date balance sheet ✓ Periodic profit & loss ✓ Ratio analysis ✓ Cash flow ✓ Fund flow ✓ Budget analysis ✓ Fixed asset management ✓ Asset purchase sale ✓ Asset scraps
Airport management	<ul style="list-style-type: none"> ✓ Fleet Management ✓ Baggage Handling ✓ Fuel Management
Safety management system	<ul style="list-style-type: none"> ✓ Provides a systematic way to identify hazards and control risks while maintaining assurance that these risk controls are effective
Simulation and training	<ul style="list-style-type: none"> ✓ Artificially re-creates aircraft flight and the environment in which it flies, for pilot training, design, or other purposes

time and also takes less execution time. The software development team collected information on various parameters of COTS component that is given in Table 2.

The software development team compares in-house build components with that of COTS, with respect to cost, reliability, delivery time and execution time. The components (COTS or in-house) which satisfy these criteria will get selected for the final development of software. Sometimes, there may be a situation that for particular module a COTS component is not available in the market, and then in-house development of that component becomes a mandatory decision. We can see from Table 2 that for modules 8 and 9 no COTS components are available, therefore, development of these components becomes a mandatory decision irrespective of the cost involved. Data set for in-house built component is given in Table 3.

The initial parameters, given in Table 3, help in estimating cost, reliabilities and development time of in-house build components. In total thirty components

Table 2 Defuzzified data set of cots components

Module	Module specification	Alternatives	Cost	Reliability	Delivery time	Execution time
M1	Reservation system	x_{11}	24	0.97	1	0.29
		x_{12}	20	0.93	2	0.35
		x_{13}	18	0.90	2	0.34
		x_{14}	26	0.98	1	0.25
		x_{15}	19	0.87	2	0.32
M2	Flight management	x_{21}	32	0.94	2	0.92
		x_{22}	36	0.94	2	0.90
M3	Marketing and planning	x_{31}	12	0.89	2	0.27
		x_{32}	18	0.92	1	0.25
		x_{33}	15	0.90	1	0.24
M4	Pricing and revenue management	x_{41}	18	0.92	2	0.35
		x_{42}	12	0.87	2	0.40
		x_{43}	16	0.91	2	0.37
M5	CRM	x_{51}	18	0.93	2	0.43
		x_{52}	17	0.92	1	0.52
		x_{53}	18	0.91	2	0.50
M6	Accounts/finance	x_{61}	6.5	0.97	1	0.57
		x_{62}	5.5	0.96	2	0.62
		x_{63}	5.3	0.93	2	0.67
M7	Airport management	x_{71}	34	0.99	4	0.82
		x_{72}	28	0.95	5	0.91
M8	Safety management system	<i>Nil</i>				
M9	Simulation and training	<i>Nil</i>				

Table 3 Defuzzified data set for in-house components

Module	Components	Cost	Development time	Execution time
M1	y_1	20	7	0.36
M2	y_2	30	8	0.5
M3	y_3	15	4	0.84
M4	y_4	13	5	0.85
M5	y_5	15	3	0.78
M6	y_6	4	2	0.92
M7	y_7	25	5	0.94
M8	y_8	20	7	0.97
M9	y_9	16	5	0.96

are available to the software development team, out of these nine are in-house build components and twenty one are COTS components. The decision maker has to evaluate each component on the basis of different parameters viz., cost, reliability, delivery time and execution time. Therefore, a scientific method of mathematical optimization has been adopted to deal with such a situation that involves too many decision variables. For this an optimization model was

formulated and is given in Sect. 4. The formulated model is solved using software package LINGO.

Solution

The aspiration and tolerance for each Objective function is assumed to be as follows:

Objective	Aspiration			Tolerance		
Reliability	0.89	0.935	0.96	0.45	0.50	0.67
Cost	335	344.5	355	372	375.5	384
Execution time	2.5	3.40	4.98	8.79	8.97	9.31

The various thresholds on different constraints were assumed as follows:

E_1	D_o	R_o	λ_i			f_i		
			λ_1	λ_2	λ_3	f_1	f_2	f_3
25	10	0.80	0.4	0.3	0.3	0.3	0.3	0.4

The fuzzy problem is developed and solved using a software package LINGO [15] and the solution thus obtained is:

Module	1	2	3	4	5	6	7	8	9
Components selected	x_{13}	x_{21}	x_{33}	x_{41}	y_{51}	x_{62}	x_{71}	y_{81}	y_{91}

Reliability	Cost	Execution time	α
0.69	360.95	4.97	1.0

7 Conclusion

In the present research, we have formulated a multi-criteria optimization model for optimal component selection using build-or-buy strategy under recovery block fault tolerant scheme. The objective functions comprise of simultaneously maximizing reliability and minimizing cost and execution time of the software system under the restriction of delivery time, threshold on execution time and threshold on modular reliability. The formulated methodology involves subjective judgment about each component from the software development team and hence the

proposed solution by crisp optimization may not produce better results. For the same reason fuzzy optimization technique is used to deal with the imprecision or vagueness caused due to subjective judgment. The utility of the model is illustrated using a case study of ‘Airline Industry’. The solution to the case study gives the optimal mix of COTS and in-house components selected that will give the optimal value of reliability, cost and execution time for the software system under the above mentioned limitations.

References

1. Cortellessa, V., Marinelli, F., Potena, P.: An optimization framework for “build-or-buy” decisions in software architecture. *J. Comput. Oper. Res.* **35**, 3090–3106 (2008)
2. Jung, H.W., Choi, B.: Optimization models for quality and cost of modular software systems. *Eur. J. Oper. Res.* **112**(3), 613–619 (1999)
3. Berman, O., Ashrafi, N.: Optimization models for reliability of modular software systems. *IEEE Trans. Softw. Eng.* **19**(11), 1119–1123 (1993)
4. Scott, R.K., Gault, J.W., McAllister, D.F.: The consensus recovery block. In: *Proceedings of Total Systems Reliability Symposium*, pp. 74–85 (1983)
5. Scott, R.K., Gault, J.W., McAllister, D.F.: Modeling fault tolerant software reliability. In: *Proceedings of Third Syrup, Reliability in Distributed Software and Database Systems*, pp. 15–27 (1983)
6. Scott, R.K., Gault, J.W., McAllister, D.F., Wiggs, J.: Experimental validation of six fault-tolerant software reliability models. In: *IEEE Fault Tolerant Computer Systems*, vol. 14, pp. 102–107 (1984)
7. Scott, R.K., Gault, J.W., McAllister, D.F.: Fault tolerant software reliability modeling. *IEEE Trans. Soft. Eng.* **13**, 582–592 (1987)
8. Levitin, G.: Reliability and performance analysis for fault-tolerant programs consisting of versions with different characteristics. *Reliab. Eng. Syst. Saf.* **86**, 75–81 (2004)
9. Tang, J.F., Mu, L.F., Kwong, C.K., Luo, X.G.: An optimization model for software component selection under multiple applications development. *Eur. J. Oper. Res.* **212**, 301–311 (2011)
10. Jadhav, A.S., Sonar, R.M.: Framework for evaluation and selection of the software packages: a hybrid knowledge based system approach. *J. Syst. Softw.* **84**, 1394–1407 (2011)
11. Daneshgar, F., Worasinchai, L., Low, G.: An investigation of ‘build vs. buy’ decision for software acquisition by small to medium enterprises. Available at: <http://ssrn.com/abstract=1867839> (2013). Last Accessed 15 Aug 2013
12. Berman, O., Kumar, U.D.: Optimization models for recovery block scheme. *Eur. J. Oper. Res.* **115**, 368–379 (1999)
13. Kumar, U.D.: Reliability analysis of fault tolerant recovery block. *OPSEARCH* **35**, 281–294 (1998)
14. Bellman, R.E., Zadeh, L.A.: Decision-making in a fuzzy environment. *Manag. Sci.* **17**(4), 141–164 (1970)
15. Thiriez, H.: OR software LINGO. *Eur. J. Oper. Res.* **124**, 655–656 (2000)

A Novel Lossless ECG Compression Technique for Transmission in GSM Networks

Diana Moses and C. Deisy

Abstract This paper presents a novel Lossless ECG Compression using Symbol substitution (LECS) deployable on low computational devices (LCD) like mobile phones for effective use in telecardiology. Of the few LCD deployable compression algorithms, even losslessly compressed ECG suffers transmission loss in Global System for Mobile (GSM) networks due to the reduced character set supported by the SMS protocols. The LECS encodes using the Standard GSM Character ETSI GSM 03.38 set for un-trimmed ECG transmission. The evaluation using MIT-BIH Arrhythmia database showed an average compression-ratio (CR) of 7.03, Percentage-Root-mean-square-Distortion (PRD) as low as 0.0211 proving superior performance in both compression and quality for real-time mobile based telecardiology applications.

Keywords Lossless ECG compression · ECG in SMS · ECG transmission · Telecardiology · ECG transmission in GSM network · Symbol substitution based compression

1 Introduction

Telecardiology is the electronic transmission of cardiac data viz the Electrocardiograph (ECG) acquired from the patient to a health care service provider for diagnostic purposes. The current availability of doctors is only 14 per 10,000 patients as indicated by the WHO [1]. Increase in aged population living alone and

D. Moses (✉) · C. Deisy
Thiagarajar College of Engineering, Madurai, India
e-mail: itsdianamoses@gmail.com

C. Deisy
e-mail: cdcse@tce.edu

lifestyle modifications are few factors that account for the inevitability of telecardiology based applications.

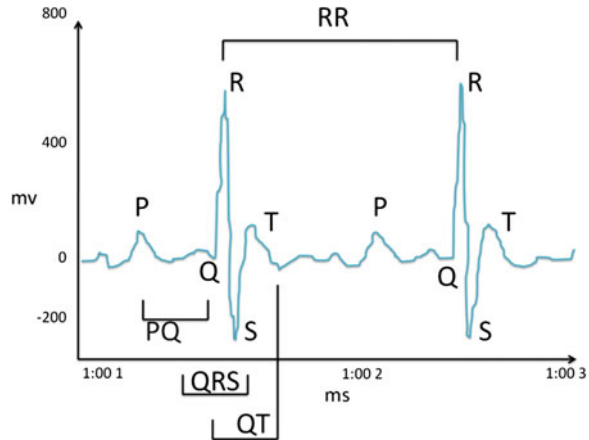
The Electrocardiogram is a noninvasive, transthoracic device used to record the electrical activity of the heart. Figure 1 shows a sample ECG waveform depicting the key features that aid in diagnosis of a heart disease. The key features used for diagnosis are HRV (Heart Rate Variability), RR Interval and Width of the QRS Complex [2]. The reflection of Cardiovascular Diseases (CVDs) occurs in the shape and size of the P-QRS-T waves randomly on the ECG timescale. Thus ECG is recorded for extended periods of time (24–48 h). This calls for compression of the acquired data both for remote monitoring and archival purposes.

The existing ECG compression algorithms can be classified into lossless and lossy based on the reconstruction potential of the algorithm. Lossless compression methods ensure no loss of information and exact reconstruction of original signal from the compressed signal. The basic approaches for lossless compression are based on Entropy, Dictionaries and Sliding windows. Entropy based methods include Huff-man, Shannon-Fano Coding and Arithmetic coding. Lempel–Ziv Welsch (LZW) and some of its variations employ dictionaries for compression. LZ77 and DEFLATE apply sliding window based compression [3, 4]. Antidictionary and ASCII based methods are also reported for lossless ECG compression. Lossless methods generally provide low compression ratios of 2:1 to 4:1 [5].

Lossy compression algorithms provide compression ratios of up to 10:1 to 20:1 where the high compression ratio is achieved by retaining only the important part of the data to be encoded while discarding others [5]. Lossy compression algorithms can be classified as: spatial, frequency, and time–frequency domain methods. Methods like direct data compression, model based and parameter extraction methods exploit redundancy in the spatial domain [5]. Model based methods such as Neural networks build mathematical models that are losslessly encoded by other coding schemes [6, 7]. Fourier transforms and Discrete Cosine Transforms exploit the redundancy at different energy levels of the transformed ECG signal [8, 9]. Multilevel resolution in both time and frequency domains is provided by wavelet transforms [9]. Wavelets are the most commonly used methods for lossy compression because of the efficient compression and the ability of the transform coefficients to intricately quantify the signal in both time and frequency domains for further analysis [9]. Generally transforms are not recommended for low computational devices such as mobile phones due to higher power consumption, requirement of buffering caused by data dependencies and increased latency.

Lossless methods are encouraged to avoid the possibility of losing biomedical signal artifacts of potential diagnostic significance. Lossless compression methods employ predictive preprocessing methods followed by an encoding scheme. Chua and Fang [10] employed a predictive discrete pulse code modulation (DPCM) and used Go-lumb Rice coding for encoding them along with a highly integrated VLSI design to losslessly compress ECG with very low power consumption. Koski adopted a structural method for recognition of ECG complexes and encoded them using complex extraction Huffman coding for improved compression performance

Fig. 1 An ideal ECG waveform and its features



[11]. Sriraam [12], proposed a correlation dimension based predictor followed by arithmetic coding for EEG signal. Srinivasan et al. [13], employed the correlation dimension based predictor on wavelet transform coefficients and encoded them using arithmetic coding. Takiharo and Morita [14], proposed a novel arithmetic coding scheme based on antidictionaries which bypassed the predictive preprocessing phase and also yielded higher compression ratio. Boucheham [15], on the other hand extended the predictive phase into two stages. Line simplification methods for short-term prediction and Curve simplification methods for long term prediction were used.

Muhopadhyay et al. [16], applied a simple but efficient ASCII based encoding technique after grouping based preprocessing. He also proposed a lossy counter part with a comparatively high quality score [17]. The method adopted the standard ASCII character set with 128 characters of which 33 are non-printable control characters. To avoid loss of data due to use of non-printable characters when encoding with the ASCII table forward and reverse grouping methods are employed to encode the characters using the available 95 printable ASCII characters. Nevertheless compressed ECG suffers data loss when transmitted in SMS due to restricted character set supported by the SMS protocol. Sufi et al. [18] proposed a novel lossy compression scheme using the extended ETSI GSM 03.38 character set defined for the SMS protocol. A computationally light, lossless compression using this character set is required to transmit ECG in real-time telemonitoring systems. Towards achieving the aforementioned goals, a simple and efficient ECG compression algorithm called Lossless ECG Compression using Symbol substitution (LECS) for lossless storage and retrieval, untrimmed transmission via messaging protocols and applicable on low computational devices is proposed. Moreover, the high compression ratio coupled with lossless compression renders the proposed algorithm for ECG archival purposes as well. The remaining portion of the paper is organized as follows: Sect. 2 explains the LECS

Compression algorithm; Sect. 3 details the LECS Decompression procedure. The experimental setup is explained in Sect. 4. The results and discussions presented in Sects. 5 and 6 concludes the work and provides directions for further research.

2 Proposed Lossless ECG Compression Using Symbol Substitution (LECS) Algorithm

Lossless compression of data ensures no data loss and exact reconstruction of original ECG from the compressed ECG. Unlike media files where data loss can be acceptable for the amount of compression offered, ECG and other health data cannot be risked with the loss of data. The proposed technique is simple and efficient. The compression is based on symbol substitution using a symbol table constructed from the 124 printable characters in ETSI GSM 03.38 character set. The use of this reduced character set ensures that data transmitted is not trimmed because of reduced character set supported by SMS protocols. The LECS algorithm involves a series of preprocessing phases followed by ECG Compaction and symbol substitution based encoding. The block representation of the proposed algorithm is given in Fig. 2. The essential steps include Normalization, Differencing, Sign epoch marking, ECG Compaction, Symbol substitution based coding. Normalization and differencing have been the basic preprocessing steps in compression in many ECG Compression methods. In the proposed approach at every phase a compression parameter is appended to the compressed ECG header to aid in exact reconstruction. The compression parameters include normalization coefficient, first value of the normalized ECG signal and Sign epoch marker symbol. These parameters are explained in the following sections.

Each sample of ECG signal is represented by floating values and the plot is shown in Fig. 3a. Normalization is the process of converting these floating point values to integers thus reducing the computational cost required for further processing. This is achieved by multiplying each of the samples by an integer called normalization coefficient (such as 200) as in Eq. 1. The Normalized ECG is shown in Fig. 3b. The normalization coefficient and the first value in the normalized ECG are appended to the normalized ECG signal.

$$ECG_n = ECG_{org} \times NC \quad (1)$$

where ECG_n is the Normalized ECG, ECG_{org} is the Original ECG, NC is the Normalization coefficient.

The normalization process reduces the number of bytes required to store the data from 4 bytes (floating point value) to 2 bytes (integer value). This also reduces the amount of computation needed, by cutting down the floating point operations. This allows the compression algorithm to be successfully implemented on a low computational device such as a mobile phone.



Fig. 2 Block diagram for LECS compression

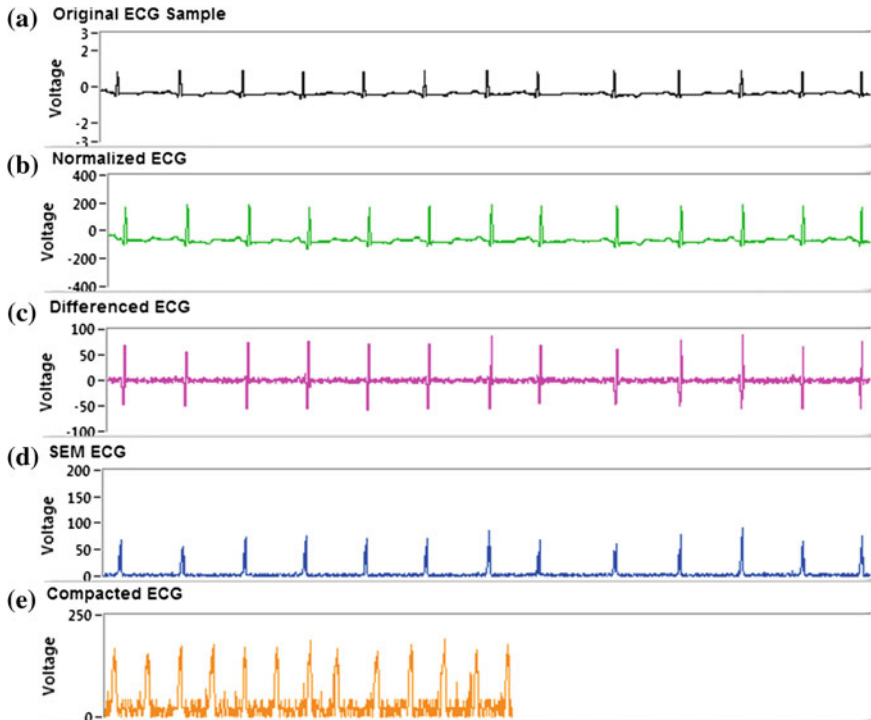


Fig. 3 Compression steps for record 100

The normalized ECG signal differenced as given in Eq. 2. The Differenced Normalized ECG, ECG_d , is shown in Fig. 3c.

$$ECG_d = ECG_n(i) - ECG_n(i + 1) \tag{2}$$

The key process before encoding is the Sign Epoch Marking (SEM). Sign Epoch Marking is done by inserting the Marker symbol at the start of a series of values with the negative sign value. After sign marking the negative values are translated as to positive given by ECG_{st} prior to encoding as shown in Fig. 3d.

The sign translated values represented by ECG_{st} are then encoded using a symbol substitution table. The table may be commonly agreed upon by both sender and receiver to avoid the transmission of the symbol substitution table for every acquired ECG signal. While compacting consecutive pairs of values are taken

from ECG_{st} . To amplify the compression, the ECG is compacted by employing the following steps.

Values are taken in pairs. If both the values in the pair are single digit values, it is then encoded using a single symbol indexed by SS_{ref} (given in Eq. 3) in the symbol substitution table.

$$SS_{ref} = ECG_{st}(i) \times 10 + ECG_{st}(i + 1) \quad (3)$$

When double digit values are encountered in ECG_{st} , each value in the pair (either single digit or double digit) value is encoded using a separate character indexed by SS_{ref1} , SS_{ref2} (Given in Eqs. 4, 5). The Compacted ECG is shown in Fig. 3e.

$$SS_{ref1} = 100 + ECG_{st}(i) \quad (4)$$

$$SS_{ref2} = 100 + ECG_{st}(i + 1) \quad (5)$$

Occasionally occurring three digit values are encoded using the symbol indexed using the whole three digit value prefixed by a fixed symbol used to denote that the symbol has been used to encode a three digit value. The output produced is a text file with symbols used in the symbol table that could be transmitted through SMS.

3 Proposed LECS Decompression Algorithm

Decompression is the exact reverse process of compression as given in Fig. 4. The compressed ECG is first decoded using the symbol table. The indices of the symbols constitute decompressed output. The values are then incorporated with sign using the marker symbol used in SEM step of LECS Compression algorithm. The compression parameters appended in the front of the compressed ECG consist of NC, first value of normalized signal. The first value of the normalized symbol is used for reverse differencing followed by division by NC called Inverse Normalization. The complete reverse of the compression method ensures exact reconstruction of the ECG signal.

4 Experimental Setup

4.1 Data Description

Experimental evaluation was carried out for all ECG samples from the Massachusetts Institute of Technology Arrhythmia database (MIT BIH adb) and MIT Compression Test Database (MIT cdb). The MIT BIH Arrhythmia database contains 48 half hour excerpts of two channel (Modified Limb Leads MLII, V1)



Fig. 4 Block diagram for LECS decompression

ambulatory ECG recordings. The modified limb leads are used since Holter monitoring is carried for extended periods of time (over 24 h), physical activity is obstructed when ECG is acquired using standard limb leads. The modified limb leads are used to avoid the interference and positioned to closely match signals acquired from the standard limb leads. Normal QRS Complexes and ectopic beats are prominent in the modified limbs. MIT ADB includes 23 random recordings and 25 selected recordings to represent clinically significant but less common arrhythmias. Each recording is sampled at 360 Hz, each sample represented by 11 bits over a 10 mV range [19].

4.2 Symbol Table from ETSI Character Set

SMS, MMS protocols support only limited character sets (like ETSI GSM 03.38) wherein even lossless compression methods are susceptible to truncation and data loss during transmission. Generally text in an SMS may be encoded using any of the three available standards: GSM 8-bit character set, GSM 7-bit character set or Unicode UCS-2. The maximum size of an SMS is limited to 1,120 bits in the GSM network. When using GSM 8-bit character each character is encoded using an 8-bit character code. Therefore when encoding using GSM 8-bit character set the maximum size of an SMS is 140 characters. Special characters in SMS text which are not present in the GSM 8-bit character set are encoded using Unicode UCS-2 where each character is represented using 16 bits reducing the maximum character limit of SMS to 70. But it is possible to squeeze in a few more characters into the SMS if the SMS contains only characters available in the GSM 7-bit character set. 160 characters encoded using GSM 7-bit code can be accommodated in the limited 140 8-bit codes. Thus using characters present in GSM 7-bit character set increases the 140 character limit and allows 160 characters to be transmitted in a single SMS. Both GSM 7-bit and GSM 8-bit character sets are defined in the ETSI GSM03.38 standard. Constructing a symbol table with only characters from the GSM 7-bit character set allows more characters to be accommodated in a single SMS and ensures untrimmed data transmission. The proposed method employs a symbol table constructed from 124 printable characters of GSM 7-bit character set defined by the ETSI GSM 03.38 standard to encode the sign translated ECG. Figure 5 shows all the 124 characters used for constructing the LECS symbol table.

Fig. 5 GSM 03.38 based reduced character set for encoding

@ £ \$ ¥ è é ù ì ò ç ø Å å Δ _ Φ Γ Λ Ω
 Π Ψ Σ Θ Ξ Æ æ ß É ! " # ¤ % & ' () * +
 , - . / 0 1 2 3 4 5 6 7 8 9: ; < + > ?
 i A B C D E F G H I J K L M N O P Q R S
 T U V W X Y Z Ä Ö Ñ Ü š ž a b c d e f g
 h i j k l m n o p q r s t u v w x y z ä
 ö ñ ü à

5 Performance Measures

Generally compression schemes are evaluated on two aspects: the compression and distortion performances. The compression performance is quantified as compression ratio, and is defined as the ratio between the number of bits in the compressed signal number of bits in the original signal (As in Eq. 6).

Compression Ratio (CR):

$$CR = n(ECG_{org}) / n(ECG_{comp}) \tag{6}$$

where n() represents the number of bits in original ECG signal (ECG_{org}) and compressed ECG(ECG_{comp}) respectively.

The distortion of the compression scheme is measured by different metrics such as Percent Root-mean-square Distortion (PRD). PRD quantifies the error percentage between the original signal and the reconstructed signal (Given in Eq. 7).

Percentage Root mean square Distortion (PRD):

$$PRD = \sqrt{[(ECG_{org} - ECG_{decomp})^2 / ECG_{org}]2} \tag{7}$$

where ECG_{org} represents samples in the original signal and ECG_{decomp} represents the decompressed signal.

However, in ECG Compression the clinical acceptability of the reconstructed signal plays an important role. The fuzzy and non-algorithmic nature of this factor propels the need for an efficient lossless compression algorithm.

6 Results and Discussions

The LECS compression, decompression algorithm was implemented in J2ME (Java Micro Edition) deployable on low computational devices such as mobile phones and the visualization of results was obtained using LabView8.5. The LECS algorithm for records The LECS was evaluated using all the records from both MIT ADB. The Compression phases and the compressed ECG data for record 100 from MIT ADB are shown in Fig. 6.

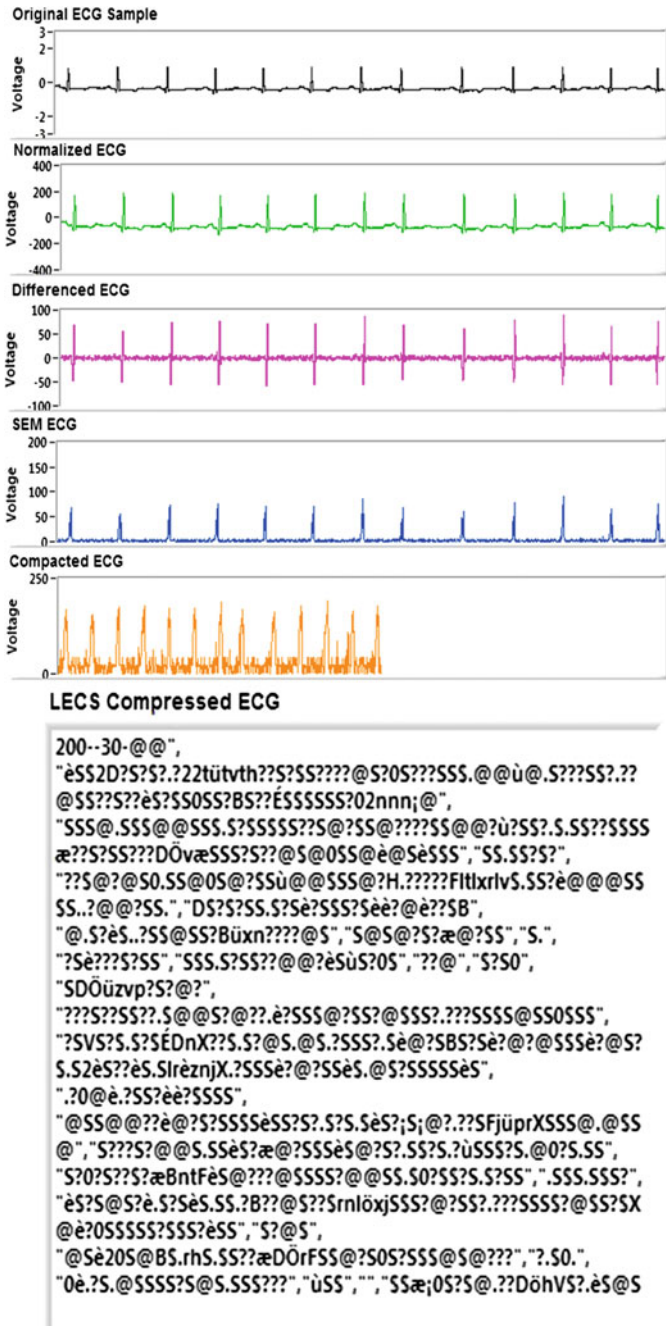


Fig. 6 LECS compression phases for MIT BIH ADB 100

Fig. 7 LECS decomposition—MITBIH ADB 100

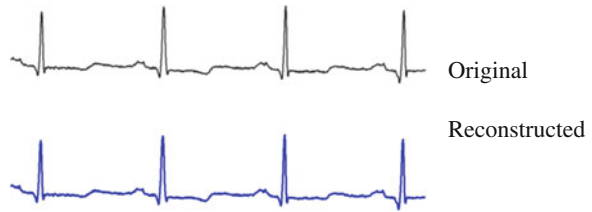


Table 1 Results of proposed LECS for MIT BIH arrhythmia dataset

Record no.	CR	PRD	Record no.	CR	PRD
100	7.42	0.0200	201	7.29	0.0200
101	7.47	0.0224	202	7.45	0.0173
102	7.35	0.0224	203	4.52	0.0224
103	7.29	0.0224	205	7.34	0.0173
104	7.09	0.0265	207	7.35	0.0245
105	7.02	0.0245	208	6.75	0.0173
106	6.92	0.0200	209	7.14	0.0265
107	6.03	0.0332	210	7.21	0.0224
108	7.27	0.0224	212	6.93	0.0245
109	6.74	0.0245	213	6.37	0.0245
111	6.99	0.0200	214	6.91	0.0141
112	7.35	0.0200	215	6.83	0.0316
113	7.17	0.0265	217	6.38	0.0141
114	7.41	0.0173	219	7.15	0.0224
115	7.38	0.0173	220	7.34	0.0224
116	6.78	0.0141	221	7.17	0.0173
117	7.32	0.0224	222	7.45	0.0200
118	6.61	0.0141	223	7.18	0.0300
119	7.11	0.0173	228	7.17	0.0283
121	7.39	0.0173	230	7.09	0.0173
122	6.90	0.0200	231	7.19	0.0200
123	7.41	0.0173	232	7.31	0.0173
124	7.25	0.0245	233	6.76	0.0141
200	6.86	0.0200	234	7.27	0.0200
Avg CR	7.04		Avg PRD		0.0211

The Comparison of original ECG signal with the reconstructed signal for the records 100 is shown in Fig. 7.

The different performance measures for records in MIT ADB (48 records) are given in Table 1. The maximum compression ratio of 7.47 was achieved for record 101. The Minimum compression ratio of 4.52 was obtained for record MIT ADB 203.

The comparison of existing near lossless, lossless with proposed algorithm is presented in Table 2. Sriraam [12], achieved an average CR of 3.23 with PRD of 0.02. Mukhopadhyay et al. [16] achieved a CR of 7.18 and PRD of 0.023 and

Table 2 Comparison of proposed LECS with existing algorithms

Author	Signal	Dataset	Algorithm		CR	PRD
			Prediction	Coding		
Sriraam [12]	EEG (12 bit)	University of Bonn & Acquired	Correlation dimension based Prediction	Arithmetic coding	3.23	0.02
Mukhopadhyay et al. [16]	ECG	PTB	Differencing, normalization grouping	ASCII based encoding	7.18	0.023
Boucheham [15]	ECG (11 bit)	MITBIH ADB	Line, curve simplification based prediction	–	7.71	0.025
Proposed	ECG (11 bit)	MITBIH ADB	Differencing, normalization & sign epoch marking	Symbol table based encoding	7.04	0.0211

Boucheham [15] achieved a CR of 7.71 and PRD of 0.025. Although the compression ratio of Mukhopadhyay et al. [16] and Boucheham [15] are higher than the proposed LECS algorithm the Percentage Root mean square Distortion (PRD) of the proposed algorithm is lower yielding higher quality in the compressed ECG.

7 Conclusions

This paper presents a simple yet efficient lossless compression algorithm using symbol substitution based encoding. The performance of the proposed scheme is evaluated using all records from MIT BIH Arrhythmia database and MIT BIH Compression Test database. The method performs well with both normal and abnormal ECG data, and achieves good CR with low distortion, preserving the morphological features in the reconstructed signal. The simplicity of the technique renders it compatible in low-computational devices such as mobile phones. The symbol table constructed using GSM 7-bit character set ensures lossless transmission through SMS in GSM Networks. Compared with some of the existing methods, the higher compression ratio coupled with higher quality and simplicity can be efficiently put to use in telecardiology and other remote monitoring systems.

Acknowledgments The authors acknowledge their cordial gratitude to Dr. Moses. K. Daniel, Professor and head of Medicine Department, Government Rajaji Hospital, Madurai, India for the valuable clinical advice and the effort and time spent in evaluating the quality of numerous reconstructed ECG samples.

References

1. World Health Organization Information.: http://www.who.int/gho/publications/world_health_statistics/2012/en/index.html
2. Sufi, F., Khalil, I.: Enforcing secured ECG transmission for realtime telemonitoring: a joint encoding, compression, encryption mechanism. *Sec. Comm. Netw.* **1**, 389–405 (2008)
3. Huffman, D.A.: A method for the construction of minimum-redundancy codes. *Proc. IRE* **40**, 1098–1101 (1952)
4. Welch, T.A.: A technique for high-performance data compression. *Computer* **17**, 8–19 (1984)
5. Jalaleddine, S.M., Hutchens, C.G., Strattan, R.D., Coberly, W.A.: ECG data compression techniques-A unified approach. *IEEE Trans. Biomed. Eng.* **37**(4), 329–343 (1990)
6. Zigel, Y., Cohen, A., Katz, A.: ECG signal compression using analysis by synthesis coding. *IEEE Trans. Biomed. Eng.* **47**(10), 1308–1316 (2000)
7. Fira, C.M., Goras, L.: An ECG signals compression method and its validation using NNs. *IEEE Trans. Biomed. Eng.* **55**(4), 1319–1326 (2008)
8. Reddy, B.R.S., Murthy, I.S.N.: ECG data compression using Fourier descriptors, *IEEE Trans. Biomed. Eng.* **BME-33**(4), 428–434 (1986)
9. Batista, L.V., Kurt Melcher, E.U., Carvalho, L.C.: Compression of ECG signals by optimized quantization of discrete cosine transform coefficients. *Med. Eng. Phys.* **23**, 127–134 (2001)
10. Chua, E., Fang, WC.: Mixed bio-signal lossless data compressor for portable brain-heart monitoring systems. *IEEE Trans. Consum. Electron.* **57**(1), 267–273 (2011)
11. Koski, A.: Lossless ECG encoding. *Comput. Methods Programs Biomed.* **52**(1), 23–33 (1997)
12. Sriraam, N.: Correlation dimension based lossless compression of EEG signals. *Biomed. Signal Process. Control* **7**(4), 379–388 (2012)
13. Srinivasan, K., Dauwels, J., Reddy, M.R.: A two-dimensional approach for lossless EEG compression. *Biomed. Signal Process. Control.* **6**(4), 387–394 (2011)
14. Takahiro, O.T.A., Morita, H.: On-line electrocardiogram lossless compression using antidictionary codes for a finite alphabet. *IEICE Trans.* **93-D**(12), 3384–3391 (2010)
15. Boucheham, B.: ShaLTERR: a contribution to short and long-term redundancy reduction in digital signals. *Sig. Process.* **87**(10), 2336–2347 (2007)
16. Mukhopadhyay, S.K., Mitra, S., Mitra, M.: A lossless ECG data compression technique using ascii character encoding. *Comput. Electric. Eng.* **37**(4), 486–497 (2011)
17. Mukhopadhyay, S.K., Mitra, S., Mitra, M.: An ECG signal compression technique using ASCII character encoding. *Measurement* **45**(6), 1651–1660 (2012)
18. Sufi, F., Fang, Q., Khalil, I., Mahmoud, S.S.: Novel methods of faster cardiovascular diagnosis in wireless telecardiology. *IEEE J. Sel. Areas Commun.* **27**(4), 537–552 (2009)
19. Goldberger, A.L., et al.: PhysioBank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000)

Disaster Relief Operations and Continuous Aid Program in Human Supply Networks: Are they congruent?—An analysis

V. G. Venkatesh, Rameshwar Dubey and Sadia Samar Ali

Abstract Humanitarian supply chain and logistics has succeeded in attracting research attention in recent years as the special field of attention. Purpose of this paper is to introduce the difference between the Continuous Aid programmes from the disaster relief chains in a Humanitarian Supply System. So far the literature was so biased towards only Disaster relief chains. Through the case study and other research, this paper is to establish and advocate that all Humanitarian Supply networks are not having the disaster management orientation and also to establish scope for further discussions and analysis to the area of Continuous aid category. This paper is primarily the conceptual framework for researching the internal operations strategy of any human Supply Network Operations. It reviews the existing literature in Humanitarian supply network and disaster relief measure to establish the current meaning of Humanitarian Supply network. It adopts the micro case based approach to authenticate the concept of including the Continuous Aid programs in “Humanitarian Supply network” which is currently biased towards Disaster relief management. The paper has found the usage of Humanitarian Supply network term in a complement way with Disaster Relief operations/management so far. However, there are operations which do not support the disaster relief measures; On the other hand—it supports the livelihood of the society. Those operations should also be called and classified under the Continuous Aid Category under Humanitarian Supply network operations. Further research and arguments are indispensable to differentiate the Operations dynamics of

V. G. Venkatesh · R. Dubey

Symbiosis Institute of Operations Management (Constituent of Symbiosis International University), Plot No. A-23, Shraavan Sector, CIDCO, New Nashik 422008, India
e-mail: Venkatesh.VG@siom.in

R. Dubey

e-mail: rameshwardubey@gmail.com

S. S. Ali (✉)

Fortune Institute of International Business, Rao Tula Ram Marg, Vasant Vihar 110057, New Delhi, India

e-mail: sadiasamarali@gmail.com

Continuous Aid Program from the disaster relief management. The discussions can help the re-orientation of research to strengthen and streamline the definitions and areas of Human Supply network. The conceptual paper is analyzing the operations of Humanitarian Supply Chains in detail and gives a new dimension and school of thought with a definition towards Humanitarian Supply networks. This is the original paper written based on the author's experience with NGOs operating in this field. The concept "Humanitarian Supply Networks" should be established and explored carefully, not to be used extensively only on the disaster relief measures, based on the facts established through this paper. The points discussed in the paper will help to widen the discussions on the Humanitarian Supply Networks field.

Keywords Humanitarian supply network · Disaster relief operations · Competitive strategy · Humanitarian logistics

1 Introduction

Humanitarian supply network organizations work to get relief services to areas of need in the shortest possible time. Disaster relief organizations, such as the International Red Cross Society, need to move quickly and effectively. This requires agile supply chains and logistics systems. The pursuit of this agility has attracted many researchers, with the development of new theories and models [1]. However, many, such as [2] argue that fragmentation and discontinuity rather than agility and effectiveness characterize humanitarian organizations' interventions during and after catastrophes.

The literature makes it clear that any firm involved in the collection and distribution of material to areas of catastrophe is part of the humanitarian supply network and such logistics account for 80 % of disaster relief [3].

Kovacs and Spens [1] shows that the challenges of humanitarian logistics depend on the types of disasters and the activities related to different phases of disaster relief. This paper argues that organizations that collect and distribute material to areas of ongoing need also need to be recognized as part of humanitarian supply networks. The objectives of such continuous aid programs and their supply chain networks are different to disaster relief networks. Based on a review of the literature around humanitarian supply networks and a case study, the paper argues that lessons from continuous aid programs could improve the speed and effectiveness of disaster relief operations.

One of the biggest challenges that organizations face relates to the resilience and effectiveness of their supply chains. For "ordinary" firms, such resilience and effectiveness can be the source of temporary competitive advantage. In times of recession and slowdown, the ability to deliver a wide range of items and services to customers at the requested times and places can maintain and increase market

share. Firms that fail to adapt to gradual structural shifts and firms whose supply chains are unresponsive to market needs lose. Timeliness and market responsiveness are vital factors in business success and failure.

According to Balcik and Beamon [4], for humanitarian organizations both the challenges of timeliness and the costs of failure are amplified by the unpredictability of demand, the suddenness of need, the high human costs involved and the lack of resources. The magnitude of the challenges underlines how much logistics and supply chain operations are pivotal in humanitarian relief and rescue operations—they determine the effectiveness and efficiency of the disaster response and relief operations.

2 Humanitarian Supply Chain: What the Researchers Say so Far?

The logistics of disaster relief and the workings of humanitarian supply chains interest researchers and practitioners and the literature is growing rapidly. The aim of this paper is to establish a re-engineered perspective on the emerging orthodoxy around humanitarian supply networks. The literature equates such networks with disaster relief: the speedy delivery of relief material to areas of acute need in a moment of crisis. Such relief management systems have been investigated well by researchers.

However, this paper argues that humanitarian supply networks could also include the delivery of material to areas of chronic need over time. For example, Thomas and Kopczak [5] define humanitarian logistics as “the process of *planning, implementing and controlling the efficient, cost-effective* flow of relief goods and materials/aid as well as related information, from *point of origin to point of consumption* for the purpose of meeting the ultimate requirements of the *end beneficiary in a situation of humanitarian concern*” (emphasis added). Such a definition allows for the inclusion of networks that operate to meet chronic rather than solely acute need. The paper argues that such continuous aid programs have both different outlooks and operations compared with disaster relief networks and chains of supply.

Ertem [6] outlines the difference between the commercial SCM and humanitarian supply network. The general logistics literature has tended to focus on leaner supply chains through improving efficiency and reducing costs [7]. More recently, the emphasis has shifted towards more agile supply chains for highly dynamic environments with more emphasis on innovation and responsiveness [8]. The analysis has been made in the lines of lead time, competency requirements, inventory requirements, performance measurements and other important points. There is a clear demarcation in terms of inner operations between the Direct or Normal supply networks and humanitarian supply network.

Such questioning is inline with Glenn's (2009) conclusion that everything from agility to working in partnership is under discussion in research around humanitarian networks. This partly fueled by comparisons between disaster relief and normal business practices and conditions. For example, [5] defines the humanitarian supply chain as:

The process of planning, implementing and controlling the efficient, cost-effective flow and storage of goods and materials, as well as related information, from the point of origin to the point of consumption for the purpose of alleviating the suffering of vulnerable people. The function encompasses a range of activities, including preparedness, planning, procurement, transport, warehousing, tracking and tracing, and customs clearance.

This perspective presents disaster relief as if it was the practice of logistics in unexceptional circumstances and conditions. However, disaster relief, by definition, should concern itself with the impact of unpredictability. As Chang et al. [9] noted, generally, the predictability of a disaster plays a role in the possibilities of responding to it and the occurrence of particular types of disasters can be tied to specific geographical areas. Humanitarian supply networks are concerned to work with the predictable unpredictable. And, according to Sowinski [10] learning from humanitarian logistics is important for business logistics and commercial supply chains, as disaster relief operations show how to manage unpredictable environments.

Thomas and Mizushima [8] notes that humanitarian logistics is a bridge between the disaster preparedness and disaster response, between procurement and distribution and between headquarters and the field. Logistical coordination is needed between humanitarian organizations and within the humanitarian supply chain [2, 11, 12].

For example, in a typical humanitarian relief mission, large stocks of a wide range of goods (water cans, tents, canned foods, clothes, blankets, cots, bedding, hygiene items and medical equipment and supplies) are inventoried in warehouses for distribution to the end beneficiaries (people affected by the calamity or natural/artificial disaster). These supplies need to be protected and transported. They also need to be part of a distribution system with reverse logistics capabilities.

Most of the literature strongly supports the idea that humanitarian logistics involves moving materials to the disaster prone area (in terms of practical/real time experience) rather than to people, communities places where the need is great. Kovacs and Spens [1] argues that humanitarian supply chains are grounded in disruptions and natural disasters that demolish markets, not simply ongoing poverty or need. They emphasize governments' role in handling the humanitarian supply network. Humanitarian logistics contributes most to disaster relief. Most reliable estimates put logistics accounting for at least 80 % of the costs of disaster relief efforts [12].

Given this emphasis, there is little discussion in the literature of supply networks operated by Non-Government Organizations (NGOs) to service the chronic rather than acute needs of communities. This is despite the work of [8] which does not differentiate between the supply networks operated in different target

environments. Continuous aid can be part of the humanitarian supply network and part of disaster relief operations and depends on the work of NGOs, argue [4]. Even so, there is no specific recognition for the NGOs who work to meet the chronic and ongoing development needs of the people. Whilst they do not feature in the literature, they act as mediating agencies or enablers for entire humanitarian supply networks.

Olorunfoba and Gray [2] cite [13] classification of humanitarian aid organizations and NGOs into two broad types of activities:

- (a) **Relief activities:** Relief for victims of large-scale emergencies. These short-term activities focus on providing goods/aid (such as food, water, shelter, clothing, bedding) and rehabilitation services to minimize immediate risks to human health and survival.
- (b) **Development activities:** These are longer-term aid activities that focus on community self-sufficiency and sustainability. These activities include, establishing permanent and reliable transportation, health care, housing, and food.

This paper argues that there needs to be greater recognition of and attention given to the various segments of humanitarian supply networks (minus the disaster relief operations). It identifies two new fields of research based on the services offered by humanitarian organizations/NGOs in mitigating disasters and managing social development.

3 A Re-look into the Typology of Humanitarian Supply Chains

Kovacs and Spens [1] argue that there are two streams of humanitarian logistics identified in the literature: continuous aid work (without any disaster orientation) and disaster relief. However, the different characteristics of these two relief chains have not been established or given much detailed attention. This gap allows scope for this paper to develop such a conceptual framework based on a detailed literature review and interviews and discussions with humanitarian logistics and supply chain managers. The conclusions are demonstrated through a micro case study of an Indian NGO.

It is very important to define a typology the humanitarian supply chains based on the views of participants and the realities of the operating environment. This paper places the entire humanitarian network into a architecture called Disaster Relief Chains and Continuous Aid Programs (CAP) and identifies two different versions of continuous aid supply chains: Disaster oriented Continuous Aid Programs (D-CAP) and Need/societal based Continuous Aid programs (N-CAP).

Kovacs and Spens [1] differentiates the participating organizations into various types based on the operations are handled by them. These go from supranational

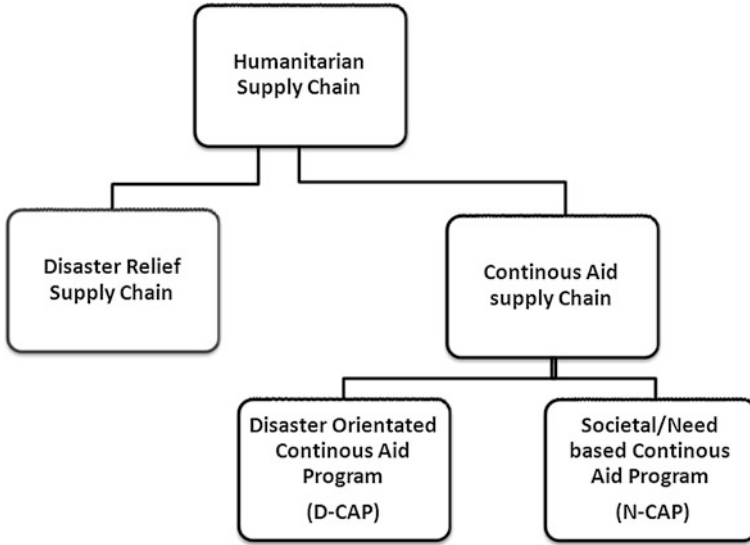


Fig. 1 Proposed new typology of humanitarian supply chains

aid agencies (such as UN agencies) and governmental organizations (GOs) to big international non-governmental organizations (BI-NGOs) and non-governmental organizations (NGOs) (Fig. 1).

3.1 Disaster Relief Supply Chains (DRC)

According to the UN Charter:

Disaster is described as an event that causes a serious disruption of the functioning of society, posing a significant, widespread threat to human life, health, property or the environment, whether caused by accident, nature or human activity, and whether developing suddenly or as a result of complex, long-term processes.

[29] and [14] present clear classifications of disasters. They argue that sudden natural disasters cannot be assessed in the same ways as most man-made disasters, as there is the possibility of preparedness. Many aid agencies are focused on relief after natural disasters [20] and for many of them the disaster is the trigger for the design and operation their entire supply chain and their use of various strategies to manage the impact of the disaster. Any disaster relief chain operates, in principle, purely on the service perspective to restore normal life in the affected area. During this process, it is dependent on its agility to collect the materials and respond to need at the site of action, argue [2].

The character of the material supply chain is more reactive in the case of relief delivered through non-established networks. They tend not to be selective in product portfolio handling and to exhibit intrinsic imbalances in the supply–demand equilibrium of the material supply chain. This phase of the operation is highly dependent upon the assessors/surveyor’s initial assessment report and there is little possibility of standardization for material handling or operations along the entire supply chain, due to the complexity of the operations and lack of appropriately trained personnel in this field of operations. The nature of relief material (food, medicinal goods and supplies) is such that they require greater attention in terms of appropriate storage and safe transportation. In this, DRC shows a better response than D-CAP and N-CAP.

There is a huge literature available on the operations of various operations in this category (sudden disaster relief chains), but the focus of this paper are the continuous aid programs/operations of NGOs and similar organizations, where there established networks of communication, supply and demand prior to any sudden disaster.

3.2 Continuous Aid Supply Chains (CAS)

Supply chains can respond differently to disasters based on their preparedness to face the situations. These NGOs and BINGOs are well exposed to their supply chain systems and delivery formats to the site of disasters. These factors help networks make agile responses to disasters, when active material collection and distribution is triggered. This type of supply network we term a Disaster-oriented Continuous supply program (D-CAP).

At the same, all the humanitarian supply is not meant for disaster relief. There are some networks, predominantly operated by NGOs, established to meet people’s ongoing basic livelihood needs. In this paper, these are termed Need based/ societal Continuous Aid Programs (N-CAPs) and presented as a separate category, with different characteristics of operation and delivery to disaster relief chains (DRCs). The overall argument is that continuous aid supply chains allow for the systematization of humanitarian operations, stronger forecasting and more effective analysis. They also offer the possibility of a completely different outlook on handling and researching operations. The following sections detail the characters of both D-CAP and N-CAP to support these conclusions.

Chandes and Paché [15] argues that it is impossible to have a standardized environment for supply chain operations in humanitarian aid and cite the Kashmir earthquake as an example. This paper argues that this need not be the case if there is a shift towards a model based on principles derived from a continuous aid environment, where there are standardized procedures for distributing aid to the needy. Such a shift would allow for greater speed/turnaround time and allow the inclusion of all contemporary practices in SCM, such as Vendor/Donor Managed Inventory (VMI) and innovative strategies in attracting the materials to the system.

The available evidence indicates that the DRC gives less scope for such standardization, experimentation and innovation.

3.2.1 Disaster Oriented Continuous Aid Program (D-CAP)

Many NGOs/BINGOs investigate disasters and their impacts using sophisticated technologies and well-researched reports. Such investigations help all partners to face the situation and establish the network to respond to it. This paper contends that disasters can be well prepared for where agencies' activities have a proactive aspect and have material collection programs throughout the year. The system might only respond after the disaster, but would be supported by much longer preparation time than that characteristic of the DRC model. This significantly helps the standardization of material handling along the supply chain, as the throughput time is really high.

Tomasini and van Wassenhove [16] said that the resources available in the D-CAP and DRC are unknown in the system and that suppliers' contribution is unpredictable. Simpson [17] argues that this creates redundancy and duplication. But NGOs may take proactive steps to cover up and lessen the negative impact created by the D-CAP and DRC approaches.

The drivers for this supply chain are entirely different from DRC. Forecasting and material handling are its core drivers and there can be some selectivity in terms of the material needs and to the product portfolio. The advantage is that it is possible to match any product to the disaster site, irrespective of preferences, as any material can be put into the system without any restriction. There is also space for products to be re-engineered due to its larger band in the throughput time.

This system is as dependent as DRCs on the surveyor's report on the impact of the disasters. It is difficult to predict the quantities of material to be loaded in the supply chain, as this depends on the impact of the disaster. However, the material handling costs will be at the higher level, as storage costs are involved in the pre-determined place as a part of the preparedness.

D-CAP can create savings in the supply chain that are not practical in the DRC model. The triggered relief measures can be operated professionally in D-CAP, as they have pre-processed goods, and this can reduce material handling costs. One of the areas of research could be the areas of costs under NGO distribution operations. For D-CAP, disaster chart analysis (including forecasts of possible disasters) is a vital tool for the distribution and handling of materials. D-CAP supply chains are handled by the professionals in the field with a cost saving orientation and more innovation in product handling.

Where NGOs assess their total distribution costs and coverage, they can take a cautious trade off approach to enter into DRC or D-CAP. The key considerations that may help them to decide are:

- Overhead costs
- Coverage area

- Total cost of distribution
- Cost of supplies
- Number of predicted donors
- Material handling ability
- Trained personnel to handle the inventory and
- Potential inventory investments (including all setups and material handling).

For example, the International Federation of Red Cross and Red Crescent Society (IFRC) has established its Regional Logistics Units (RLUs) in Kuala Lumpur and Dubai. They cover the entire South East and Middle—East Asia regions respectively during any major natural disaster. Although, this pre-positioning gives them the ability to get material to the site, they are still dependent on the assessor/surveyor's report and the supply chain managers. Even so, the D-CAP model allows for the testing or application of best practices to supply chains, less of an option with DRC. For example, the best distribution practices exist at International Red Cross in Forecasting, Fleet operations Monitoring and Humanitarian Logistics System (HLS), best Coordination practices are all core operations that can be implemented in D-CAP and N-CAP too. However, even these initiatives do not resolve the issue of standardization across all phases of supply chain.

3.2.2 Need Based/Societal-Continuous Aid Program (N-CAP)

The next category in the humanitarian framework is the Need based/Societal Continuous Aid Program (N-CAP). This research shows that humanitarian networks have not taken on board this concept. N-CAPs tend to be run by NGOs with strong social distribution networks. NGOs that collect material throughout the year according to emerging social needs can be seen as part of N-CAP supply chains. N-CAP exhibits many characteristics that differentiate it from DRC and D-CAP. Some of these are described below.

Demand forecasting can be very accurate and the products necessary to meet particular demands can be sourced and provided. N-CAP assessments can be precise, as supply chain programs can be monitored and controlled and there is a high degree of delivery flexibility in terms of planning and distribution, as they are triggered by any disaster/event. The profit outlook can also be considered and calculated rather better than with DRC and D-CAP. Even though in DRC there are fewer holding costs and in D-CAP there is a close monitoring step on the same.

But in N-CAP, collection and delivery can be planned through assessments at the various phases on the cost of the operations. The demand–supply equilibrium can be well maintained through continuous supply and close follow ups with the donors. There can also be a higher degree of control over the quality and quantity of the products and a greater probability that they can be delivered according to the needs than there are with the D-CAP and DRC models. There is evidence that the DRC and D-CAP systems can include non-useful material. The N-CAP approach

allows NGOs both to raise and distribute existing materials necessary on the ground and to innovate and re-engineer products into the system.

Steps of quality check and grading processes are the unique differentiating and advantage operations in N-CAP operations. These can result in better efficiency in terms of usage of the product. Less coordination is required, as everything can be well organized/planned prior to any actual disaster and distribution planning can be very precise. Above all, there is no pressure in N-CAP to retrieve the livelihood in comparison with the magnitude of the operations handled by DRC and D-CAP.

This is very similar to the normal supply chain, except for the quantity of material and the practice of inventory. Still, it carries the uncertainties in terms of material supply, which has to be focused differently. It also offers the flexibility of handling the disaster as the DRC lead member. In Tsunami 2004 event, many of the NGOs took the charge of the particular coastal areas in Indonesia, India and Sri Lanka. Along with D-CAP, N-CAP offers a clearer platform than DRC to assess their performance through the metric system, due to its transparency and control in the entire supply chain. A supply chain based on N-CAP offers checks on inventory value and returns, spoilages, cost spent on the particular aid recipient, donor/acceptor—amount received per time and year-long fill rate achievement. It gives the room for the application of IT to improve the processes and increase the efficiency of the system, unlike the DRC and D-CAP models.

The attributes of the networks can be analyzed using Charles' [18] frame work. Charles identifies various attributes of the supply chain from control through to flexibility and the comparison highlights the overall gains in efficiency, effectiveness, economy and equity of N-CAP over both DRC and D-CAP (Table 1).

4 *Goonji*: Micro Case Study on N-CAP

The case study focuses on an Indian NGO, *Goonji*. Based in New Delhi, it was established in 1998 to deliver “clothing for all”. It now acts as the lead organization/initiator in a humanitarian network that transfers used clothing and household goods from India's rich to its poorest communities. Its story facilitates an understanding of N-CAP and its place in the entire supply chain system.

Goonji is primarily focused on distributing materials to needy people through a humanitarian supply chain system supported by more than 250 partner organizations, including Indian Army. 75 % of these are primary partners who work with *Goonji* throughout the year and the remaining 25 % give support during disasters.

Initially, *Goonji* collected and distributed clothes to remote villages of the interior. It founders felt that NGOs often forgot the basic need for clothing as part of their disaster relief measures. However during 2001 Gujarat earthquake and the 2004 Tsunami relief operations, NGOs were flooded with clothing surpluses that were not needed at the disaster sites, whilst *Goonji* distributed their continuously collected product in the far flung villages of India with the dignity and pride.

Table 1 Supply chain attributes of human supply chain network

NGO continuous aid operations—Human supply chain network	
Supply chain attributes	N-CAP D-CAP and DRC
Supply Chain Control	Costs and delivery can be well controlled by 'fine tuning' factors and the delivery planning can be handled/optimized accordingly
Volume flexibility	Profit orientation can be the priority The volume handled depends upon the cost optimization and distribution planning. There is a greater possibility of effective demand forecasting and alignment of resources
Supply system flexibility	Needs can be met efficiently as delivery from the supply system partners can be completely programmed. It allows some flexibility in handling the materials in terms of their variety and usefulness in the network Fewer coordination mechanisms are required because of the set planning system of logistics
Supply chain reactivity	The response can be called reactive compared with DRC. This works in a passive way. It is relatively easy to predict and to control the response. The impact and pressure of target delivery schedules are lower than DRC
Flexibility product portfolio	Assessments for the goods needed are done and collection programs can be designed accordingly. Specific products can be put in the demand system and collected throughout the year. This allows for more concentration on the material handling part in tune with the needs of the society. Lead time consideration in collecting products cannot have that impact on the collection program

Cost cannot be controlled. It is a secondary priority to the delivery of material to the points of need. This is irrespective of the various constraints available in the market

Delivery orientation can be the SOLE priority
Due to the high levels of uncertainty, effective demand forecasting is difficult. Demand is often determined and fixed post-event by surveyors' assessments. There is room for errors of judgment

Response to disasters may preclude considerations of flexibility in terms of collection according to the product type and variety. It can give freedom to control to certain extent of the supply system by playing with the collection mechanisms

Strong coordination of participating members is essential to set the logistics

There is a need for proactive steps for mobilizing the goods before in disaster prone areas and/or their quick re-positioning for delivery to the relief area

There is some flexibility in terms of collections. However, decisions about the product portfolio require detailed attention so that they can be managed in tune with the lead time available and extent of the need

The target customers for *Goonji* are the rural poor, who are often unable to afford two per day let alone fresh clothing. It also works to supply poor school children with stationary through an initiative largely managed by urban school children.

The nature of poverty in the remote villages encouraged *Goonji* to deal with their social needs in a different way. It wanted to establish a humanitarian supply chain system by partnering with the Channel (for collection and distribution) partners wanted to charge for the clothes that it distributed. It needs to be emphasized that both of these decisions are new in the Indian context: most NGOs distribute material free of charge and deal with disaster relief or different forms of social development.

Goonji has developed an extensive network of initiatives to collect clothing and materials in partnership with major retailers, manufacturers, volunteering organizations, schools and colleges from all parts of India throughout the year. There are specific and seasonal campaigns to maximize the resource collection before and after climatic seasons and there are more than 8 distribution and processing centers.

Goonji's track record of effective partnerships and flexibility has meant that it also acts as the leading NGO in many DRC and D-CAP activities, such as responses to the 2004 tsunami and annual flood relief measures in Northern India.

4.1 Network Operations

Goonji has built up a network of partners across twenty-one Indian states and holds yearly summits for them. It maintains its partner list as a dynamic document in its network operations and take great care in selecting them. Potential partners are usually assessed on the basis of their credibility in the local community, media reports, strong local references and previous local program implementation experience. *Goonji's* works with partners because it needs their strengths in particular localities and to provide a distribution infrastructure to reach far reaching communities across a much wider range than it could manage by itself.

Whilst *Goonji* is active 365 days a year collecting materials, their distribution system works in a different way. It has a centralized warehouse and processing center, where materials are sorted, graded, reengineered into new products and packed for distribution. The center processes 80 tons of materials (clothing, stationery and furniture) every month. Re-engineering is an important aspect of this process. For example, unusable clothes are converted into cloth-based sanitary napkins. 1.8–2.4 million of these are made per year. Newspapers are converted into paper bags and note books re-engineered from one side used papers from offices.

In *Goonji* forecasting works on the basis of a continuous feedback system with partner organizations. The partners work in particular locations to assess material needs. The details can be matched with the forecasting figures for the supply

system studying the details of village and its needs based on the population and age-group. The figures are always validated before any distribution exercise by site visits from *Goonji* volunteers and the partner. The partner NGOs are responsible for the administration of the distribution network, whilst *Goonji* works to establish its supply chain system to act on the inventory management and its backlog maintenance. *Goonji* keeps communicating with partners about their warehoused materials and encourages them to explore possibilities for their use. Every year *Goonji*'s performance is calculated based on the number of programs implemented in a particular region, the total amount of materials distributed against the goods received and the turn over times in managing supply requests from partners. A key feature of their work is continuous nature of their collection, communication, assessment and distribution. This allows for better planning and more responsive and effective interventions.

Another important feature of this humanitarian network that distinguishes it from DRC and D-CAP is that the materials are NOT free to beneficiaries. They have to execute the social and necessary project in their community area and in turn, they get the material for their contribution in work. The school materials are distributed in a similar way. The way in which the product portfolio is managed is highly relevant here. It closely monitors distribution costs through better planning and the use of domestic trucking companies (such as, Transport Corporation of India, AFL and Safe Express). The service providers give a reasonable discount for *Goonji*'s distribution operations and *Goonji* passes on these costs to local partners.

4.2 Learning and Discussions from Goonj

Many NGOs operate humanitarian supply networks only during disasters. They are judged on their ability to handle and operate the relief supply chains. However, operations like *Goonji* offer evidence that there should be the separate category of relief supply chain that operates continuously. It could be seen as a nodal or Mother NGO according to Spens framework [1] (Table 2).

The above points support the recognition of *Goonji* in the N-CAP framework and underline the importance of a separate field for research in humanitarian supply chain management where continuous aid programs have different operating objectives to disaster relief chains.

5 Concluding Remarks

The research field of humanitarian logistics is undergoing strong and dynamic development, as evidenced by the recent increase in publications around the humanitarian supply chain system. However, those research outputs concentrate on disaster relief chains and that leaves scope for the analysis of continuous aid

Table 2 Analysis of *Goonji* network under humanitarian network characteristics

Aim of the network	To serve the basic clothing and materials to the mankind
Actor structure	<i>Goonji</i> has more than 250 partners, of which 75 % have the primary purpose of supporting its humanitarian supply network
Set-up phases	Collection, sorting, re-engineering and distribution to the villages
Basic features	The goods are not given freely. They are exchanged for work rather than money The service operates throughout the year to match the needs of deprived communities with available resources The infrastructure also has disaster relief capabilities
Supply chain philosophy	The supply chain works through the ongoing coordination of supply with demand by continuous communication with partners and necessary adjustments
Time effects	The appropriate and timely distribution of goods meets ongoing social and material needs, but does not have the ‘impact’ of DRC
Control aspects	Demand and transparency in the supply chain are well controlled. Donors can be informed of product movement at any stage and costs are also monitored very closely

programs (CAP) as a separate field of research. This paper is the first step towards a separate category of humanitarian network and discussions around NGO participation in such networks.

There are few characters that can be established for the continuous aid operations. Nevertheless, the paper opines the fact of the established facts of the Humanitarian Supply Chain framework and tried to analyze the entire framework in a different angle aimed for the knowledge development in the particular field of research. Further, the paper can lead to the discussions of establishing the Key performance metrics/indices for the Continuous aid operations through the N-CAP model. Further, it would leave the argument of calling all NGO operated relief chains to be put under “Humanitarian Supply Chain Operations”. It welcomes the discussions from the practitioners across the world for the useful deliberations, considering the importance of the discussions put forwarded in this paper.

References

1. Kovacs, G., Spens, K.M.: Identifying challenges in humanitarian logistics. *Int. J. Phys. Distrib. Logistics Manage.* **39**(6), 506–528 (2009)
2. Oloruntoba, R., Gray, R.: Humanitarian aid: an agile supply chain? *Supply Chain Manage.* **11**(2), 115–120 (2006)
3. Trunick, P.A.: Special report: delivering relief to tsunami victims. *Logistics Today* **46**(2), 1–3 (2005)
4. Balcik, B., Beamon, B.M.: Performance measurement in humanitarian relief chains. *Int. J. Public Sect. Manage.* **21**(1), 4–25 (2008)
5. Thomas, A., Kopczak, L.: *From Logistics to Supply Chain Management: The Path Forward in the Humanitarian Sector*. Fritz Institute, San Francisco, CA (2005)

6. Ertem, A.M., Buyurgan, N., Manuel D.R.: Multiple-buyer procurement auctions framework for humanitarian supply chain management. *Int. J. Phys. Distrib. Logistics Manage.* **40**(3), 202–227 (2010)
7. Lee, H.L.: The triple-a supply chain. *Harvard Bus. Rev.* **82**, 102–112 (2004)
8. Thomas, A., Mizushima, M.: Logistics training: necessity or luxury? *Forced Migr. Rev.* **22**, 60–61 (2005)
9. Chang, M.-S., Tseng, Y.-L., Chen, J.-W.: A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. *Transp. Res. Part E* **43**(6), 737–754 (2007)
10. Sowinski, L.L.: The lean, mean supply chain and its human counterpart. *World Trade* **16**(6), 18 (2003)
11. Altay, N., Prasad, S., Sounderpandian, J.: Strategic planning for disaster relief logistics: lessons from supply chain management. *Int. J. Serv. Sci.* **2**(2), 142–161 (2009)
12. Wassenhove, L.N.: Humanitarian aid logistics: supply chain management in high gear. *J. Oper. Res. Soc.* **57**(5), 475–489 (2006)
13. Byman, D., Lesser, I.O., Pirnie, B.R., Benard, C., Waxman, M.: *Strengthening the Partnership: Improving Military Coordination with Relief Agencies and Allies in Humanitarian Operations*. Rand Corporation, Santa Monica, CA. Available at: www.rand.org/pubs/monograph%20reports/MR1185/ (accessed April, 2012)
14. Pettit, S.J., Beresford, A.K.C.: Emergency relief logistics: an evaluation of military, non-military, and composite response models. *Int. J. Logistics Res. Appl.* **8**(4), 313–331 (2006)
15. Chandes, J., Paché, G.: Investigating humanitarian logistics issues: from operations management to strategic action. *J. Manuf. Technol. Manage.* **21**(3), 320–340 (2010)
16. Tomasini, R.M., van Wassenhove, L.N.: Pan-American health organization’s humanitarian supply management system: de-politicization of the humanitarian supply chain by creating accountability. *J. Public Procurement* **4**(3), 437–449 (2004)
17. Simpson, G.R.: Just in time: in year of disasters, experts bring order to chaos of relief; logistics pros lend know-how to volunteer operations; leasing a fleet of forklifts; bottlenecks on the tarmac. *Wall Street J.* (Eastern edition), Nov 22, p. A1 (2005)
18. Charles, A.: Apprentissage croisé entre secteur humanitaire et industriel pour une meilleure gestion des chaînes logistiques, Actes du IXe’ Congre’s EDSYS, Toulouse, pp. 1–6 (CD-rom) (2008)
19. Gupta, A.: *Clothes for Work*. Outlook Business (2009)
20. Long, D.C., Wood, D.F.: The logistics of famine relief. *J. Bus. Logistics* **16**(1), 213–229 (1995)
21. Bhasin, S.: *The Material Man*. Tomorrow, India (2011)
22. Thomas, A.: *Humanitarian Logistics: Enabling Disaster Response*. White Paper, Fritz Institute, San Francisco, CA (2004)
23. Towill, D., Christopher, M.: The supply chain strategy conundrum: to be lean or agile or to be lean and agile. *Int. J. Logistics: Res. Appl.* **5**, 299–309 (2002)
24. Rajpal, U.: *The benevolent gatherer*. Log. India (2009)

Graduate School Application Advisor Based on Neural Classification System

Devarsh Bhonde, T. Sri Kalyan and Hari Sai Krishna Kanth

Abstract Neural classification systems are widely used in many fields for making logical decisions. This paper envisages a neural classification system based on back propagation algorithm to suggest an advisory model for graduate school admissions. It uses real and synthetically generated data to advise the students about the group of graduate schools where they have the maximum probability of getting selected. The system takes into consideration all the important aspects of the student's application such as: the GPA, GRE score, number of publications, professor recommendation, parent institute rating and work experience in order as to suggest the group of potential schools. A new parameter named Student Rating Index (SRI) is also defined for a better representation of the quality of professor recommendation. The system comprises of a two-layer feed-forward network, with sigmoid hidden and output neurons to classify the data sets. The results are verified using mean square error method, Receiver Operator Characteristic (ROC) curve and confusion matrices. The verification confirms that the proposed system is an accurate and reliable representation. Thus the proposed advisory system can be used by the students to make more focused applications in the graduate schools.

Keywords Neural classification · Graduate school application · Advisory model · Back propagation algorithm

1 Introduction

Across the globe, there may be numerous schools offering a particular graduate program. The students generally shortlist the graduate schools based on their rankings and on the recommendation of peers currently studying in those schools. There

D. Bhonde (✉) · T. Sri Kalyan · H. S. Krishna Kanth
Department of Civil Engineering, Indian Institute of Technology Kharagpur,
Kharagpur, India
e-mail: devarshbhonde@gmail.com

are discussion forums and admission counselors available for admission advice but no advisory model is present for the students to suggest the graduate schools where their chances of getting selected are the best. The total cost of completing the application procedure for different potential graduate schools can be very high.

In this project a neural classification system based on back propagation algorithm is proposed to advise the students about the group of graduate schools where have maximum probability of getting selected. The system takes into consideration all the important aspects of an application, such as: the Grade Point Average (GPA) of the student, GRE score, professor recommendation, number of publications, parent institute rating and work experience to suggest the result. A new parameter named Student Rating Index (SRI) is defined for a better representation of the quality of professor recommendation. The system is trained based on data available from prominent sources and it predicts the suitability of the student to get selected into each defined group of graduate schools. The system proposed is useful for the students to make more focused applications in the graduate schools where their chances of getting selected are high. It may also be used to reduce unnecessary expenditure on application costs for the group of graduate schools where their probability of selection is not good.

2 Neural Classification System

The neural network classification system helps in classifying various cases into a set of target categories based on various input parameters that represent the input cases. It has wide variety of applications in market forecasting, mortgage screening, loan advising etc. The graduate applications consists of various parameters that represents their academic and research performance during their undergraduate study. Due to the large number of input parameters available, a neural classification system based on back-propagation algorithm using real and synthetic data is developed. Artificial neural networks analyze data sets one by one, and learn by comparing the predicted classification of the data set with its actual classification. The calculated errors from the initial classification of the first set are fed back into the network, and are then used to modify the networks algorithm the second time and this process is repeated for n iterations. This process of learning from the error and updating the model is the basis of back-propagation algorithm.

2.1 Model Formulation

The model is formulated based on back-propagation algorithm using MATLAB's Neural Network Pattern Recognition Tool. This tool helps in developing a neural network to classify inputs into a set of target classes. The tool employs a two-layer feed-forward network, with sigmoid hidden and output neurons to classify the data sets.

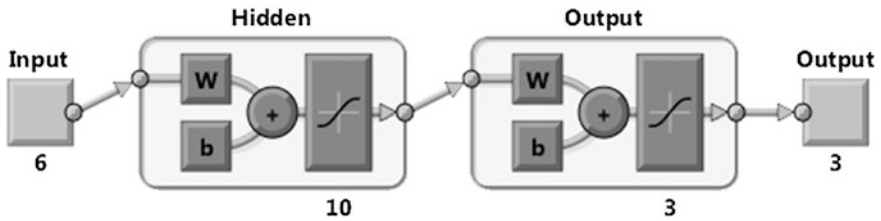


Fig. 1 Schematic of neural network (Source MATLAB Neural Network Pattern Recognition toolbox)

A total of 1,000 data sets with 6 input parameters, 10 neurons in the hidden layer and 3 neurons in the output layer are used for the creation of the model as shown in Fig. 1. In the model 70 % of the data sets are used for training, 15 % of the sets are used for validation and the rest 15 % are used for testing of the model. The input parameters used for each data set are:

1. Grade Point Average (GPA) of the student
2. Graduate Record Examination (GRE) score
3. Student Rating Index
4. Number of Publications
5. Parent Institute Rating
6. Work Experience.

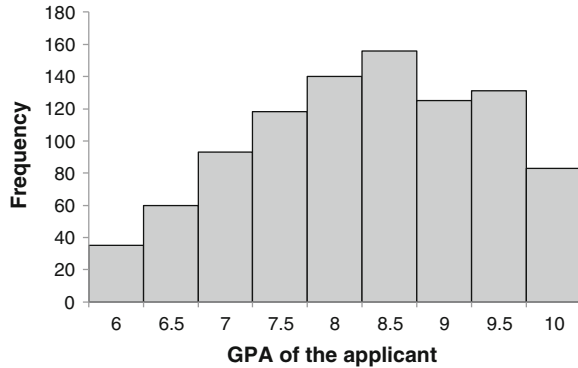
2.2 Result Representation

The proposed neural classification system takes the input parameters for an applicant into consideration and suggests the group of schools suitable for his profile. In order to group the schools, three grades of schools namely: Grade A, Grade B and Grade C are defined based on the ranking of schools available from the reputed ranking organization: QS Rankings (<http://www.topuniversities.com/>) Grade A schools have been defined as the graduate schools with ranking between 1 and 30, Grade B schools comprise of the schools having ranking between 31 and 70 and Grade C schools are defined as the schools having ranking between 71 and 100.

3 Input Data Collection

The input data sets required for training and formulating the neural network are collected from prominent sources or generated synthetically on the basis of observed trends. Detailed explanation of the input data used for representing different input parameters are as follows.

Fig. 2 Histogram depicting the GPA distribution among the applicants



3.1 Grade Point Average (GPA) of the Student

The GPA of a student is one of the most influential parameter in deciding the outcome of the application and hence is employed in the formulation of the proposed neural network model. The GPA of students studying at various graduate schools are collected from the online discussion forums: Gradcafe (<http://forum.thegradcafe.com/>) and Edulix (<http://www.edulix.com/forum/index.php>). From the data collected, it is observed that the GPA varies over the range of 6–10 on a 10 point scale. A histogram representing the range of GPA used for data description is shown in Fig. 2.

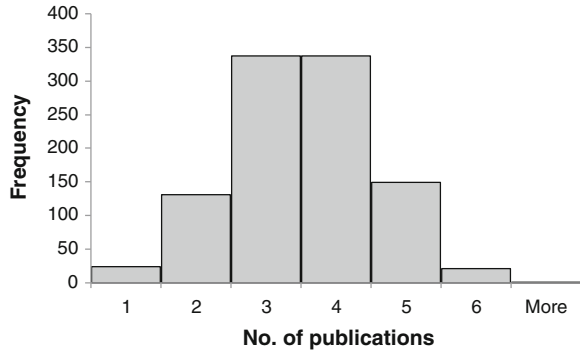
3.2 Graduate Record Examination (GRE) Score

The GRE score is a prerequisite for many universities and hence forms an essential part of every application. The GRE scores for previous year applicants are also collected from the online discussion forums: Gradcafe (<http://forum.thegradcafe.com/>) and Edulix (<http://www.edulix.com/forum/index.php>). The range of score observed in the data set is from 300 to 340 mostly ranging between 315 and 325.

3.3 Student Rating Index

A new parameter called as Student Rating Index is defined to account for more accurate representation of the professor recommendation value. This index is dependent on the rating of student given by the professor and also on the reputation of professor in their field of research. The h-index values of the professors measures the productivity and impact of their published work. The h-index values for various professors are collected from the citation website Scopus (<https://www>.

Fig. 3 Histogram showing the distribution of no. of publications among the applicants



scopus.com/home.url). Thus the final student rating index values ranging from 0 to 1 is formulated based on the following equation:

$$\text{Student rating index} = 0.7 * (\text{rating of student}) + 0.3 * (\text{normalized h-index}) \tag{1}$$

Where the rating of student is done on scale of 0–1 and normalized h-index values are also on a scale of 0–1. It is evident from Eq. 1 that more weightage is given to the rating given by the professor to the student.

3.4 Number of Publications

The publications of an applicant represents their research background and is given a high weightage while judging different applications. On account to the inbuilt difficulty in collecting real world data for the number of publications, it is assumed that it follows a normal probability distribution function with mean of three ($\mu = 3$) and standard deviation of one ($\sigma = 1$). The histogram depicting the distribution can be seen in Fig. 3.

3.5 Parent Institute Rating

The reputation of an applicant’s parent institute is considered an influential parameters towards the selection of an application as it represents the level of the competition they experienced and in-depth exposure to their respective fields. The rating values of various universities are collected from the ranking organization: QS Rankings (<http://www.topuniversities.com/>).

Fig. 4 Histogram depicting the distribution of work experience among the applicants



3.6 Work Experience

The work experience represents first hand application of a student's knowledge in a given field, which can be in the form of a job or an internship. Large data sets of work experience for different applicants was not available in the resources, hence synthetic data is generated for its representation. On observing the available trends of work experience for various applicants, a skewed distribution like lognormal probability distribution function is found to be more suitable to represent this parameter, since a large number of applicants have work experience mostly ranging from 1 to 2 years and there are fewer applicants with higher work experience. The lognormal distribution used for its representation can be seen in Fig. 4.

4 Example Application

The contribution of the applicant's GPA, GRE score, number of publications, parent institute rating, work experience and professor recommendation (which is reflected in the student rating index) in determining the probable grade of schools in which they can get admitted can be seen in Table 1. In the first example application, the student has a high GPA, two publications and good work experience. Even though the applicant has a GRE score which may be termed as low, the rest of his profile is too good to be rejected by a top notch university. The system takes these factors into consideration and suggests that the applicant has high probability of getting selected in Grade A schools. Similar trends are observed in the real world data, where the academic profile is given a higher importance than the GRE score. In the second case, the applicant has a low profile (low GPA and other important factors) and hence is suggested applying to Grade C schools with a suitability of 1. In the third case, the student has a good profile with moderate GPA as a result, the system suggests him to apply for Grade B schools to

Table 1 Sample example results for the proposed model

Sr. No	CPA	GRE store	No. of publication	Parent institute rating	Work experience	Student rating index	Suitability of getting admitted into grade			Probable grade of school
							A	B	C	
1	9	316	2	3	4	0.8	1	0	0	A
2	6.7	313	0	5	0	0.6	0	0	1	C
3	8.2	320	1	3	2	0.7	0	0.998	0.002	B
4	8.76	327	2	4	1	0.9	1	3E-04	0	A
5	8	315	1	4	1	0.75	0	0.624	0.376	B
6	8.5	320	2	4	2	0.75	0.57	0.43	0	A

improve his selection chances. In the fourth case, the overall profile of the applicant is great, which is in-turn reflected in the professor recommendation index too, hence the system recommends him to apply for Grade A schools. In the fifth case, the applicant has an above average profile, which is more appropriate for Grade B and Grade C schools. The system rightly predicts the suitability of the student getting admitted into a Grade B and Grade C schools as 0.624 and 0.376 respectively so that the student is advised to apply for both Grade B and Grade C schools for best results. Similar results are observed for the sixth case where the applicant’s profile is a border case between Grade A and Grade B schools. The model duly predicts the suitability of getting admitted to Grade A and Grade B schools as 0.57 and 0.43 respectively. Hence the results obtained are in agreement with the statistics available from various universities regarding their graduate admissions.

4.1 Verification of Results

The performance of the system is verified by determining the mean square errors, the Receiver Operator Characteristic (ROC) curve and by computing the confusion matrix.

The Mean Square Error (MSE). It is the average squared difference between the target and the outputs which indicates how accurate a model is. The MSE values obtained in the model during training, validation and testing are $4.27e-3$, $6.64e-3$ and $4.83e-3$ respectively (as shown in Table 2), indicating that the model is very accurate. The percent error which indicates the fraction of samples misclassified, has a very low value of $6.67e-1$ percent for the proposed model. It implies that the system fails just 6 in 1,000 times (or has accuracy of 99.33 %) thereby verifying the accuracy and the reliability of the results.

Receiver Operator Characteristic (ROC) curve. Another useful diagnostic tool used to get an idea about the accuracy of the model is the Receiver Operator Characteristic (ROC) curve. If threshold values are assigned to output in the range of 0–1 for each class of the classifier, the ROC represents the curve of true positive

Table 2 The mean square error and the percentage error for the proposed system (Source MATLAB Neural Network Pattern Recognition toolbox)

Process	Samples	MSE	Error (%)
Training	700	4.27709e-3	0
Validation	150	6.64836e-3	6.66666e-1
Testing	150	4.83072e-3	6.66666e-1

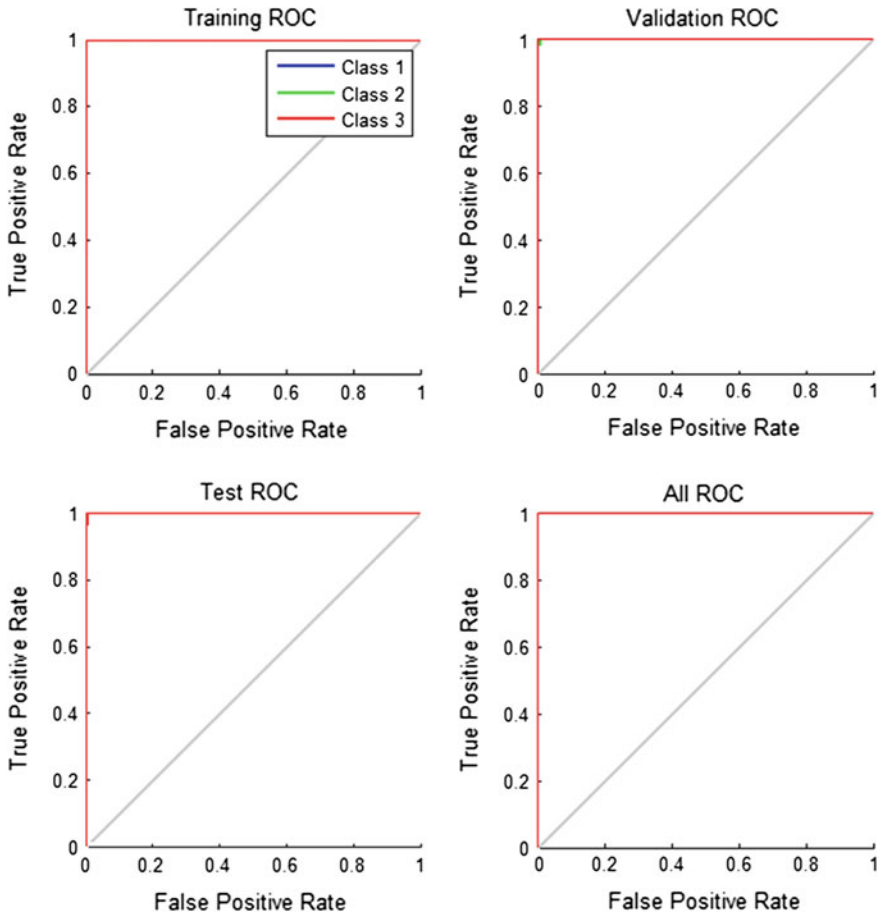


Fig. 5 Receiver operator characteristic curve for the proposed system (Source MATLAB Neural Network Pattern Recognition toolbox)

rate against the false positive rate, where the false positive rate is the ratio of the number of output values that are less than the threshold to the number of targets having a value of 0, and the true positive rate represents the ratio of the number of output values greater than or equal to the threshold to the number of targets having



Fig. 6 Confusion matrices for the proposed system (Source MATLAB Neural Network Pattern Recognition toolbox)

a value 1. This curve represents the inherent capacity of the model to discriminate different classes of outputs. From Fig. 5 it can clearly be seen that the upper left corner points have near 100 % specificity (false positive rate) and almost 100 % sensitivity (true positive rate). Hence it can be concluded that the model can accurately distinguish a particular class from the others.

Confusion matrices. The confusion matrices for training, validating, testing and the overall process can be seen in Fig. 6. It is observed that the output is accurate as the number of correct responses which are indicated in the green squares [squares with indices (1, 1), (2, 2), (3, 3)] are high and the number of incorrect responses represented in the red squares [squares with indices (1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 2)] are low. The overall accuracies indicated in the

lower right blue squares [squares with index (4, 4)] are high, justifying the reliability of the system.

It can be inferred from the verification of the results that the graduate school application advisory model based on neural classification system developed in this project is an accurate and reliable model which closely resembles the statistics available.

5 Conclusion

An applicant can have numerous options while applying for graduate school programs. In the present scenario, the applicants finalize potential graduate schools based on their rankings and advice of peers currently studying in those schools. The total cost of completing the application procedure for different schools can be very high. In this project a neural classification system is proposed to advise the applicants about the graduate schools where their chances of selection are good based on the student's application aspects namely: the Grade Point Average (GPA) of the student, GRE Score, professor recommendation, number of publications, parent institute rating and work experience to suggest the result. A new parameter named Student Rating Index (SRI) is defined for a better representation of the quality of professor recommendation. The model is trained based on the data from prominent sources and the results are verified using mean square error method, Receiver Operator Characteristic (ROC) curve and confusion matrices. The verification confirms that the proposed system is an accurate and reliable representation. Hence it can be used by the applicants to make more focused applications in the graduate schools where their chances of selection are the best. The model can also be used by the applicant as an advisor to reduce expenditure on the application costs of those graduate schools where his chances of selection are not good.

Acknowledgments The authors would like to thank Mr. Pushpal Mazumder of the Department of Civil Engineering at the Indian Institute of Technology Kharagpur for his help in the data extraction process.

References

1. Huang, M.H.: Opening the black box of QS World University Rankings. *Res. Eval.* **21**(1), 71–78 (2012)
2. Raghunathan, K.: Demystifying the American graduate admissions process (Online). Available <http://nlp.stanford.edu/~rkarthik/DAGAP.pdf> (2010). Accessed 5 June 2013
3. MathWorks: MATLAB 7.12 (R2011bSPM12), The Language of Technical Computing. The MathWorks, Inc., Natick, Massachusetts (2011)
4. Meho, L.I., Rogers, Y.: Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and web of science. *J. Am. Soc. Inf. Sci. Technol.* **59**(11), 1711–1726 (2008)

5. Pratihari, D.K.: Soft Computing. Alpha Science International Ltd (2007)
6. Quacquarelli Symonds: QS Top Universities (Online). Available <http://www.topuniversities.com/> (2013). Accessed 1 June 2013
7. SCOPUS 2013 (Online). Available <http://www.scopus.com/home.url> (2013). Accessed 29 May 2013

About the Editors

Dr. Millie Pant is Associate Professor with the Department of Paper Technology, Indian Institute of Technology Roorkee, Roorkee, India. She has to her credit several research papers in journals of national and international repute and is a well known figure in the field of Swarm Intelligence and Evolutionary Algorithms.

Prof. Kusum Deep is with the Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee, India. Over the last 25 years, her research is increasingly well-cited making her a central International figure in the area of Nature Inspired Optimization Techniques, Genetic Algorithms and Particle Swarm Optimization.

Dr. Jagdish Chand Bansal is Assistant Professor with the South Asian University, New Delhi, India. He is holding an excellent academic record and an excellent researcher in the field of Swarm Intelligence at the National and International Level. He has several research papers in journals of national and international repute.

Prof. Atulya Nagar holds the Foundation Chair of Computer and Mathematical Sciences and Heads the Department of Computer Science at Liverpool Hope University, Liverpool, UK. Prof. Nagar is an internationally recognised scholar working at the cutting edge of theoretical computer science, applied mathematical analysis, operations research, and systems engineering and his work is underpinned by strong complexity-theoretic foundations.

Author Index

A

Acharjya, D. P., 767
Agarwal, Himanshu, 577
Agarwal, Namrata, 801
Agarwal, Vernika, 883
Aggarwal, Mukul, 447
Aggarwal, Sugandha, 905
Ahmad, Abdul Nafey, 625
Akila, R. G., 31
Ali, Sadia Samar, 959
Aman, 265
Arora, Pratibha, 173
Arora, Sakshi, 601
Arora, Sonam, 217
Arun Bhaskar, M., 31
Arya, Anoop, 471
Arya, Leena, 167
Arya, Rajeev, 207

B

Bansal, Jagdish Chand, 787
Bera, Tushar Kanti, 377, 689, 703
Bhateja, Ashok K., 509
Bhonde, Devarsh, 975
Biswas, Samir Kumar, 703
Borisagar, Hetal G., 225
Boughaci, Dalila, 589

C

Chaturvedi, Ratnesh N., 613
Chaudhur, Biplab, 649
Chaurasia, Brijesh Kumar, 779

D

Daga, Pranam, 767
Daga, Pranjal, 767

Darbari, Jyoti Dhingra, 883
Das, Kedar Nath, 497, 649
Dash, S. S., 31
Deb, Suman, 113, 125
Deep, Kusum, 55
Deepa, S. N., 87
Deisy, C., 947
Deshpande, Prachi, 635
Din, Maiya, 509
Dubey, Manisha, 471
Dubey, Rameshwar, 959

G

Gaba, Navneet, 225
Galar, Diego, 867
Gandhi, Kanika, 867
Garg, M. L., 601
Garg, N. K., 529
Garg, Ritu, 133
Gaur, Madhu Sharma, 553
Gharpure, Damayanti, 337
Ghosh, Sanjoy, 349
Ghune, Nitish, 565
Goel, Samiksha, 831
Govindan, Kannan, 867
Gupta, Ankit, 433
Gupta, Anshu, 905
Gupta, Deepak, 423
Gupta, Jitendra Kumar, 279
Gupta, Prateek, 125
Gupta, Vaibhav, 741
Gupta, Vishal, 45

H

Hanumanthappa, M., 815
Hardaha, Mahesh Kumar, 237
Husain, Iqbal, 673, 845

J

Jadon, Shimpi Singh, 529, 787
 Jagadeesh Kumar, M., 31
 Jain, Shalini, 315
 Jain, Vikas Kumar, 845
 Jangid, Neeraj, 125
 Jaya Laxmi, A., 539
 Jha, P. C., 867, 883, 905, 929
 Joshi, A., 411
 Joshi, Ramesh, 757

K

Kamaraj, T., 457
 Katta, Ravinder, 577
 Kaul, Arshia, 905
 Kaur, Ramandeep, 929
 Kekre, H. B., 613
 Khanna, Anupam, 173
 Khare, Deepak, 237, 365, 391
 Kohli, Shruti, 217, 433
 Kripakaran, P., 87
 Krishna Kanth, Hari Sai, 975
 Krishna Murthy, A., 253
 Kulkarni, S. V., 757
 Kumar, Amit, 517
 Kumar, Pavan, 349
 Kumar, Shishir, 517
 Kumar, Yogendra, 471
 Kumari, Pushpa, 411
 Kumbhakar, C., 411
 Kundu, Sananda, 365, 391

L

Lal, A. K., 715
 Lala, Archana, 279
 Lalitha, D., 457
 Lokanadham, D., 63

M

Madan, Sushila, 929
 Mahajan, Varun, 303
 Malik, Latesh, 1
 Mamatha, M., 815
 Mani, Naveen, 45
 Meena, P. K., 237, 365
 Mehta, Neha, 167
 Menghal, P. M., 539
 Mirnalinee, T. T., 195, 403
 Mishra, P. K., 99, 365, 391
 Mishra, Prasanna R., 225
 Mondal, Arun, 237, 365, 391

Moses, Diana, 947
 Moshayedi, Ata Jahangir, 337

N

Nagaraju, J., 377, 689, 703
 Nailwal, Kewal Krishan, 423
 Naraina, A., 87
 Narula, Sonam, 929
 Nauriyal, D. K., 303
 Negi, Yuvraj Singh, 823
 Nitu, 133

P

Palwalia, D. K., 529
 Panchal, V. K., 831
 Pandey, R. K., 489
 Pandeya, Tarun, 325
 Pant, Bhaskar, 553
 Pant, Millie, 729, 801, 823
 Panwar, Poonam, 715
 Patel, Divyesh, 183
 Patel, Rahila, 1
 Patrai, Kalika, 143
 Peddoju, S. K., 635
 Prakash, B. R., 815
 Prasad, SVAV, 167

R

Raghunandan, K. S., 253
 Raghuwanshi, M. M., 1
 Rajan, K., 703
 Rajoria, Yogendra Kumar, 155
 Raman, Balasubramanian, 577
 Ramteke, Manojkumar, 565
 Ranjan, Jayanthi, 325
 Rao, C. S. P., 63
 Ratan, Ram, 509
 Rathor, Abhilasha Singh, 99
 Rezoug, Abdellah, 589

S

Saini, Seema, 155
 Sanju, P., 195
 Saxena, Neha, 17
 Sekhar, Chandra, 625
 Senthil, J., 767
 Shahi, Awanish, 779
 Sharma, Anurag, 75
 Sharma, Arpita, 831
 Sharma, Harish, 529, 787

Sharma, Kamal Kant, 447
Sharma, Madhavi, 279
Sharma, S. C., 207, 635
Sharma, Sameer, 423
Sharma, Swati, 289
Shrivastava, Tanuja, 183
Singathiya, R., 125
Singh, Amarjeet, 55
Singh, Gagandeep, 75
Singh, H. P., 75
Singh, Jai Prakash, 247
Singh, Koushendra K., 489
Singh, Manoj, 757
Singh, Pawan, 729
Singh, S. P., 303
Singh, S. R., 17, 155, 289, 315
Singh, Tapan Kumar, 497
Sinha, Subarna, 113
Sreeramulu, D., 63
Sri Kalyan, T., 975
Srivastava, Santosh Kumar, 673
Srivastava, Shilpa, 801
Subramani, C., 31
Suman, Bhupendra, 741
Suman, S., 489
Sunita, 741
Suresha, S., 253

T

Thepade, Sudeep D., 613
Thomas, D. G., 457
Tiwari, Ritu, 787
Tripathi, Pooja, 325
Trivedi, Kiran, 757
Trivedi, Vibhu, 565
Tyagi, Renu, 823

U

Uprety, Indu, 143

V

Venkatesh, V. G., 959
Verma, Shekhar, 779
Verma, Shiv Kumar, 183
Vora, Megha, 403

Y

Yadav, Dilip Kumar, 661
Yadav, Harikesh Bahadur, 661
Yadav, Neha, 447
Yadav, Siyaram, 265
Yashkar, Abhinav, 625